

TESIS DOCTORAL

**MODELADO ADAPTATIVO DEL MEDIO PARA LA
NAVEGACIÓN DE ROBOTS AUTÓNOMOS UTILIZANDO
ALGORITMOS BASADOS EN EL CENTRO DE ÁREAS**

José Manuel Cuadra Troncoso

Licenciado en Ciencias Matemáticas,
especialidad de Estadística e Investigación Operativa

Departamento de Inteligencia Artificial
Escuela Técnica Superior de Ingeniería Informática
Universidad Nacional de Educación a Distancia



Madrid, mayo 2011

Departamento de Inteligencia Artificial
Escuela Técnica Superior de Ingeniería Informática-UNED

**MODELADO ADAPTATIVO DEL MEDIO PARA LA
NAVEGACIÓN DE ROBOTS AUTÓNOMOS UTILIZANDO
ALGORITMOS BASADOS EN EL CENTRO DE ÁREAS**

Autor José Manuel Cuadra Troncoso

Licenciado en Ciencias Matemáticas,
especialidad de Estadística e Investigación Operativa

Director de la tesis: Dr. D. José Ramón Álvarez Sánchez

Co-Director de la tesis: Dr. D. Félix de la Paz Sánchez

TESIS DOCTORAL

**MODELADO ADAPTATIVO DEL MEDIO PARA LA
NAVEGACIÓN DE ROBOTS AUTÓNOMOS UTILIZANDO
ALGORITMOS BASADOS EN EL CENTRO DE ÁREAS**

Autor José Manuel Cuadra Troncoso

Director de la tesis: Dr. D. José Ramón Álvarez Sánchez

Co-Director de la tesis: Dr. D. Félix de la Paz Sánchez

Tribunal calificador

Presidente: Dr. D.

Secretario: Dr. D.

Vocal Primero: Dr. D.

Vocal Segundo: Dr. D.

Vocal Tercero: Dr. D.

Madrid, de mayo de 2011

Agradecimientos

Quisiera dedicar esta tesis a la memoria de nuestro amigo y maestro el Profesor José Mira, cuyas ideas sobre la Inteligencia Artificial impregnan por completo el espíritu de este trabajo.

Quiero agradecer la ayuda prestada por mis dos directores de tesis José Ramón Álvarez y Félix de la Paz, las interesantes discusiones y su colaboración en los experimentos. También a los becarios Dani, Ángel y Manuel por su ayuda en los experimentos. Y a todos los compañeros del departamento de Inteligencia Artificial de la Uned.

Expreso también mis agradecimientos a los proyectos TIN2004-07661-C02-01 y TIN-2007-67586-C02-01 del Ministerio de Ciencia e Innovación por la financiación parcial de esta tesis.

Gracias a mi padre por su ayuda para realizar los estudios de Doctorado.

Y especialmente quiero agradecer a mi mujer, Isabel, los innumerables sacrificios realizados durante el largo tiempo de preparación y realización de esta tesis.

*a Isabel y Natacha por el tiempo que no hemos podido estar juntos,
con todo mi amor.*

Índice general

1. Introducción general y objetivos	1
1.1. Motivación y objetivos	2
1.2. Estructura de la memoria	4
I Modelado adaptativo del medio	7
2. Estado del arte en segmentación	9
2.1. Planteamiento del problema	9
2.2. Clustering	10
2.2.1. Clustering basado en umbral de distancias	11
2.2.1.1. Detector Adaptativo de Puntos de Ruptura	11
2.2.1.2. Criterio de Dietmayer	11
2.2.1.3. Criterio de Santos	13
2.2.2. Otros tipos de umbral	13
2.2.3. Scale-space	13
2.2.3.1. Formulación teórica	14
2.2.3.2. Discretización	15
2.2.3.3. Árbol de intervalos	15
2.3. Extracción de líneas	16
2.3.1. Split and Merge	16
2.3.2. Line tracking	17
2.3.3. Random Consensus Sample (RANSAC)	17
2.3.4. Transformada de Hough	17
2.3.5. Expectation Maximization	18
2.4. Procedimientos de estimación	18
2.4.1. El método de regresión clásico	18
2.4.1.1. Mínimos cuadrados ordinarios (MCO)	19
2.4.1.2. Mínimos cuadrados generalizados (MCG) y heterocedasticidad	20
2.4.1.3. Modelos no lineales	21
2.4.2. Mínimos cuadrados totales	22

3. Filtros de Kalman para la estimación de líneas	25
3.1. Introducción	25
3.2. Estimación de modelos lineales mediante filtros de Kalman	25
3.2.1. Formulación del modelo de regresión lineal como sistema dinámico	26
3.2.2. Estimación MCO mediante filtro de Kalman	27
3.2.3. Estimación MCP mediante filtro de Kalman	29
3.2.3.1. Normalización de la ganancia del filtro	31
3.3. EKF para la estimación de líneas rectas en coordenadas polares	31
3.3.1. El EKF	32
3.3.2. Estimación de líneas	32
3.4. Experimentos lineales	35
3.4.1. Experimentos MCO	36
3.4.2. Experimentos MCG	36
3.4.3. Estudio del factor de normalización de ganancia	40
3.4.3.1. Primeros experimentos	40
3.4.3.2. Método de MonteCarlo	45
3.4.3.3. El caso general	51
3.5. Experimentos en coordenadas polares	55
3.5.1. Descripción	55
3.5.2. Discusión de resultados	57
3.5.2.1. 181 puntos	57
3.5.2.2. 35 puntos	64
3.5.3. El procedimiento bilateral	71
3.6. Demostraciones	72
3.6.1. MCP mediante filtro de Kalman	72
3.6.1.1. Relaciones entre estimaciones	72
3.6.1.2. Test de detección de outliers	73
3.6.1.3. Normalización de ganancia	74
4. Clustering mediante scale-space	75
4.1. Curvas de nivel	76
4.1.1. Discretización	77
4.1.1.1. Discretización de la convolución	77
4.1.1.2. Coste computacional de cálculo de diferencias	79
4.1.1.3. Niveles de la escala	80
4.1.2. Trazado	82
4.1.3. Filtrado del ruido	84
4.1.3.1. Área y altura de las curvas de nivel	84
4.1.3.2. Número de curvas de nivel en función del ruido	90
4.2. Árboles de intervalos	90
4.2.1. Rectángulos estables	90
4.2.2. Obtención de las semillas	92
4.3. Experimentos	93

4.3.1.	Descripción	93
4.3.2.	Resultados	95
4.3.2.1.	Usando el tamaño máximo para los núcleos	95
4.3.2.2.	Reducción del tamaño máximo núcleo y del número de niveles	96
5.	Extracción de líneas	99
5.1.	Introducción	99
5.2.	Inicialización del filtro: el método de la mediana repetida	100
5.2.1.	Razones para su empleo	100
5.2.2.	El tamaño de la semilla	101
5.3.	Estimación de segmentos	101
5.3.1.	Detección de extremos de segmentos	101
5.3.2.	Estimación y resolución de solapes	103
5.3.3.	Mezcla de segmentos	104
5.4.	Experimentos sobre fiabilidad	106
5.4.1.	Datos	106
5.4.2.	Estimación de segmentos por Mínimos Cuadrados Totales Generalizados Factibles	107
5.4.3.	Criterios de comparación	108
5.4.3.1.	Contrastes de parámetros	109
5.4.3.2.	Contraste de varianza	110
5.4.4.	Resultados	111
5.4.4.1.	M180	111
5.4.4.2.	M360	117
5.4.4.3.	M720	121
5.4.4.4.	Resumen	121
5.5.	Experimentos sobre parámetros de scale-space y eficiencia	121
5.5.1.	Experimentos sobre parámetros de scale-space	121
5.5.2.	Experimentos sobre eficiencia	123

II El método del centro de áreas 127

6.	Navegación reactiva	129
6.1.	Introducción	129
6.2.	Controles continuos	130
6.2.1.	Métodos de potencial	130
6.2.2.	Histograma de campo vectorial	131
6.2.3.	Ventana dinámica	131
6.3.	Aprendizaje	131
6.3.1.	Aprendizaje por refuerzo	131
6.3.2.	Redes neuronales	132
6.3.3.	Lógica borrosa	132

6.3.4. Algoritmos genéticos	132
6.4. El centro de áreas total	133
7. Centro de áreas básico	135
7.1. Introducción	135
7.2. Descripción del método	135
7.2.1. La hipótesis de trabajo	135
7.2.2. El proceso de división para centros de área inaccesibles	137
7.2.2.1. Descripción	137
7.2.3. Teniendo en cuenta el tamaño del robot	140
7.3. Alcance de objetivos	142
7.4. El control del movimiento	143
7.5. Experimentos simulados	144
7.5.1. Descripción del robot	144
7.5.2. Comparación con el procedimiento de wandering de la librería Aria	145
7.5.3. Evitación simple de obstáculos para alcanzar objetivos	147
7.5.4. Evitación de obstáculos complejos para alcanzar objetivos	149
7.6. Experimentos con robots reales	152
7.6.1. Descripción general	152
7.6.2. Localización del objetivo	152
7.6.3. Resultados	154
8. Centro de áreas con segmentación	157
8.1. Localización adecuada de los puntos de división	157
8.2. Puntos de división dinámicos	159
8.3. Experimentos	162
9. Conclusiones y trabajos futuros	165
9.1. Conclusiones y aportaciones	165
9.2. Trabajos futuros	167

Nomenclatura

CA/AC Centro de áreas/Area Center

DWA Dynamic Window Approach

EKF Extended Kalman Filter

EM Expectation maximization

FFT Fast Fourier Transform

LM Levenberg-Marquardt

M180 Mundo con 180 mediciones

M360 Mundo con 360 mediciones

M720 Mundo con 720 mediciones

MCG Mínimos Cuadrados Generalizados

MCO Mínimos Cuadrados Ordinarios

MCP Mínimos Cuadrados Poderados

MCT Mínimos Cuadrados Totales

NPE Número de puntos emparejados

NSE Número de segmentos emparejados

NSNE Número de segmentos no emparejados

NVC Número de varianzas correctas

RANSAC Random Sample Consensus

SCR Suma de cuadrados residual

SM Split and Merge

TLE Longitud total emparejada

VFF Virtual Forces Field

VFH Vectorial Field Histogram

Índice de figuras

2.1. Determinación del umbral adaptativo.	12
3.1. Representaciones de la recta	33
3.2. Estimación de una línea recta por MCO-Kalman	37
3.3. Estimación de una línea recta por MCG-Kalman 1	38
3.4. Estimación de una línea recta por MCG-Kalman 2	39
3.5. Evolución de los estadísticos de contraste	43
3.6. Comportamiento de la ganancia en función del número de puntos incluidos en el filtro, n y del factor de compensación ω . En a) para el parámetro a y en b) para b	46
3.7. Comportamiento de la ganancia	47
3.8. Errores en la estimación de rectas horizontales	48
3.9. Histogramas de d_t con 181 puntos	49
3.10. Histogramas de los errores no horizontal	53
3.11. Errores radiales	56
3.12. Comparación Kalman con LM 1	60
3.13. Comparación Kalman con LM 2	61
3.14. Comparación Kalman con LM 3	63
3.15. Comparación Kalman con LM 4	65
3.16. Comparación Kalman con LM 5	67
3.17. Comparación Kalman con LM 6	70
4.1. Mundo en coordenadas polares y rectangulares	76
4.2. Curvas de nivel continuas	78
4.3. Curvas de nivel continuas con ruido	81
4.4. Curvas de nivel discretizadas	83
4.5. Histogramas de áreas	85
4.6. Histogramas de las alturas 1	87
4.7. Histogramas de las alturas 2	88
4.8. Curvas de nivel filtradas	89
4.9. Rectángulos estables	91
4.10. Semillas obtenidas	95
4.11. Curvas de nivel filtradas 2	97

5.1. Inclusión de semillas	100
5.2. Errores en semillas	102
5.3. Secuencia de segmentación	105
5.4. Comparativa χ^2 en M180	112
5.5. Comparativa F en M180	113
5.6. Reconstrucción de M180 1	114
5.7. Reconstrucción de M180 2	115
5.8. Comparativa χ^2 en M360	118
5.9. Comparativa F en M360	119
5.10. Comparativa χ^2 en M720	120
5.11. Comparativa F en M720	122
5.12. Reconstrucción de M720 1	123
5.13. Reconstrucción de M720 2	124
5.14. Pruebas de velocidad	125
6.1. Campo de fuerzas generados	134
7.1. Fuentes de inspiración	136
7.2. Evitación de obstáculo	138
7.3. Divisiones sucesivas	141
7.4. Alcance de objetivos	142
7.5. Control de movimiento	144
7.6. Trayectorias en entorno doméstico	146
7.7. Trayectorias en circuito cerrado	147
7.8. Obstáculos circulares	148
7.9. Obstáculos cóncavos	148
7.10. Alcanzando objetivos en mundos complejos 1	149
7.11. Alcanzando objetivos en mundos complejos 2	150
7.12. Histogramas de rendimiento	151
7.13. Experimento con un robot real	153
8.1. Mala situación del centro de áreas 1	158
8.2. Mala situación del centro de áreas 2	160
8.3. Punto de división estáticos y dinámicos	161
8.4. Mundo atestado de obstáculos	163

Índice de algoritmos

2.1. Clustering adaptativo	11
3.1. Filtro de Kalman	26
3.2. MCO mediante filtro de Kalman	28
3.3. MCP mediante filtro de Kalman	30
3.4. Filtro extendido de Kalman EKF	32
3.5. MCP no lineal mediante filtro de Kalman	34

Índice de cuadros

3.1. Estimaciones de varias rectas horizontales	42
3.2. Medidas de los errores	52
3.3. Medidas de los errores	58
3.4. Medidas de los errores	66
3.5. Medidas de los errores 35 puntos	71
5.1. Inclusión de semillas	101
5.2. Magnitudes usadas en la comparación entre procedimientos de segmentación.	109

Capítulo 1

Introducción general y objetivos

Esta tesis tiene como objetivo la investigación de nuevos métodos de navegación reactiva para robots autónomos, considerada como base de cualquier otra forma de navegación de más alto nivel. En la navegación reactiva el robot se mueve en función únicamente de lo que percibe, no hay una planificación previa de la ruta a seguir. El tipo de robot que consideraremos en este trabajo es el robot terrestre equipado con sensores de rango 2D. Estos sensores proporcionan al robot mediciones de las distancias a las que se encuentran los objetos que le rodean, rescindiendo dichas mediciones a un plano paralelo al suelo. Estamos por lo tanto, en un contexto de movimientos y mediciones bidimensionales.

La memoria se estructura en dos partes unidas por el objetivo de navegación descrito en el párrafo anterior: el modelado del medio y la navegación mediante el método del centro de áreas. Un modelo del medio es una abstracción, normalmente realizada para algún propósito concreto. Con centro de áreas nos referimos al centro de masas o gravedad, en 2D, del área libre percibida por el robot. Aunque la memoria se estructura en dos partes, la investigación que se expone en este trabajo se realizó en tres fases. La primera fue el desarrollo del sistema de navegación reactiva basado en el centro de áreas usando un modelo del medio muy básico, sin ningún tipo de preprocesamiento de las medidas sensoriales. Tras una extensa investigación se vio que para superar varios problemas que el método presentaba, era necesario emplear un modelo del medio más adecuado, por lo que se inició la investigación sobre el modelado del medio. Una vez se dispuso una forma adecuada de modelar el entorno, se usó para mejorar el método de navegación reactiva mediante el centro de áreas.

Ha parecido más adecuado, por claridad y unidad en la exposición, no seguir el orden cronológico de la investigación sino agrupar los hitos de la investigación según las áreas a las que pertenecen: el modelado y la navegación. Por esta razón, se han escritos estados el arte distintos para cada área que se incluyen al principio cada una de las dos partes en la que se estructura la memoria, capítulos 2 y 6, durante el resto de esta introducción remitimos a ellos al lector para la posible consulta de términos y referencias. Pasaremos a describir las motivaciones que condujeron al desarrollo de esta tesis y los objetivos perseguidos en ese desarrollo.

1.1. Motivación y objetivos

En la navegación puramente reactiva, en robots o animales, no hay posibilidades de planificación de trayectorias más allá de un corto espacio, trayectorias locales. Supongamos que nos encontramos en un bosque espeso del que podemos escapar siguiendo el tañido de una campana situada fuera de él. Árboles, arbustos y rocas nos impedirían diseñar trayectorias de más de pocos metros en nuestro camino, el sonido de la campana solo no da una dirección aproximada, que se desea seguir. En una situación como la descrita, bosque espeso, nos vemos muy limitados en el uso de nuestra principal arma para diseñar trayectorias: la visión. La trayectoria vendrá dada por el modelo del medio del que se disponga y éste, a su vez, por las características de sistema sensorial utilizado, una hormiga necesita tocar un obstáculo con las antenas para percibirlo, un humano lo ve antes y lo evita. El comportamiento de la hormiga es similar al de los robots simples equipados con sensores de rango de corto alcance, en los que los sensores actúan al modo de antenas. Podríamos decir en este caso se tiene un modelo del entorno con dos estados: obstáculo (puntual) o área libre y, simplificada, las trayectorias generadas por el modelo serán la concatenación de segmentos rectilíneos. En el caso de los humanos, para explicar la acción a distancia de los obstáculos, se puede pensar un modelo en el que los obstáculos generan un campo continuo de fuerzas repulsivas que desvían la trayectoria, ahora las trayectorias serán curvas suaves, ver subsección 6.2.1.

El problema que se plantea es hacia dónde se dirigen dichas fuerzas, o dicho de otra manera, cuál es la forma del campo. Si se dispone de un plano geométrico del entorno, aunque solo sea una sucesión de planos locales, es posible calcular la expresión del campo bajo alguna condición: campo de fuerzas central, campo derivado de un potencial, etc. Sin embargo, si solo se dispone de un conjunto de mediciones tomadas en un muestreo de los sensores, se debería crear dicho mapa para poder definir el campo. Supongamos que se tuviera tal mapa, aún queda por dilucidar la cuestión de la forma del campo creado por un objeto determinado, cuál es la dirección apropiada de las fuerzas para el objetivo de navegación. Pensemos en la primera solución que se viene a la mente, vectores perpendiculares a la superficie del objeto en cada punto, está bien para evitar chocar con el obstáculo, pero tiene un problema, es local, no tiene en cuenta los otros objetos. Aunque el campo efectivo total de todos los obstáculos integra los campos locales depende de dichos campos, por ejemplo si la dirección del campo se definiera como orientada al robot en cada punto el movimiento resultante sería bastante distinto. Por esta razón surge la cuestión de si es posible encontrar una manera “natural” definir el campo, que tenga en cuenta la distribución de obstáculos del entorno y el propósito de navegación.

La solución adoptada para el problema de la elección del campo consiste en un cambio del punto de vista en la forma de modelar el entorno, en lugar de modelar los obstáculos se modela el espacio libre que dejan. Si el objetivo es evitar los obstáculos, parece una buena idea fijarse en el “centro” del área libre, es fácil que esté lejos de los obstáculos, como sumidero de dicho campo de manera que atraiga al robot hacia él. No son los obstáculos los que ejercen fuerzas repulsivas, es el centro del área libre percibida el que ejerce una atracción que determina el movimiento. Dado que el centro del área viene determinado

por las posiciones de todos los puntos de su borde, las direcciones de campos locales en la superficie de los objetos serán interdependientes. Una forma de dar una expresión de dicho campo, o mejor dicho de sus efectos, es hacer que el movimiento del robot dependa de su posición con respecto al centro del área libre. Esta definición del campo no supone conocida la forma de los obstáculos que delimitan el área libre, se puede obtener rápida y fácilmente de las mediciones sensoriales directas, no precisa mapa. El modelo del medio que se tiene es, el conjunto mediciones, que se puede representar como un polígono con tantos vértices como mediciones, y un campo que en todo punto apunta hacia el centro de áreas del polígono.

Como se verá en el desarrollo de esta tesis, ésta forma de definir el campo, aunque solo usa mediciones directas sin procesar, supera en rendimiento a los métodos que elaboran alguna especie de mapa para calcular el campo. La evitación de obstáculos no es el propósito de la navegación, es una condición, el principal propósito es alcanzar un objetivo. En el modelo que describe los obstáculos como repulsivos los objetivos son definidos como puntos de atracción. En el método que se propone ya se tiene el centro del área libre como atractor, por lo que añadir los objetivos como atractores puede ser confuso y contradictorio. La forma en la que se introducen los objetivos en el modelo del entorno es la siguiente: la dirección del objetivo induce una división del área libre en dos sectores y se elige el que está en la dirección del objeto, este planteamiento recuerda algo al de los métodos de navegación no basados en campos, ver subsecciones 6.2.2 y 6.2.3.

Los resultados del método se pueden mejorar si las mediciones se agrupan en entidades con significado para la navegación, estas entidades suelen ser segmentos. En este caso los extremos de los segmentos constituyen las características del entorno a tener cuenta para la navegación, lo difícil de evitar no es la parte plana de un segmento sino las esquinas salientes. Denominaremos al procedimiento para hallar los segmentos segmentación, ver sección 2.2 para una aclaración del término.

Los procedimientos de segmentación se han usado con el objetivo de creación de mapas y para la localización del robot. En la localización se intenta emparejar el conjunto de medidas tomado desde una posición con el mapa del entorno, principalmente usando algún método estadístico. No es el objetivo de esta tesis llegar a plantear la cuestión de la localización, aunque el método que se propone esté sólidamente diseñado desde el punto de vista estadístico, sino usar la segmentación para obtener un modelo del medio más eficiente para la navegación reactiva, que el que proporcionan las medidas sin procesar.

Durante estudio de los métodos de segmentación llamó la atención el tipo de entornos que usualmente se realizan experimentos de segmentación: interiores de edificios normalmente caracterizados por paredes lisas que se cortan el ángulo recto. En estas situaciones el modelo de segmento coincide bastante con el objeto físico que representa, la pared plana, incluso visto de cerca. Sin embargo hay que separarse un tanto de un seto recto para verlo como un segmento rectilíneo. Por lo que se planteó la cuestión de intentar diseñar un procedimiento que fuera robusto, adaptativo, ante lo que difieran las mediciones con respecto al modelo, las mediciones pueden diferir bastante del modelo aunque el sensor sea preciso. El procedimiento se ha diseñado para una forma particular de tomar las mediciones, radialmente, de manera que el procedimiento se adapte al funcionamiento a los sensores

de rango que toman medidas de esa forma. Otra cuestión relacionada con la robustez y adaptabilidad de un método, es si posee o no parámetros de efecto crítico que el usuario deba ajustar, en caso de poseerlos es difícil conseguir la robustez y la adaptabilidad.

1.2. Estructura de la memoria

La memoria de esta tesis se estructura en los siguientes capítulos:

1. Introducción general y objetivos
2. Estado del arte en segmentación
3. Filtros de Kalman para la estimación de líneas
4. Clustering mediante scale-space
5. Extracción de líneas
6. Navegación reactiva
7. Centro de áreas básico
8. Centro de áreas con segmentación
9. Conclusiones y trabajos futuros

Aparte de este primer capítulo y del último, la memoria se estructura en dos partes: “Modelado adaptativo del medio”, capítulos 2 a 5, y “El método del centro de áreas”, capítulos 6 a 8.

La primera parte, como se comentó al principio de este capítulo, trata del problema de la segmentación.

En el capítulo *Estado del arte en segmentación* se estudia el problema de la segmentación en dos dimensiones, se proporciona un estado del arte y se describen las bases teóricas de las que parten de los métodos de segmentación que se han desarrollado para esta tesis.

En el capítulo *Filtros de Kalman para la estimación de líneas* se desarrolla un filtro de Kalman extendido, que permite la estimación de líneas utilizando la ecuación de la recta en coordenadas polares. El diseño del filtro lo hace equivalente a un modelo de regresión clásico, con lo que se consigue disponer de todas las técnicas y resultados de la teoría de la regresión clásica: propiedades de estimadores, contrastes de hipótesis, detección de outliers, etc. El diseño permite ajustar el filtro a distintos tipos de modelos de ruido, en los que el nivel de éste puede variar. Se describen pruebas bastante exhaustivas realizadas para comprobar el rendimiento del método.

En *Clustering mediante scale-space* se desarrolla un sistema adaptativo de filtrado de ruido para la técnica agrupación (clustering) conocida como scale-space (espacio de escalas). Mediante procedimientos estadísticos se realiza un filtrado que logra eliminar gran

parte de la fragmentación de grupos, segmentos, que produce el método al ser aplicado al tipo de señales objeto de este estudio. El objetivo del procedimiento de clustering es determinar zonas donde comenzar la estimación mediante los filtros de Kalman desarrollados en el capítulo anterior.

En *Extracción de líneas* se utilizan los procedimientos expuestos en los dos capítulos anteriores para conseguir una segmentación del entorno. La capacidad de detección de outliers del filtro de Kalman se usa para la detección de los extremos de los segmentos y se utiliza un test, tomado de la teoría de la regresión clásica, para unir segmentos contiguos similares. También se describen experimentos de comparación estadística con otros métodos, en los que se estudió el comportamiento frente al diversos niveles de ruido.

La segunda parte trata de la cuestión de la navegación mediante centros de áreas.

En el capítulo *Navegación reactiva* se hace una introducción al problema de la navegación reactiva, se repasa el estado del arte de la cuestión y se exponen los trabajos sobre el centro de áreas previos a esta tesis.

En *Centro de áreas básico* se describe el método de navegación basado en el centro de áreas en su versión básica, esto es, sin procesamiento de los sensores. Se trata la cuestión de los centros de área inaccesibles y la modificación del método para alcanzar objetivos. Experimentos con resultados cuantitativos y cualitativos son descritos.

En *Centro de áreas con segmentación* se usan los métodos de segmentación, desarrollados en la primera parte, para resolver distintos problemas observados en el método básico.

Para terminar en *Conclusiones y trabajos futuros* se comentan los resultados obtenidos, las aportaciones originales y se dan líneas futuras de investigación que puede abrir esta tesis.

Parte I

Modelado adaptativo del medio

Capítulo 2

Estado del arte en segmentación

2.1. Planteamiento del problema

Cuando se nos pide dibujar un plano geométrico de un lugar tendemos a usar formas simples en su descripción, o sea, realizamos un modelo geométrico simplificado del lugar, el grado de simplicidad o complicación vendrá determinado por la tarea a realizar con dicho plano. Si se dispone de mediciones de rango y se intenta construir un determinado modelo geométrico a partir de ellas, nos encontramos con la cuestión del error de dichas mediciones con respecto al modelo. Vamos a aclarar el concepto de error de la medición con respecto al modelo, lo denominaremos error intrínseco de la superficie, que no es el mismo que el error de medición del dispositivo. Una pared recta lisa, una superficie lateral de un seto recto o un lienzo de muralla recto derruido cuando se observan a cierta distancia, siguen el modelo de línea recta o más concretamente de segmento rectilíneo. Sin embargo las diferencias de las posibles mediciones que se realicen sobre estos objetos, tendrán una variación con respecto a la línea recta ideal que cambiará de forma apreciable de un caso a otro. Este tipo de error puede variar aunque el error del dispositivo de medición fuera el mismo al medir sobre una pared, una hoja o un bloque de piedra basta.

Aunque el error del dispositivo influye en el intrínseco, puede ser mucho más pequeño dada la alta precisión de muchos dispositivos para mediciones de rango. Al error intrínseco descrito se pueden unir otras muchas fuentes de error como, características de reflectancia de la superficie, iluminación, distancia al objetivo, etc. Por esto existen muchos ambientes, especialmente en exteriores, donde el error del dispositivo queda totalmente enmascarado por el intrínseco y los debidos a otras fuentes. Es el objetivo de esta parte de la tesis el proponer un método de modelado del entorno que pueda funcionar en entornos donde el error intrínseco pueda ser variable y desconocido a priori, por lo que en la revisión del estado del arte se hará hincapié en estos aspectos. Si un procedimiento pretende ser tan flexible y robusto como se requiere, no debe poseer parámetros que tengan una influencia crítica en los resultados y que deban ser ajustados por el usuario, el procedimiento debe ser todo lo adaptativo posible, este aspecto también será seguido con interés en la revisión que se realizará.

Tras estas consideraciones pasamos a plantear en términos más concretos el término segmentación, restringiéndonos especialmente a los segmentos rectilíneos. El problema de modelar un conjunto de mediciones como segmentos, que denominaremos segmentación, plantea tres cuestiones según Forsyth and Ponce (2003): el número de líneas existentes, la pertenencia de un determinado punto a un segmento y la estimación de los parámetros de las líneas. La primera y segunda cuestión se pueden unir en una sola: detectar los extremos o puntos de ruptura (break points) de todos los segmentos. La detección de extremos y la estimación se pueden abordar de forma separada o conjunta. En la primera forma dos procedimientos ejecutados secuencialmente se encargarían de cada una de las tareas, en la segunda un único procedimiento realizaría ambas tareas conjuntamente. Un procedimiento que solo busque extremos de segmentos es denominado procedimiento de agrupación de puntos (clustering), usaremos su traducción inglesa por ser el término habitual al uso. Un procedimiento que busque extremos de segmentos a la vez que estima sus parámetros, es denominado procedimiento de extracción de líneas (line extraction). Es posible aplicar un procedimiento de clustering previo al de extracción para intentar obtener más información sobre los extremos. Es este trabajo se usará el término segmentación para indicar un procedimiento que proporcione segmentos estimados, ya sea aplicando clustering más estimación, extracción de líneas o ambos.

A continuación se describirán los algoritmos más usuales de clustering, de extracción de líneas y las técnicas de estimación especialmente orientados a la obtención de segmentos a partir de medidas de rango. La exposición está orientada a segmentar un conjunto de medidas obtenidas en un solo barrido, no las provenientes de la agrupación de varios.

2.2. Clustering

En esta sección se va a realizar un repaso de las técnicas de clustering para la obtención de segmentos a partir de medidas de rango, se comentarán las más usuales y otras relacionadas con la empleada en este trabajo, que se describirá en el capítulo 4.

La gran mayoría de los procedimientos de clustering, para decidir si un punto P_{n+1} pertenece al segmento formado por los puntos anteriores del haz establecen un valor umbral D , normalmente la distancia al punto anterior P_n , de manera que si dicha distancia es superior a D el punto P_{n+1} se considera como el inicio de un nuevo segmento. A la hora de la elección del umbral se plantean dos cuestiones. La primera es que para puntos proyectados radialmente sobre una línea no existe una cota para la distancia entre dos puntos consecutivos, depende de los ángulos de incidencia. La segunda es que puntos, cercanos a los extremos, de segmentos consecutivos no concurrentes pueden estar bastante próximos. Si el valor umbral es fijo y se le va un valor pequeño se podrán separar segmentos próximos pero las zonas de los segmentos donde el ángulo de incidencia sea un poco elevado no serán clasificadas, si se le da un valor mayor se corre el riesgo de perder segmentos. Por esta razón se han desarrollado procedimientos con umbral adaptativo.

En nuestro trabajo no se ha usado una técnica de clustering basada en umbral sino técnicas, usadas en visión artificial y en tratamiento de señales, de multi-escala (scale-

space), usaremos su traducción inglesa por ser el término habitual al uso. Por esta razón incluiremos una subsección para las descripción de estas técnicas.

2.2.1. Clustering basado en umbral de distancias

El clustering basado en umbral de distancias básico, no adaptativo, ya ha sido descrito y comentado en la introducción de esta sección, es muy simple y padece los problemas comentados, pero aún así ha sido utilizado en algún trabajo Mastrogiovanni et al. (2005); Nguyen et al. (2007).

En cuanto al clustering con umbral adaptativo D_n , su descripción puede verse en el algoritmo 2.1. Criterios para el cálculo del umbral se describen a continuación.

Algoritmo 2.1 Clustering adaptativo

```

1: for  $n = 1$  to  $numPoints - 1$  do
2:   calcular  $D_n$  según criterio
3:   if  $dist(P_n, P_{n+1}) > D_n$  then
4:      $P_{n+1}$  comienza nuevo segmento
5:   else
6:      $P_{n+1}$  se incluye en el segmento actual
7:   end if
8: end for

```

2.2.1.1. Detector Adaptativo de Puntos de Ruptura

El Detector Adaptativo de Puntos de Ruptura (Adaptive Break Point Detector) fue propuesto en Borges and Aldon (2004). Utiliza un parámetro, λ , que es el ángulo complementario del ángulo de incidencia máximo que el método o el dispositivo fijen. Consideremos la recta que pasa por P_n y forma un ángulo λ con el con el rayo del haz que pasa por P_n y la recta definida por el siguiente rayo del haz, su intersección nos proporciona un punto Q , se toma entonces $D_n = dist(P_n, Q)$, ver figura 2.1 a).

2.2.1.2. Criterio de Dietmayer

En Dietmayer et al. (2001) el umbral se define así $D_n = C_0 + C_1 \min(r_n, r_{n+1})$, donde r representa la distancia medida, C_0 es una constante relacionada con el ruido y $C_1 = \sqrt{2(1 - \cos \Delta\varphi)}$, $\Delta\varphi$ es la resolución angular. $C_1 \min(r_n, r_{n+1})$ es la distancia entre P_n y el punto, P'_n , situado en la línea definida por el siguiente rayo del haz y a la misma distancia del origen que P_n , si $r_n < r_{n+1}$, en caso contrario se usa P_{n+1} en lugar de P_n , ver figura 2.1 b).

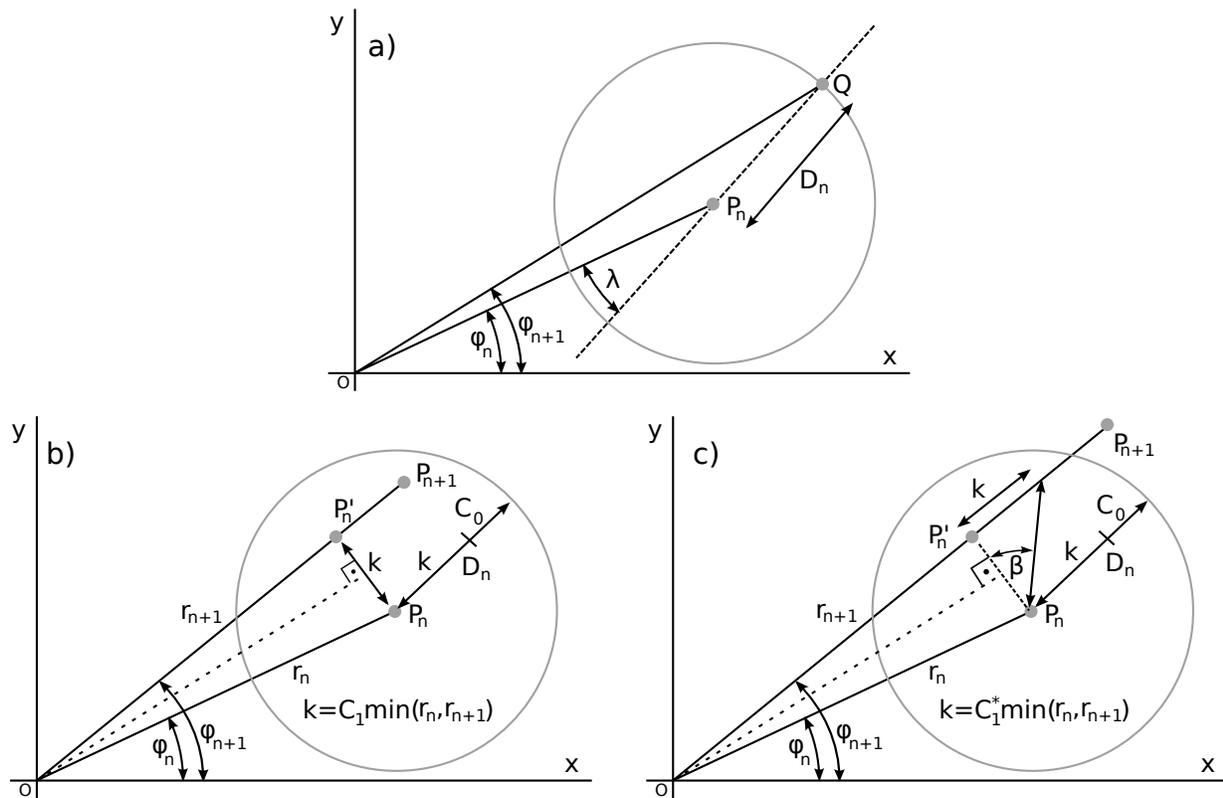


Figura 2.1: Determinación del umbral adaptativo. En a) el criterio usado por el Detector Adaptativo de Puntos de Ruptura, en b) el criterio de Dietmayer y en c) el criterio de Santos.

2.2.1.3. Criterio de Santos

Dado que el criterio de Dietmayer depende de la distancia al objeto en Santos et al. (2003) se cambia la expresión del parámetro C_1 de la siguiente manera:

$$C_1^* = \frac{\sqrt{2(1 - \cos \Delta\varphi)}}{\cot(\beta) \cos\left(\frac{\Delta\varphi}{2}\right) - \sin\left(\frac{\Delta\varphi}{2}\right)}$$

donde β es un ángulo, a fijar, que representa la máxima inclinación que la superficie de un objeto, conteniendo a P_{n+1} , puede presentar con respecto a la bisectriz de las líneas definidas por P_n y P_{n+1} , para poder ser considerado como perteneciente al segmento de P_n , ver figura 2.1 c).

2.2.2. Otros tipos de umbral

Harati and Siegwart (2007) proponen un umbral basado en ángulos además de en distancias. Se define el ángulo de orientación γ_n (bearing) como el ángulo formado por el rayo del haz correspondiente a P_n y la línea definida por P_n y P_{n-1} . Se establece un valor umbral fijo Φ , si $|\gamma_n - \gamma_{n+1}| < \Phi$ y no se alcanza el umbral de distancia, P_{n+1} pasa a formar parte del segmento de P_n .

Para incrementar la robustez del procedimiento frente al ruido los autores proponen usar otro ángulo denominado ángulo de movimiento (running) $\phi_n = \gamma_n + \varphi_n$, como siempre φ representa el ángulo del rayo en el sistema de coordenadas de robot. A los valores de ϕ se les aplica la transformada de wavelets discreta Stollnitz et al. (1996), y se calculan los valores $\tilde{\phi}$ a partir de una reconstrucción suavizada de los ϕ , el nivel de suavización viene determinado por dos parámetros a fijar. El criterio de clustering queda $|\tilde{\phi}_n - \tilde{\phi}_{n+1}| < \Phi$.

2.2.3. Scale-space

Al plantear el problema que nos ocupa en la sección 2.1, no se mencionó la cuestión de la escala que se debía usar en el plano, o dicho de otra manera el nivel de detalle. Visto desde esta perspectiva el problema del clustering se convierte en encontrar la escala apropiada (natural) para “ver” el mundo como un conjunto de segmentos. La cuestión de las escalas naturales y métodos para intentar hallarlas son analizados en Rosin (1994). En esta subsección sólo nos dedicaremos a explicar la técnica seleccionada para este trabajo scale-space. Pruebas apuntan a una relación bastante aproximada entre los operadores de la teoría y los perfiles de los campos receptivos en la retina y el cortex visual de los mamíferos Lindeberg (2009).

El concepto de scale-space aparece por primera vez en Witkin (1984), donde una señal 1D $f(x)$ es convolucionada con núcleos gaussianos de varianza t creciente para generar una superficie 2D, $F(x, t)$. En esta superficie los cambios de la señal aparecen como ceros de las primeras derivadas de F , $\frac{\partial^k F}{\partial x^k}$, máximos, mínimos y puntos de inflexión. Dado que en este trabajo se manejarán las mediciones en coordenadas polares tendremos siempre señales

1D por lo que restringiremos esta exposición a dicho tipo de señales. Posteriormente se ha desarrollado la teoría matemática del scale-space, principalmente por los trabajos de Lindeberg (1990, 1991, 1992, 1993, 1997), pasamos a comentar brevemente esta teoría.

Para no repetir figuras remitiremos al lector a las del capítulo 4.

2.2.3.1. Formulación teórica

Dada una señal unidimensional $f(x)$ de denomina representación scale-space lineal de $f : \mathbb{R} \rightarrow \mathbb{R}$ a $F(x, t) = T(u; t) \star f(x)$, donde $t > 0$ es el parámetro de escala y $T(u; t)$ los núcleos de convolución. Al incrementarse t se van produciendo versiones cada vez más suavizadas de f que conforman una superficie bidimensional, ver figura 4.2, allí t es designado σ_s^2 . Cuando los núcleos de convolución son gaussianos $T(u; t) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{u^2}{2t}\right)$, t es la varianza, se obtienen ciertas propiedades conocidas como axiomas de scale-space, indicaremos mediante $T_t(f)$ al anterior producto de convolución:

1. Linealidad: $T_t(af + bg) = aT_t(f) + bT_t(g)$, donde $a, b \in \mathbb{R}$ y f y g son funciones reales.
2. Invariancia frente a traslaciones: si S_a designa al operador traslación $S_a(f(x)) = f(x - a)$ se tiene que $T_t(S_a(f)) = S_a(T_t(f))$.
3. Estructura de semigrupo: $T(u; t_1) * T(u; t_2) = T(u; t_1 + t_2)$, que lleva asociada la propiedad de suavización en cascada, $F(x, t_2) = T(u; t_2 - t_1) * F(x, t_1)$.
4. Existencia de generador infinitesimal A : $\frac{\partial F(x; t)}{\partial t} = (AF)(x; t)$
5. No creación de extremos locales, esta propiedad solo se da en una dimensión.
6. No se realzan los extremos locales $\frac{\partial F(x; t)}{\partial t} \leq 0$ en los máximos espaciales y $\frac{\partial F(x; t)}{\partial t} \geq 0$ en los mínimos espaciales, esta propiedad se da en cualquier número de dimensiones.
7. Simetría rotacional: existe una función h tal que $T(u; t) = h(u^2; t)$.
8. Invariancia de escala: $\hat{T}(\omega_u; t) = \hat{h}\left(\frac{\omega_u}{\varphi(t)}\right)$, donde \hat{T} es la transformada de Fourier de T , φ es cierta función y \hat{h} su transformada.
9. Positividad. $T(x; t) \geq 0$
10. Normalización: $\int_{-\infty}^{\infty} T(u; t) dt = 1$.

La propiedad 5 permite establecer una relación de la teoría con ecuaciones en derivadas parciales como la ecuación de difusión del calor, podríamos imaginar que t es el tiempo durante el que se está “derritiendo” f , produciéndose sus versiones suavizadas. Nos asegura que no aparecerán nuevos extremos a escalas inferiores que no estén ya presentes en escalas superiores. También garantizan las propiedades que dado un extremo localizado en una escala determinada será posible rastrear su posición hasta las escalas inferiores. Dado que

el problema de encontrar extremos equivales al de hallar las curvas de nivel cero (zero-crossing) de $\frac{\partial^k F(x;t)}{\partial t^k}$, se tendrá que todas estas curvas han de comenzar y terminar en el nivel inferior.

2.2.3.2. Discretización

Para aplicar el procedimiento a señales discretas $f : \mathbb{Z} \rightarrow \mathbb{R}$ es necesario encontrar un tipo de núcleos para los que se cumplan las propiedades citadas en la subsección 2.2.3.1. El único tipo de núcleos para los que cumplen dichas propiedades son los análogos discretos de los núcleos gaussianos, que no hay que confundir con núcleos gaussianos muestreados en los que se toman los valores para los núcleos de la función de densidad de probabilidad normal.

La expresión de la superficie F como convolución discreta vendrá dada por

$$F(x_n, t) = T(m; t) * f(x_n) = \sum_{m=-\infty}^{\infty} T(m, t) f(x_n - m) \quad (2.1)$$

$T(m, t)$ son los análogos discretos de los núcleos gaussianos:

$$T(m, t) = e^{-\alpha t} I_m(\alpha t) = e^{-t} \sum_{k=0}^{\infty} \frac{1}{k! \Gamma(k + m + 1)} \left(\frac{t}{2}\right)^{2k+m} \quad (2.2)$$

donde $I_m(t)$ representa las funciones modificadas de Bessel de orden entero m y $\alpha \geq 0$, se puede hacer $\alpha = 1$ sin pérdida de generalidad.

Otra cuestión es el necesario truncamiento del sumatorio de la convolución para hacerlo computable, de manera que el cálculo efectivo se realiza así:

$$F(x_n, t) = T * f = \sum_{m=-M}^M T(m, t) f(x_n - m)$$

M suele tomarse como $M = \lceil c\sqrt{t} \rceil + 1$, o $M = \lceil c\sqrt{t} \rceil$ si $c\sqrt{t}$ es entero, $\lceil \cdot \rceil$ denota la parte entera, con c entre 3 y 6, estos valores corresponden a probabilidades superiores a 0,999. Denominaremos a M como la semi-amplitud del núcleo.

Otras aproximaciones, que ya no cumplen todas las propiedades deseadas son posibles como: núcleos gaussianos muestreados, filtros cuya función de transferencia aproxima a la de $T(m, t)$ o aproximaciones por pirámides Lindeberg and Bretzner (2003).

2.2.3.3. Árbol de intervalos

Una vez obtenida la representación multi-escala de la señal, se plantea el problema de determinar las escalas naturales. La solución ofrecida por Witkin (1984) se basa en las propiedades comentadas en la subsección 2.2.3.1, debido a estas propiedades las curvas de nivel comienzan y terminan en el nivel inferior de la escala por lo que están unas contenidas en otras como se aprecia en la figura 4.2, las curvas que llegan al nivel superior representado

se cerrarían en algún nivel superior, por esto definirán una estructura en forma de árbol. Dicho de otro modo características observadas como separadas a un determinado nivel de escala pueden aparecer mezcladas en una sola característica en un nivel superior de la escala.

El procedimiento consiste asociar a cada curva de nivel un rectángulo, cuya posición en el eje horizontal viene determinada por el intervalo definido por los pies de la curva, la posición superior en el eje vertical viene determinada por el mayor nivel de escala alcanzado por la curva y la posición inferior por el mayor nivel alcanzado por las curvas contenidas dentro de dicha curva. Para las curvas que llegan al nivel superior, y por lo tanto no se sabe cuál de las otras curvas que llegan al nivel superior es su continuación, se le asocia el rectángulo con posición horizontal determinada por el intervalo definido por el pie de dicha curva y el pie de la siguiente que llega al nivel superior, la posición vertical viene determinada por el mayor nivel alcanzado por las curvas cuyos pies están contenidos dentro del intervalo definido por el rectángulo. De esta manera se tiene un árbol ternario que particiona el espacio de escalas en rectángulos, comenzando por los originados por las curvas que alcanzan los niveles superiores y descendiendo hasta el nivel inferior. Dado un criterio de estabilidad para los rectángulos, por ejemplo la altura, se recorre el árbol buscando los rectángulos que maximizan la estabilidad a lo largo de la rama del árbol a la que pertenecen, ver figura 4.9.

La presencia de ruido puede hacer necesario un procedimiento de mezcla de características que disminuya la fragmentación Rosin (1994).

2.3. Extracción de líneas

Como se comentó en la sección 2.1 los métodos de extracción de líneas van estimando la ecuación del segmento a medida que incorporan las mediciones, por lo tanto, aparte de disponer de un método de estimación de los parámetros, deben disponer de un sistema para la detección de los extremos de los segmentos. La cuestión de la estimación de los parámetros será estudiada en la sección 2.4. Muchos de estos algoritmos usan la distancia un punto al segmento para decidir si se le considera perteneciente al segmento, esta distancia puede ser la euclídea o la de Mahalanobis Mahalanobis (1936), ver subsección 5.4.3.1. En esta sección se comentarán los métodos más habituales para la extracción de líneas haciendo hincapié en la cuestión de la detección de extremos.

Los nombre se estos algoritmos suelen usarse en inglés y en muchos casos intentar una traducción podría llevar a confusión al lector, por lo que se usarán sus nombres en inglés cuando proceda.

2.3.1. Split and Merge

Split and Merge (SM) Pavlidis and Horowitz (1974), dividir y mezclar, es quizás el algoritmo de seguimiento de líneas más popular, su sencillez y eficiencia y buenos resultados le avalan Nguyen et al. (2007). Ha sido usado en robótica por Castellanos and Tardós

(1996); Siadat et al. (1997); Baltzakis and Trahanias (2002); Borges and Aldon (2004) entre otros. SM se compone, como su nombre indica de dos fases. En la fase de división se comienza por estimar un segmento con todos los puntos disponibles y se calcula qué punto queda a mayor distancia de él, si esta distancia supera un umbral se divide el segmento en dos partes por dicho punto. Este proceso se continúa hasta que no es posible dividir ningún segmento. En la fase de mezcla de segmentos colineales, el criterio para considerar dos segmentos colineales suele estar basado en la distancia de Mahalanobis entre dos distribuciones, se mezclan los segmentos adyacentes que cumplan el criterio de colinealidad.

Si en lugar de estimar los parámetros de los segmentos por un procedimiento de ajuste, simplemente se calculan mediante la expresión de la línea que une los extremos, entonces el procedimiento se denomina Iterative End-Point Fit Duda and Hart (1973). Ha sido usado en robótica en Siadat et al. (1997); Borges and Aldon (2004); Tan et al. (2010) .

2.3.2. Line tracking

El Seguimiento de Líneas o algoritmo Incremental, comienza con los primeros dos puntos estima la recta se selecciona el punto siguiente y se añade a la recta, si se cumple cierto criterio se sigue con el punto siguiente, si no se comienza un nuevo segmento. El criterio suele ser un umbral superior para las varianzas de las estimaciones de los parámetros. Ha sido usado en Taylor and Probert (1996); Vandorpe et al. (1996).

2.3.3. Random Consensus Sample (RANSAC)

El Consenso por Muestreo Aleatorio Fischler and Bolles (1981) (RANSAC), elige aleatoriamente dos puntos y calcula la línea que pasa por ellos. A continuación calcula las distancias de los demás puntos de subconjunto a la recta y los separa en dos clases, próximos a al segmento (inliers) o lejanos (outliers) usando un umbral. Si se encuentran suficientes inliers, otro umbral, se recalcula el segmento y se guarda. Este proceso se repite hasta que se alcance un número máximo de iteraciones o queden pocos puntos. Algoritmo básico se han propuesto modificaciones como en Fontanelli et al. (2007).

2.3.4. Transformada de Hough

La transformada de Hough, Duda and Hart (1972), es un algoritmo muy empleado en Visión Artificial. La transformada usa una malla, el acumulador, que discretiza el conjunto de valores para el radio y ángulo de las mediciones. Una celda del acumulador se incrementa si una determinada medición verifica la ecuación de la recta para los parámetros de la malla. Las celdas con mayor número de votos en el acumulador definen los valores de los parámetros de los segmentos. La transformada de Hough o alguna de sus variantes se ha empleado para extracción de líneas en medidas de rango en Norouzi et al. (2009); Alempijevic and Dissanayake (2004); Forsberg et al. (1995); Fernández et al. (2010).

2.3.5. Expectation Maximization

La maximización de expectativas (EM), Dempster et al. (1977), usa un procedimiento iterativo con el que mejora un modelo previo. Tiene dos fase, en la primera se calculan las expectativas de las mediciones con respecto al modelo, a continuación se buscan los valores de los parámetros que maximizan las expectativas, bajo restricciones impuestas por las expectativas previamente calculadas. Ha sido empleado en robótica por Pfister et al. (2003).

2.4. Procedimientos de estimación

Las coordenadas polares resultan la manera más natural de expresar las medidas radiales tomadas por un láser. Aunque existe una transformación geométrica entre las coordenadas coordenadas polares y las cartesianas de manera que resultan equivalentes, en términos estadísticos no se verifica dicha equivalencia. Las variables ángulo del rayo φ y distancia medida r son claramente independientes, pero sus contrapartidas cartesianas no lo son ya que están relacionadas por el teorema de Pitágoras $x^2 + y^2 = r^2$. Además al ser función de r , son aleatorias, por lo que la variable independiente x incumple la condición de ser determinística, sección 2.4.1. En estas condiciones los estimadores obtenidos son sesgados y las distribuciones estadísticas asociadas asintóticas.

La condición, para x de ser determinística se puede relajar si los intervalos entre sus valores debieran ser constantes, aunque pueden estar alterados aleatoriamente, como ocurre cuando provienen de un aparato de medición que, teóricamente, deba dar medidas equiespaciadas. Este es el caso del ángulo en las medidas de rango proporcionadas por un láser. Luego las coordenadas polares constituyen un marco adecuado para abordar el problema de estimación de modelos de regresión con medidas proporcionadas por los dispositivos comentados.

Dado que en el capítulo 3 se hará uso intensivo de la estimación por mínimos cuadrados clásica, se expondrán aquí las bases teóricas necesarias, esta exposición se pueden encontrar en libros de texto como por ejemplo Johnston and DiNardo's (1996). A continuación se expondrán otros métodos de estimación usados en el problema que nos ocupa.

2.4.1. El método de regresión clásico

Un modelo de regresión lineal relaciona una variable dependiente y con un conjunto de k variables X mediante la relación

$$y = X\beta + \varepsilon \tag{2.3}$$

Donde β es un conjunto de k parámetros y ε un término de perturbación aleatoria normal (gaussiana). Esto hace que la y sea también aleatoria normal, pero por el momento de pedirá que las variables contenidas en X sean determinísticas (no aleatorias).

El objetivo es estimar el vector de parámetros a partir de un conjunto de n observaciones de y y X para que hagan mínima la suma de los cuadrados de los errores (suma de cuadrados residual SCR) cometidos en la estimación, precisando, la expresión vectorial de la ecuación (2.3) con los tamaños matriciales¹ queda:

$$y_{n \times 1} = X_{n \times k} \beta_{k \times 1} + \varepsilon_{n \times 1} \quad (2.4)$$

en este caso tendremos que $\varepsilon \sim N(0_{n \times 1}, Q_{n \times n})$ con matriz de varianzas y covarianzas, Q , definida positiva y pediremos que $\text{rang}(X) = k$, así $X'X$ será invertible.

2.4.1.1. Mínimos cuadrados ordinarios (MCO)

Si denominamos $\hat{\beta}$ al estimador de β , obtendremos la expresión de los estimadores de los valores de y , \hat{y} , y a partir de ella la de los errores de estimación e .

$$\begin{aligned} \hat{y} &= X \hat{\beta} \\ e &= y - \hat{y} \\ &= y - X \hat{\beta} \end{aligned} \quad (2.5)$$

La expresión de SCR queda

$$SCR = \sum_{i=1}^n e_i^2 = e'e = y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$$

derivando con respecto a $\hat{\beta}$ e igualando a 0 se obtiene la expresión del estimador mínimo cuadrático ordinario (MCO):

$$\hat{\beta}_{MCO} = (X'X)^{-1}X'y \quad (2.6)$$

y por lo tanto queda

$$SCR = y'y - \beta'X'y$$

Bajo las hipótesis del teorema de Gauss-Markov, perturbaciones normales, de media cero, homocedásticas e incorreladas, o sea que $Q = \sigma^2 I_n$, donde σ^2 es la varianza constante de las perturbaciones e I_n la matriz identidad de orden n , los estimadores MCO son lineales, insesgados y de mínima varianza. La expresión de su varianza es

$$\text{var}(\hat{\beta}_{MCO}) = \Sigma_{MCO} = \sigma^2 (X'X)^{-1} \quad (2.7)$$

La estimación insesgada $\hat{\sigma}^2$ de la varianza de la perturbación es,

$$\hat{\sigma}^2 = \frac{e'e}{n - k} \quad (2.8)$$

¹Los vectores siempre se consideran dispuestos en columna, para disponerlos en fila se trasponen.

y la distribución de probabilidad asociada a ella es,

$$\frac{e'e}{\sigma^2} = \frac{(n-k)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-k) \quad (2.9)$$

La estimación insesgada de la varianza de los estimadores es por lo tanto:

$$\text{var}(\hat{\beta}_{MCO}) = \hat{\Sigma}_{MCO} = \hat{\sigma}^2(X'X)^{-1} \quad (2.10)$$

Con estos resultados se puede construir un test sobre las predicciones realizadas por el modelo comparandolas con los valores realmente observados. Así si X_{n+1} representa el vector con las nuevas observaciones de las variables X e y_{n+1} la nueva observación de y , el modelo que representa la relación entre estas observaciones es:

$$y_{n+1} = X'_{n+1}\beta + \varepsilon_{n+1}$$

donde $\varepsilon_{n+1} \sim N(0, \sigma^2)$ y está incorrelada con ε .

En estas condiciones el error cometido en la predicción de Y_{n+1} , es

$$e_{n+1} = y_{n+1} - X'_{n+1}\hat{\beta}$$

y la distribución de probabilidad asociada es

$$\frac{e_{n+1}}{\hat{\sigma}\sqrt{1 + X'_{n+1}(X'X)^{-1}X_{n+1}}} \sim t(n-2) \quad (2.11)$$

2.4.1.2. Mínimos cuadrados generalizados (MCG) y heterocedasticidad

Estudiaremos ahora el caso en que la varianza de la perturbación no es homocedástica, aunque mantendremos la hipótesis de normalidad, ahora es $\varepsilon \sim N(0, \sigma^2 Q)$ con Q simétrica, definida positiva y conocida. Para Q es posible encontrar una matriz P invertible tal que $Q = PP'$, entonces multiplicando la ecuación (2.4) por P^{-1} transformamos el modelo

$$P^{-1}y = P^{-1}X\beta + P^{-1}\varepsilon \quad (2.12)$$

que podemos escribir

$$y^* = X^*\beta + \varepsilon^*$$

donde $\varepsilon^* \sim N(0, \sigma^2 I_n)$, por lo que los estimadores obtenidos aplicando MCO al modelo transformado tiene las propiedades descritas en el teorema de Gauss-Markov, subsección 2.4.1.1, en este caso se conoce como teorema de Aitken. Dicho estimador se denomina estimador mínimo cuadrático generalizado (MCG):

$$\hat{\beta}_{MCG} = (X^{*'}X^*)^{-1}X^{*'}y^* = (X'Q^{-1}X)^{-1}X'Q^{-1}y$$

con varianza

$$\text{var}(\hat{\beta}_{MCG}) = \sigma^2(X'Q^{-1}X)^{-1} \quad (2.13)$$

Los errores transformados $e^* = P^{-1}e$ cumplen propiedades análogas a las comentadas en la subsección 2.4.1.1.

En el primer párrafo de esta sección se pedía que Q fuera conocida, este hecho no se suele dar en las aplicaciones reales, por lo que es necesario estimar Q . Dado la cantidad de elementos distintos que puede tener, $(n+n^2)/2$, hay que realizar hipótesis restrictivas sobre la forma de dicha matriz, en estos casos de habla de estimación MCG factible. En este papel nos vamos a ocupar del caso en que las perturbaciones no tienen varianza constante pero están incorreladas, heterocedasticidad, con lo que Q y P^{-1} tienen forma diagonal:

$$Q = \begin{pmatrix} \frac{1}{w_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{w_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{w_n^2} \end{pmatrix} \quad P^{-1} = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{pmatrix}$$

El procedimiento de estimación por MCG se suele denominar en este caso estimación por mínimos cuadrados ponderados (MCP). Las varianzas de las perturbaciones son $\sigma_i^2 = \sigma^2/w_i^2$. Los pesos w_i pueden depender de las variables X , entonces serían calculables a priori. También pueden depender de la variable y , con lo que no serían calculables a priori, otras formas de dependencia más compleja pueden darse.

Para el caso en que las varianzas dependan de y , o más precisamente de su valor esperado $E[y]$, $\sigma_i^2 = \frac{\sigma^2}{w^2(y_i)}$ será $w_i = |w(y_i)|$, la técnica de estimación involucra dos etapas. Una primera estimación por MCO a partir de la cual se obtienen estimaciones de y , \hat{y}_i . Con estas estimaciones se estiman los pesos $w_i = |w(\hat{y}_i)|$, se transforma el modelo como en la ecuación (2.12). En este caso el producto por P^{-1} equivale a multiplicar cada observación de y y de las X por el peso correspondiente. Aplicando nuevamente MCO al modelo transformado obtenemos los estimadores MCG, este procedimiento se denomina estimación por MCG factibles.

2.4.1.3. Modelos no lineales

La técnica de estimación de modelos no lineales que se va a seguir en este papel está basada en el método de Gauss-Newton, consiste en realizar una aproximación de Taylor de primer orden del modelo y trabajar con esta aproximación como si se tratase de un modelo lineal. O sea, si tenemos un modelo

$$y = h(X; \beta) + \varepsilon \quad (2.14)$$

calculamos el jacobiano de f , $J = \frac{\partial h(X; \beta)}{\partial \beta}$ que ocupará el lugar de la matrix X en las ecuaciones (2.6), (2.7), (2.10) y (2.11). Por ejemplo de la ecuación (2.6) obtendríamos una fórmula para el estimador no lineal (MCONL)

$$\hat{\beta}_{MCONL} = (J'J)^{-1}J'y \quad (2.15)$$

El método de Gauss-Newton necesita arrancar de unos valores iniciales, como cualquier otro basado en aproximaciones, y suele iterarse varias veces hasta alcanzar la precisión deseada o superar un límite de iteraciones.

La validez de los resultados estadísticos comentados en la subsección 2.4.1.1 tiene carácter asintótico, o sea, disminuye con el tamaño de la muestra, Gallant (1975); Kimura (1990). El método en dos etapas descrito en la subsección 2.4.1.2 también se puede aplicar a los modelos no lineales como se describe en Beal and Sheiner (1988).

2.4.2. Mínimos cuadrados totales

El método de los Mínimos Cuadrados Totales (MCT) Markovsky and Huffel (2007); Tellinghuisen (2010), en su versión más sencilla los errores son definidos como las distancias de los puntos a la recta, en lugar de diferencias de variable dependiente, como hace la teoría clásica. MCT es capaz de contemplar el caso en el que tanto las variables dependientes como las independientes de un modelo de regresión puedan estar afectadas por errores. Por estas razones es conocido también como: regresión ortogonal, errores-en-variables y errores de medida. La teoría de los MCT ofrece un paralelismo con la de regresión clásica, ver subsecciones 2.4.1.1 y 2.4.1.2 al disponer de un modelo con varianza del ruido constante, como en MCO, y otros para varianzas no constante, ver Markovsky and Huffel (2007).

La formulación del caso lineal de los MCT, expuesta para el caso de una sola variable aleatoria para apreciar mejor los parecidos y diferencias con el caso clásico expuesto en la subsección 2.4.1, es la siguiente. En el método clásico, a partir de la ecuación 2.5, se puede escribir $X\hat{\beta}_{MCO} = y - e$ y entonces se busca $\hat{\beta}$ de manera que minimice la suma de los cuadrados de los errores. En MCT puede haber errores en las X , luego ahora se tendrá $(X - e_x)\hat{\beta}_{MCT} = y - e_y$ y habrá que minimizar los dos errores en el sentido de la norma de Frobenius. Otra forma de verlo es que se minimiza la siguiente suma de cuadrados ponderada $\frac{\sum e^2}{\|X\|^2+1}$.

La solución se puede obtener de dos maneras. Sea denomina $C = \begin{pmatrix} X_{n \times k} & y_{n \times 1} \end{pmatrix}$, se obtiene la descomposición en valores singulares de $C = U\Sigma V'$ con Σ diagonal. Se particionan Σ y V separando la parte correspondiente a X de la de y , o sea,

$$V = \begin{pmatrix} V_{XX} & V_{Xy} \\ V_{Xy} & V_{yy} \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{XX} & 0 \\ 0 & \Sigma_{yy} \end{pmatrix}$$

entonces el estimador MTC es $\hat{\beta}_{MCT} = -V_{Xy}V_{yy}^{-1}$ y la matriz de errores $\begin{pmatrix} e_x & e_y \end{pmatrix} = -U \text{diag} \begin{pmatrix} 0 & \Sigma_{yy} \end{pmatrix} V'$.

La segunda solución es $\hat{\beta}_{MCT} = (X'X - \sigma_y^2 I) X'y$, donde σ_y^2 es el único elemento de Σ_{yy} , compárese con la ecuación 2.6.

El estimador MCT es débilmente consistente.

Este método ha sido usado en diversas formas en el problema que nos ocupa Krystek and Anton (2007); Arras and Siegwart (1997); Núñez et al. (2008), en ellos se usa la

ecuación de la recta en forma polar de la siguiente manera $r \cos(\varphi - \theta) = \rho$, ρ y θ son los parámetros de la recta, r y φ las coordenadas polares de los puntos. Ciertas estimaciones serán realizadas en los experimentos de esta tesis mediante el procedimiento descrito en Arras and Siegwart (1997), pasamos a exponer brevemente este procedimiento.

Para determinar los estimadores se pueden emplear estas fórmulas:

$$\begin{aligned} \tan(2\hat{\theta}) &= \frac{-2 \sum w_n (y_n - \bar{y}_w)(x_n - \bar{x}_w)}{\sum w_n [(y_n - \bar{y}_w)^2 - (x_n - \bar{x}_w)^2]} \\ \hat{\rho} &= \bar{x}_w \cos \hat{\theta} + \bar{y}_w \sin \hat{\theta} \end{aligned} \quad (2.16)$$

siendo w_n pesos y \bar{x}_w e \bar{y}_w las medias ponderadas. También es posible determinar la matriz de varianzas y covarianzas de los estimadores, en la que aparece como factor de proporcionalidad la varianza de las medidas r , los pesos se eligen como los inversos de esas varianzas, por lo que éstas deben ser conocidas, determinadas experimentalmente o fijadas *ad hoc*.

Capítulo 3

Filtros de Kalman para la estimación de líneas

3.1. Introducción

En esta capítulo se mostrará cómo distintos modelos regresión pueden ser estimados mediante filtros de Kalman y cómo obtener un test de detección de valores anómalos (*outliers*). Los casos a estudio serán los necesarios para realizar una adecuada obtención de segmentos rectilíneos a partir de medidas de rango tomadas desde un punto central. Primero se describirá la estimación de modelos lineales homocedásticos (varianza de perturbación constante), en segundo lugar la de modelos lineales con perturbaciones heterocedásticas (varianza de perturbación no constante), para llegar a la estimación de segmentos rectilíneos en coordenadas polares, modelo no lineal, que razonaremos suele ser la forma más apropiada de representar mediciones de rango para esta clase de estimación. La formulación que se usará hace a este procedimiento independiente de la varianza de los errores de medida del láser, pudiendo detectar segmentos sobre paredes rectas muy bien acabadas o sobre un seto de un jardín, aunque claro está cuanto mayor sea la varianza menor será la fiabilidad y precisión de los resultados.

Para emplear estos procedimientos es necesario que el sensor de rango produzca un número suficiente de medidas como para poder realizar contrastes de hipótesis con cierta verosimilitud. Por este motivo nos vamos a centrar en mediciones realizadas con láser.

3.2. Estimación de modelos lineales mediante filtros de Kalman

La formulación del filtro de Kalman para la estimación recursiva de modelos lineales está tomada de Peña (1985)¹, que referiremos en la subsección 3.2.1, primero describiremos

¹Este artículo está publicado en castellano, sospecho que eso le ha hecho pasar desapercibido.

los filtros de Kalman adaptando la notación de Thrun et al. (2005), para fijar ideas. Consideremos un sistema dinámico cuyo estado, inobservable, en el instante t viene representado por la variable s_t , sobre este sistema se pueden ejercer acciones de control simbolizadas por la variable u_t y por último se toman medidas del entorno, y_t , que dependen del estado. Las ecuaciones que rigen las transiciones del estado y las mediciones del entorno son las siguientes:

$$s_t = A_t s_{t-1} + B_t u_t + \delta_t \quad (3.1)$$

$$y_t = C_t s_t + \varepsilon_t \quad (3.2)$$

donde A_t , B_t y C_t son matrices conocidas de tamaños $k \times k$, $k \times m$ y $p \times k$, respectivamente, siendo k , m y p los tamaños de los vectores s_t , u_t e y_t , respectivamente. Las perturbaciones aleatorias tiene las siguientes distribuciones de probabilidad, $\delta_t \sim N(0, R_t)$ y $\varepsilon_t \sim N(0, Q_t)$. Dado que el proceso es recursivo hará falta conocer el estado inicial s_0 .

El algoritmo del filtro de Kalman permite la estimación recursiva de los parámetros de la distribución del estado condicionada a la acción de control y a la medición actuales, bajo las condiciones de linealidad y normalidad del sistema esta distribución es $N(\mu_t, \Sigma_t)$. El algoritmo del filtro de Kalman se describe en el algoritmo 3.1.

Algoritmo 3.1 Filtro de Kalman. Las líneas 1 y 2 se denominan de actualización del estado tras la acción de control u_t , en 3 se calcula la ganancia de Kalman K_t que determina el grado en que la nueva medición se incorpora a la estimación del estado, en 4 y 5 se actualiza el estado incorporando la medición y_t , e_t se denomina la innovación y corresponde al error de predicción de la medida.

Filtro_de_Kalman(μ_{t-1} , Σ_{t-1} , u_t , y_t)

valores iniciales μ_0 y Σ_0

1: $\bar{\mu}_t = A_t \mu_{t-1} + B_t u_t$

2: $\bar{\Sigma}_t = A_t \Sigma_{t-1} A_t' + R_t$

3: $K_t = \bar{\Sigma}_t C_t' (C_t \bar{\Sigma}_t C_t' + Q_t)^{-1}$

4: $\mu_t = \bar{\mu}_t + K_t \underbrace{\left(y_t - C_t \bar{\mu}_t \right)}_{e_t}$

5: $\Sigma_t = (I - K_t C_t) \bar{\Sigma}_t$

6: **return** μ_t , Σ_t

3.2.1. Formulación del modelo de regresión lineal como sistema dinámico

La formulación dada en Peña (1985) para el modelo lineal es la siguiente:

- Los estado los constituyen el conjunto de parámetros, $s_t = \beta \forall t$.

3.2. ESTIMACIÓN DE MODELOS LINEALES MEDIANTE FILTROS DE KALMAN 27

- No hay acciones de control, $u_t = 0$ y $B_t = 0 \forall t$.
- La ecuación de la medición representa al modelo lineal para la observación t -ésima de las variable dependiente y_t y de las independientes X_t .

lo que expresado ecuacionalmente queda,

$$\beta_t = \beta_{t-1} \quad (3.3)$$

$$y_t = X_t' \beta_t + \varepsilon_t \quad (3.4)$$

por lo tanto k vuelve a ser el número de parámetros del modelo, $m = 0$, $p = 1$, $A_t = I$, $\delta_t = 0$, $R_t = 0$, $C_t = X_t'$ y Q_t se reduce a un escalar. Bajo estas condiciones la distribución del estado condicionada por la medición es $N(\beta, \Sigma)$, o sea, el filtro de Kalman calcula los estimadores de los parámetros del modelo y estima la matriz de varianzas de dicho estimadores, los que, por ejemplo, en la subsección 2.4.1.1 se denominaron $\hat{\beta}_{MCO}$ y $\hat{\Sigma}_{MCO}$, ver ecuaciones (2.6) y (2.10). En lo que sigue se omitirá el acento circunflejo sobre el estimador $\hat{\beta}$, por no sobrecargar la notación, quedará claro por el contexto los pocos casos en los que β representará a los parámetros en lugar de a sus estimaciones.

3.2.2. Estimación MCO mediante filtro de Kalman

Cambiamos el subíndice t , usado en la exposición de los filtros de Kalman, por el subíndice n usado en la exposición de los modelos de regresión, principalmente para no confundirlo con la distribución t de Student.

En el trabajo citado se desarrolla la estimación por MCO, o sea suponiendo que $Q_n = \sigma^2$, obteniendo los resultados estadísticos comentados en la subsección 2.4.1.1. La presencia del factor de escala desconocido σ^2 se evita redefiniendo Σ_n como $\frac{\Sigma_n}{\sigma^2}$, a partir de ahora simplemente Σ_n , de esta manera las ecuaciones del algoritmo se hacen independientes de σ^2 , queda entonces $Q_n = 1$. Sustituyendo lo comentado anteriormente en el algoritmo 3.1 queda el algoritmo 3.2 para la estimación recursiva de modelos lineales con perturbaciones homocedásticas.

Una cuestión importante para las estimación de la varianza de la perturbación es el cálculo de la SCR tras la incorporación de una nueva observación, reescribimos la ecuación (2.8),

$$\hat{\sigma}_n^2 = \frac{SCR_n}{n - k} = \frac{\sum_{i=1}^n \overbrace{(y_i - X_i' \beta_n)}^{e_{i/n}}}{n - k} \quad (3.5)$$

el uso de esta fórmula tal cual presenta una cuestión de eficiencia, ya que respecto al número de observaciones, n , la complejidad del algoritmo, si emplea la fórmula, es $O(n^2)$. En Peña (1985) no aclara esta cuestión, pero acudiendo al artículo seminal Plackett (1950), en su

Algoritmo 3.2 Regresión lineal por MCO mediante filtro de Kalman. Las líneas 1 y 2 son triviales y sus resultados han sido sustituidos en las siguientes, en adelante serán suprimidas. La inversión en 3 sólo afecta al escalar a_n . Puede verse la similitud entre dicho escalar y el que aparece bajo el signo radical en la ecuación (2.11), teniendo en cuenta la (2.7). En la línea 6 se realiza el cálculo recursivo de la SCR.

MCO_de_Kalman($\beta_{n-1}, \Sigma_{n-1}, y_n, SCR_{n-1}$)

valores iniciales β_0 y Σ_0

$SCR_0 = 0$

1: $\bar{\beta}_n = \beta_{n-1}$

2: $\bar{\Sigma}_n = \Sigma_{n-1}$

3: $K_n = \Sigma_{n-1} X_n \underbrace{\left(1 + X_n' \Sigma_{n-1} X_n\right)}_{a_n}^{-1}$

4: $\beta_n = \beta_{n-1} + K_n \underbrace{\left(y_n - X_n' \beta_{n-1}\right)}_{e_{n/n-1}}$

5: $\Sigma_n = (I - K_n X_n') \Sigma_{n-1}$

6: $SCR_n = SCR_{n-1} + \frac{e_n^2}{a_n}$

7: **return** β_n, Σ_n, SCR_n

punto 10, se proporciona una fórmula para la estimación recursiva de la SCR, por lo que la complejidad queda $O(n)$:

$$SCR_n = SCR_{n-1} + \frac{e_n^2}{\underbrace{1 + X_n' \Sigma_{n-1} X_n}_{a_n}} \quad (3.6)$$

con $e_n = y_n - X_n' \beta_n$ y $SCR_0 = 0$.

De la distribución presentada en la ecuación (2.11) se obtiene un test de detección de outliers que es igualmente válido cuando se realiza la estimación mediante filtro de Kalman, el estadístico de contraste para decidir si la observación n -ésima es un outlier queda:

$$t_{exp,n} = \frac{e_{n/n-1}}{\hat{\sigma}_{n-1} \sqrt{\underbrace{1 + X_n' \Sigma_{n-1} X_n}_{a_n}}} \sim t(n-1-k) \quad (3.7)$$

con $e_{n/n-1} = y_n - X_n' \beta_{n-1}$. Para un nivel de significación α , la región crítica del test para considerar una nueva observación n , a partir de la observación $k+2$, como outlier es $|t_{exp,n}| > t_{n-1-k, \frac{\alpha}{2}}$, siendo $t_{n-1-k, \frac{\alpha}{2}}$ el valor de la distribución $t(n-1-k)$ que deja por encima de sí una probabilidad $\frac{\alpha}{2}$.

Obsérvese que con la redefinición de Σ_n se hicieron las ecuaciones del filtro independientes de σ^2 , pero nos es posible obtener una estimación de dicho parámetro y por lo tanto de la verdadera matriz de varianzas de los estimadores $\hat{\sigma}_n^2 \Sigma_n$.

Al evaluar la complejidad temporal del algoritmo tendremos que tener en cuenta que al aplicarlo a segmentación, el número, k , de parámetros β será pequeño y el de observaciones, n , grande. Con respecto a k la operación que involucra más cálculos es el producto de matrices $k \times k$ en la línea 5 del algoritmo 3.2, cuya complejidad naive es $O(k^3)$. Con respecto a n ya se comentó que la complejidad es $O(n)$.

3.2.3. Estimación MCP mediante filtro de Kalman

Para la estimación MCP, cuando los pesos dependen de la variable y o de su valor esperado $E[y]$, mediante filtro de Kalman, proponemos realizar un procedimiento similar al descrito en la subsección 2.4.1.2, ver algoritmo 3.3. En lo que sigue los superíndices o y p , indican MCO, MCP, respectivamente.

En una primera etapa se realiza la estimación por MCO descrita en la subsección 3.2.2 de la que se obtienen β_n^o y Σ_n^o . El peso, w_n , se obtiene a partir de la estimación MCO de y , \hat{y}_n^o , o sea, $w_i = |w(\hat{y}_i^o)|$, donde w es una función conocida salvo por un factor de proporcionalidad. En la segunda etapa, MCP, hay que tener en cuenta que la matriz de varianzas del término de perturbación ε es ahora $Q_n = 1/w_n^2$. El haber formulado el modelo de regresión con un filtro de Kalman nos evita transformar los datos, ver sección 3.6.1.1, basta con sustituir la nueva expresión de Q_n en las ecuaciones del filtro de Kalman y proceder como en la primera etapa para obtener los estimadores MCP, β_n^p , y la estimación de su matriz de varianzas, salvo el factor σ^2 , Σ_n^p .

Al igual que en la subsección 3.2.2 las SCR puede ser calculada secuencialmente mediante una fórmula análoga a la ecuación (3.6), ver sección 3.6.1.2,

$$SCR_n^p = SCR_{n-1}^p + \frac{(e_n^p)^2}{\underbrace{\frac{1}{w_n^2} + X_n' \Sigma_{n-1}^p X_n}_{a_n^p}} \quad (3.8)$$

análogamente a la subsección 2.4.1.2 tendremos que el estimador de la varianza de la perturbaciones es

$$(\hat{\sigma}_n^p)^2 = \frac{SCR_n^p}{n - k} \quad (3.9)$$

En la subsección 2.4.1.2 se comentó que los errores transformados seguían las distribuciones de probabilidad comentadas en 2.4.1.1, por lo que ahora el estadístico de contraste para decidir si la observación n -ésima es un outlier queda

$$t_{exp,n} = \frac{e_{n/n-1}^p}{\hat{\sigma}_{n-1}^p \sqrt{\underbrace{\frac{1}{w_n^2} + X_n' \Sigma_{n-1}^p X_n}_{a_n^p}}} \sim t(n - 1 - k) \quad (3.10)$$

con $e_{n/n-1}^p = y_n - X_n' \beta_{n-1}^p$. Para un nivel de significación α , la región crítica del test para considerar una nueva observación n , a partir de la observación $k + 2$, como outlier es

Algoritmo 3.3 Regresión lineal MCP mediante filtro de Kalman. En las tres primeras líneas se realiza la estimación MCO descrita en el algoritmo 3.2. En la cuarta se calcula el peso. Y en las tres últimas se realiza un proceso similar al realizado en las tres primeras, la salvedad más importante es el inverso de w_n^2 en a_n^p en lugar del 1 que aparece en a_n^o . En la línea 4 se ha introducido el factor de escalamiento de pesos ω .

MCP_de_Kalman($\beta_{n-1}^o, \Sigma_{n-1}^o, \beta_{n-1}^p, \Sigma_{n-1}^p, y_n, SCR_{n-1}^p$)

valores iniciales β_0^o y Σ_0^o

β_0^p y Σ_0^p se inicializan con β_1^o y Σ_1^o , respectivamente

$SCR_0^p = 0$

- 1: $K_n^o = \Sigma_{n-1}^o X_n \left(\underbrace{1 + X_n' \Sigma_{n-1}^o X_n}_{a_n^o} \right)^{-1}$
 - 2: $\beta_n^o = \beta_{n-1}^o + K_n^o \left(\underbrace{y_n - X_n' \beta_{n-1}^o}_{e_n^o} \right)$
 - 3: $\Sigma_n^o = (I - K_n^o X_n') \Sigma_{n-1}^o$
 - 4: $w_n = \omega |w(\hat{y}_n^o)| = \omega |w(X_n' \beta_{n-1}^o)|$
 - 5: $K_n^p = \Sigma_{n-1}^p X_n \left(\underbrace{\frac{1}{w_n^2} + X_n' \Sigma_{n-1}^p X_n}_{a_n^p} \right)^{-1}$
 - 6: $\beta_n^p = \beta_{n-1}^p + K_n^p \left(\underbrace{y_n - X_n' \beta_{n-1}^p}_{e_{n/n-1}^p} \right)$
 - 7: $\Sigma_n^p = (I - K_n^p X_n') \Sigma_{n-1}^p$
 - 8: $SCR_n^p = SCR_{n-1}^p + \frac{(e_n^p)^2}{\omega^2 a_n^p}$
 - 9: **return** $\beta_n^o, \Sigma_n^o, \beta_n^p, \Sigma_n^p, SCR_n^p$
-

$|t_{exp}| > t_{n-1-k, \frac{\alpha}{2}}$, siendo $t_{n-1-k, \frac{\alpha}{2}}$ el valor de la distribución $t(n-1-k)$ que deja por encima de sí una probabilidad $\frac{\alpha}{2}$.

La estimación de la verdadera matriz de varianzas de los estimadores MCP será $(\hat{\sigma}_n^p)^2 \Sigma_n^p$. La complejidad es la misma que para el caso MCO, subsección 3.2.2.

3.2.3.1. Normalización de la ganancia del filtro

En los primeros experimentos se comprobó que si los pesos asociados a las primeras observaciones eran apreciablemente menores que la unidad, el filtro podía ser incapaz de corregir sesgos iniciales debido a que la ganancia de Kalman, K_n^p , se hacía pequeña, por lo que el filtro tenía poco en cuenta la innovación aportada por nuevas observaciones, que podría corregir el problema, ver subsección 3.4.3 para una discusión más detallada. La solución encontrada fue escalar los pesos de manera que sean próximos a 1 para los primeros datos añadidos al filtro, así los pesos se calculan como $w_i = \omega |w(\hat{y}_i^p)|$. Valores de w_i , al inicio de la estimación, apreciablemente superiores a la unidad hacen al filtro un tanto inestable, más cuanto mayor sea los w_i , aumentando la producción de falsos positivos en la detección de outliers. Estas características hace de ω , que denominaremos factor de normalización de ganancia o de escala de pesos, un parámetro crítico para el correcto funcionamiento del filtro, por ello se puede ver como un parámetro de sintonización (*tuning*) del filtro y debe ser siempre $\omega > 0$.

Este no cambia en nada las expresiones obtenidas anteriormente, ver subsección 3.6.1.3, salvo que la verdadera estimación de σ^2 que queda como $\frac{(\hat{\sigma}_n^p)^2}{\omega^2}$ y este cambio también se refleja en estimación de la verdadera matriz de varianzas de los estimadores MCP, ahora $\frac{(\hat{\sigma}_n^p)^2}{\omega^2} \Sigma_n^p$.

3.3. EKF para la estimación de líneas rectas en coordenadas polares

Las coordenadas polares resultan la manera más natural de expresar las medidas radiales tomadas por un láser. Aunque existe una transformación geométrica entre ambos sistemas de coordenadas de manera que resultan equivalentes, en términos estadísticos no se verifica dicha equivalencia. Las variables ángulo del rayo φ y distancia medida r son claramente independientes, pero sus contrapartidas cartesianas no lo son ya que están relacionadas por el teorema de Pitágoras $x^2 + y^2 = r^2$. Además al ser función de r , son aleatorias, por lo que x incumple la condición de ser determinística, sección 2.4.1. En estas condiciones los estimadores obtenidos son sesgados y las distribuciones estadísticas asociadas asintóticas.

La condición, para x de ser determinística se puede relajar si los intervalos entre sus valores debieran ser constantes, aunque pueden estar alterados aleatoriamente, como ocurre cuando provienen de un aparato de medición que, teóricamente, deba dar medidas

equiespaciadas. Este es el caso del ángulo en las medidas de rango proporcionadas por un láser.

Luego las coordenadas polares constituyen un marco adecuado para abordar el problema de estimación de modelos de regresión con medidas proporcionadas por los dispositivos comentados. Sin embargo presenta el problema de la no linealidad, por lo que debe ser abordado con extensiones del filtro de Kalman para sistemas no lineales, en este caso se usará el filtro de Kalman extendido (EKF).

3.3.1. El EKF

El EKF usa la técnica comentada en 2.4.1.3, para linealizar las funciones g y h que aparecen en las ecuaciones del sistema,

$$s_t = g(u_t, s_{t-1}) + \delta_t \tag{3.11}$$

$$y_t = h(s_t) + \varepsilon_t \tag{3.12}$$

usando sus gradientes $G_t = \left(\frac{\partial g(u_t, s_{t-1})}{\partial s_{t-1}} \right)_{\mu_{t-1}}$ y $H_t = \left(\frac{\partial h(s_t)}{\partial s_t} \right)_{\mu_{t-1}}$. El procedimiento del EKF se muestra en el algoritmo 3.4.

Algoritmo 3.4 Filtro extendido de Kalman EKF, para sistemas no lineales. Es similar al filtro básico, algoritmo 3.1, aunque usa los gradientes G_t y H_t en lugar de A_t y C_t respectivamente, y usa las funciones g y h al predecir las distintas medias del estado.

EKF($\mu_{t-1}, \Sigma_{t-1}, u_t, y_t$)

valores iniciales μ_0 y Σ_0

- 1: $\bar{\mu}_t = g(\mu_{t-1}, u_t)$
 - 2: $\bar{\Sigma}_t = G_t \Sigma_{t-1} G_t' + R_t$
 - 3: $K_t = \bar{\Sigma}_t H_t' (H_t \bar{\Sigma}_t H_t' + Q_t)^{-1}$
 - 4: $\mu_t = \bar{\mu}_t + K_t \underbrace{\left(y_t - h(\bar{\mu}_t) \right)}_{\varepsilon_t}$
 - 5: $\Sigma_t = (I - K_t H_t) \bar{\Sigma}_t$
 - 6: **return** μ_t, Σ_t
-

3.3.2. Estimación de líneas

En coordenadas polares, (r, φ) , la ecuación de una recta que no pasa por el polo (el origen) tiene la siguiente expresión:

$$r = h(\varphi) = \frac{\rho}{\cos(\varphi - \theta)} \quad -\frac{\pi}{2} < \varphi - \theta < \frac{\pi}{2} \tag{3.13}$$

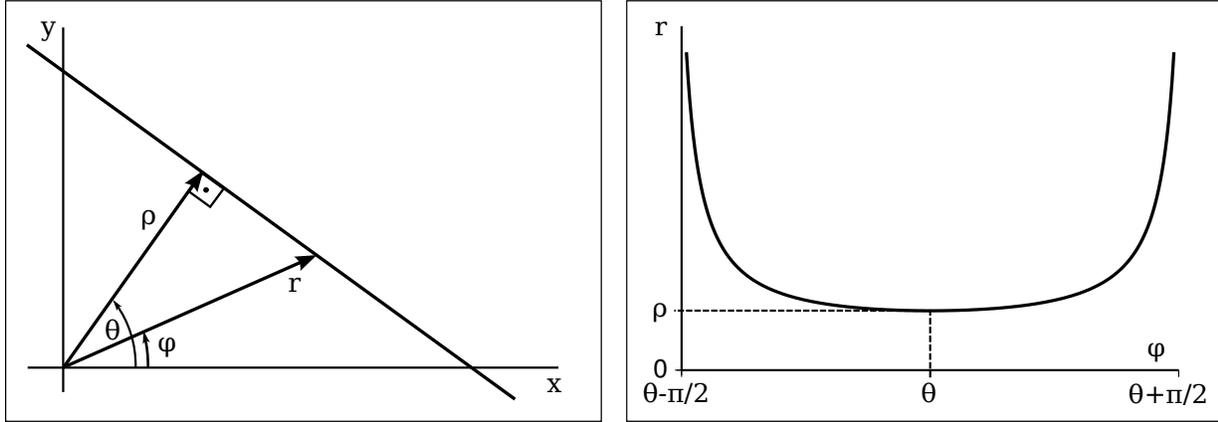


Figura 3.1: Representaciones de la recta en coordenadas rectangulares a) y en coordenadas polares b). Todas las rectas, que no pasan por el polo, tienen en coordenadas polares exactamente la misma forma, solo varía la posición del mínimo.

donde ρ es la distancia de la recta al origen y θ el ángulo que forma una perpendicular a la recta con el eje de abscisas, ver figura 3.1, que constituyen los parámetros a estimar.

Formulando el modelo como un sistema dinámico, denominaremos $\beta = \begin{pmatrix} \rho \\ \theta \end{pmatrix}$ al vector de parámetros del modelo, y usando de nuevo el subíndice n tendremos:

$$\beta_n = \beta_{n-1} \quad (3.14)$$

$$r_n = \frac{\rho_n}{\cos(\varphi_n - \theta_n)} + \varepsilon_n \quad (3.15)$$

luego la función g es $g(\rho, \theta) = I_2 \begin{pmatrix} \rho \\ \theta \end{pmatrix}$ y por lo tanto $G_n = I_2$ y

$$H_n = \begin{pmatrix} \frac{\partial r}{\partial \rho} \\ \frac{\partial r}{\partial \theta} \end{pmatrix}_{(\rho_{n-1}, \theta_{n-1})} = \begin{pmatrix} \frac{1}{\cos(\varphi_n - \theta_{n-1})} \\ -\frac{\rho_{n-1} \sin(\varphi_n - \theta_{n-1})}{\cos^2(\varphi_n - \theta_{n-1})} \end{pmatrix}$$

este gradiente puede ser calculado en el punto formado por los estimadores MCO de la etapa anterior o en el formado por los MCP, los distinguiremos con los superíndices o y p .

En el caso de perturbaciones homocedásticas bastará sustituir X_n por H_n , salvo donde se estimen valores de r , en el algoritmo 3.2 para tener el algoritmo EKF para estimación MCO no lineal. En el caso de perturbaciones heterocedásticas con varianza dependiente de r , $\sigma_i^2 = \frac{\sigma^2}{w^2(r_i)}$, podremos tratarlo en dos etapas como el caso de MCP lineal, subsección 3.2.3, haciendo los pesos $w_i = |w(r_i)|$, ver algoritmo 3.5. Las distribuciones de probabilidad y contrastes del caso lineal, subsecciones 3.2.2 y 3.2.3, tienen ahora carácter asintótico, como se comentó en la subsección 2.4.1.3.

Para el correcto funcionamiento de este procedimiento es necesario aplicar el factor escalamiento de pesos, ω , comentado en la subsección 3.2.3.

Algoritmo 3.5 Regresión no lineal MCP mediante filtro de Kalman, para la estimación de una recta en coordenadas polares. Es análogo al algoritmo 3.3 salvo que en algunas líneas se usa el gradiente de la función de medida, H_n^o o H_n^p , en el lugar que ocupaba X_t , y que cuando se estima un valor de r se emplea la propia función de medida h . En la línea 4 se ha introducido el factor de escalamiento de pesos ω .

EKF_polar($\beta_{n-1}, \Sigma_{n-1}, r_n, SCR_{n-1}^p$),
 valores iniciales β_0 y Σ_0
 β_0^p y Σ_0^p se inicializan con β_1^o y Σ_1^o , respectivamente
 $SCR_0^p = 0$

- 1: $K_n^o = \Sigma_{n-1}^o H_n^o \left(\underbrace{1 + H_n^{o'} \Sigma_{n-1}^o H_n^o}_{a_n^o} \right)^{-1}$
 - 2: $\beta_n^o = \beta_{n-1}^o + K_n^o \left(\underbrace{r_n - h(\varphi_n; \beta_{n-1}^o)}_{e_n^o} \right)$
 - 3: $\Sigma_n^o = (I - K_n^o H_n^{o'}) \Sigma_{n-1}^o$
 - 4: $w_n = \omega |w(\hat{r}_n^o)| = \omega |w(h(\varphi_n; \beta_{n-1}^o))|$
 - 5: $K_n^p = \Sigma_{n-1}^p H_n^p \left(\underbrace{\frac{1}{w_n^2} + H_n^{p'} \Sigma_{n-1}^p H_n^p}_{a_n^p} \right)^{-1}$
 - 6: $\beta_n^p = \beta_{n-1}^p + K_n^p \left(\underbrace{r_n - h(\varphi_n; \beta_{n-1}^p)}_{e_n^p} \right)$
 - 7: $\Sigma_n^p = (I - K_n^p H_n^{p'}) \Sigma_{n-1}^p$
 - 8: $SCR_n^p = SCR_{n-1}^p + \frac{(e_n^p)^2}{a_n^p}$
 - 9: **return** $\beta_n^o, \Sigma_n^o, \beta_n^p, \Sigma_n^p, SCR_n^p$
-

Una cuestión a tener en cuenta al evaluar los jacobianos es la determinación del ángulo $\varphi_n - \theta_{n-1}$, ya que de acuerdo con la ecuación (3.13) debe ser $-\frac{\pi}{2} < \varphi_n - \theta_{n-1} < \frac{\pi}{2}$. Si esto no se tiene en cuenta pueden aparecer cambios de signos en los jacobianos debidos a diferentes determinaciones de $\varphi_n - \theta_{n-1}$, por lo que puede ser necesario cambiar la determinación de $\varphi_n - \theta_{n-1}$, sólo en el cálculo de jacobianos, de la siguiente forma:

$$\varphi_n - \theta_{n-1} = \begin{cases} \varphi_n - \theta_{n-1} + \pi & \text{si } \varphi_n - \theta_{n-1} \leq -\frac{\pi}{2} \\ \varphi_n - \theta_{n-1} - \pi & \text{si } \varphi_n - \theta_{n-1} \geq \frac{\pi}{2} \end{cases}$$

dicho de otra forma, la diferencia $\varphi_n - \theta_{n-1}$ se mide según el sentido del giro más corto que lleva un valor sobre el otro.

3.4. Experimentos lineales

En las subsecciones 3.4.1, 3.4.2 y 3.4.3 se detallan pruebas realizadas con los procedimientos comentados anteriormente. Para crear las muestras se generaron 181 valores equiespaciados en el intervalo $[0, 2, 2]$, x_n , correspondientes a la variable x y a continuación se generaron los valores y_n de la variable y según el esquema

$$y_n = a + bx_n + \varepsilon_n .$$

Donde la perturbación, ruido, ε_n se genera a partir de otra perturbación, $\delta_n \sim N(0, 1)$ que denominaremos perturbación base. En el caso de MCO es $\varepsilon_n = \sigma\delta_n \sim N(0, \sigma^2)$ y en el caso de MCG el comportamiento heterocedástico se hace dependiente de la variable y así: $\varepsilon_n = \sigma E[y_n]\delta_n \sim N(0, \sigma^2 E^2[y_n])$. El vector de parámetros es ahora $\beta = \begin{pmatrix} a \\ b \end{pmatrix}$, por lo que tenemos $k = 2$ y la matriz X queda:

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} .$$

Los nueve primeros valores de cada serie, se usaron como puntos semilla, $\{(x_s, y_s)\}$, para la obtención de una estimación inicial para la inicialización del filtro de Kalman. El procedimiento empleado para el cálculo de esa estimación inicial fue el de la mediana repetida, ver sección 5.2. Cuando se detecta un outlier se desprecia.

La separación empleada para los valores de x es del orden de las usales entre ángulos de los rayos de un haz láser medidas en radianes. En los ejes de abcisas de las figuras se ha optado por representar en orden de las x , en lugar de su valor, para dar una información más clara sobre el número de puntos usados en la estimación en cada bucle de los algoritmos.

Hay que tener en cuenta que el error cometido con σ , si no es excesivo, no es tan importante para nuestros objetivos como en cometido con los parámetros de las ecuaciones, a la hora de analizar los resultados de este estudio. Para los parámetros deseamos la mayor precisión posible. En el caso de σ los errores más grandes se dan, con diferencia, por

sobreestimación, así que teniendo en cuenta que $\chi_{exp}^2 = \frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2)$, la región crítica unilateral a nivel α que indica sobreestimación será $\chi_{exp}^2 > \chi_{n-2,\alpha}^2$ que equivale a que el error relativo en términos de la desviación típica σ sea $\frac{\hat{\sigma}-\sigma}{\sigma} > \sqrt{\frac{\chi_{n-2,\alpha}^2}{n-2}} - 1$. Para los tamaños muestrales usados en este estudio el error tolerado para σ queda entre el 12 %, para las muestras grandes, y 30 % para las pequeñas. Valores mayores del error podrían ser admitidos si no comprometen demasiado la detección de outliers.

3.4.1. Experimentos MCO

Se aplicó el algoritmo 3.2 a tres conjuntos de datos generados a partir de la recta horizontal $y = 10$, para los valores de la desviación típica del ruido, $\sigma = 0,002, 0,02$ y $0,2$. En los tres casos los valores del ruido base, δ , fueron los mismos.

En la figura 3.2 se muestran los resultados de las estimaciones. En las subfiguras a), c) y e) aparecen las nubes de puntos, las rectas teóricas en azul y las estimadas en rojo. Bajo estas subfiguras aparece una tabla con los resultados de las estimaciones. Se muestran el valor fijado para σ , los valores iniciales, a_0 y b_0 y las estimaciones finales \hat{a} , \hat{b} y $\hat{\sigma}$. Puede verse que el aumento del error absoluto de las estimaciones es igual al aumento de σ , como se desprende de las ecuaciones 2.8 y 2.10. En las subfiguras b), d) y f) aparecen los valores de los estadísticos para la detección de outliers, t_{exp} , las líneas rojas muestran los valores críticos para un nivel de significación $\alpha = 0,01$. Puede verse que las tres gráficas son iguales, lo que nos muestra el hecho de que el procedimiento es independiente de la varianza de las perturbaciones, salvo por su influencia en el error de las estimaciones. En la subfigura g) se muestra el ruido base, δ , entre las líneas rojas se hallarían comprendidos el 99 % de los valores normales, fuera de ellas estarían los considerados outliers. Se aprecia como en b), d) y f) se reproduce aproximadamente el comportamiento observado en g), lo que da una muestra de la eficacia del procedimiento al filtrar el ruido del proceso y detectar outliers.

3.4.2. Experimentos MCG

Comentaremos ahora unos experimentos en los que se pone a prueba de forma extrema la capacidad del algoritmo 3.3 para filtrar ruido de varianza no constante, al intentar estimar los parámetros de rectas casi verticales de pendientes ± 50 . En estos experimentos la desviación típica del ruido, σ , aumentaba o disminuía en un factor de aproximadamente 10 entre los extremos de las rectas. El factor de compensación de ganancia, ω , utilizado ha sido el máximo de los valores de la variable y en el conjunto de puntos semilla usados para la estimación inicial, $\max\{y_s\}$. Experimentos mucho más detallados se presentan en la subsección 3.4.3.

En las figuras 3.3 y 3.4 se muestran los resultados de las estimaciones de las rectas $y = 10 + 50x$ e $y = 110 - 50x$ para distintos valores de la desviación típica de ruido, σ . Los valores de la semilla corresponden a los más pequeños para la primera recta y a los más grandes para la segunda. Por razones de espacio no se han incluido en las figuras los casos de $\sigma = 0,002$ cuyos resultados se detallan a continuación:

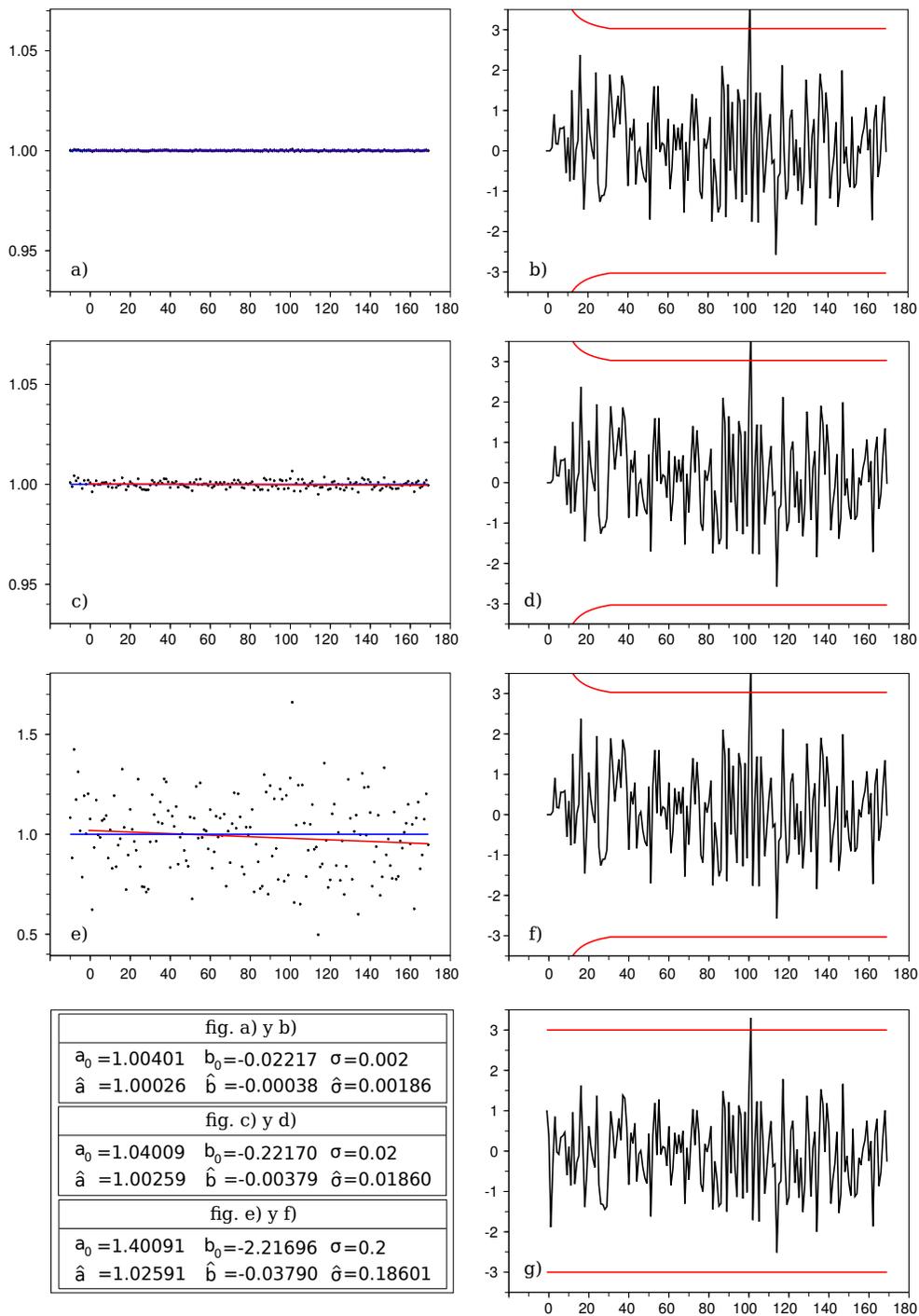


Figura 3.2: Estimación de una línea recta por MCO usando el algoritmo 3.2 para distintas desviaciones típicas de las perturbaciones σ . Los verdaderos parámetros de la línea $y = a + bx$, son $a = 10$ y $b = 0$. Ver comentario en el texto, subsección 3.4.1.

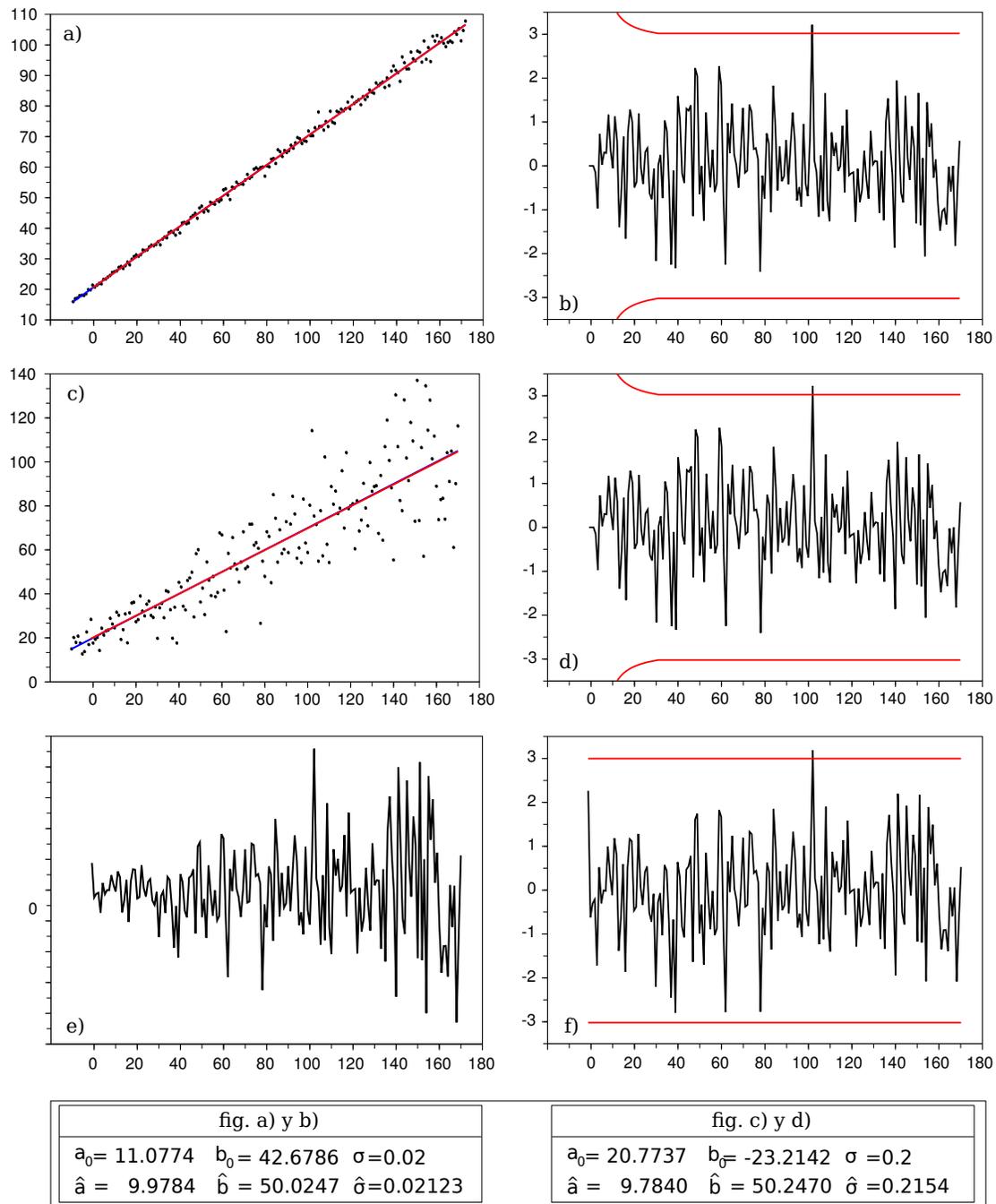


Figura 3.3: Estimación de una línea recta por MCG usando el algoritmo 3.3 para distintas desviaciones típicas, por unidad de y , de las perturbaciones σ . Los verdaderos parámetros de la línea $y = a + bx$, son $a = 10$ y $b = 50$, por lo que el ángulo determinado por b es $\angle b = 88,8542^\circ$. Ver comentario en el texto.

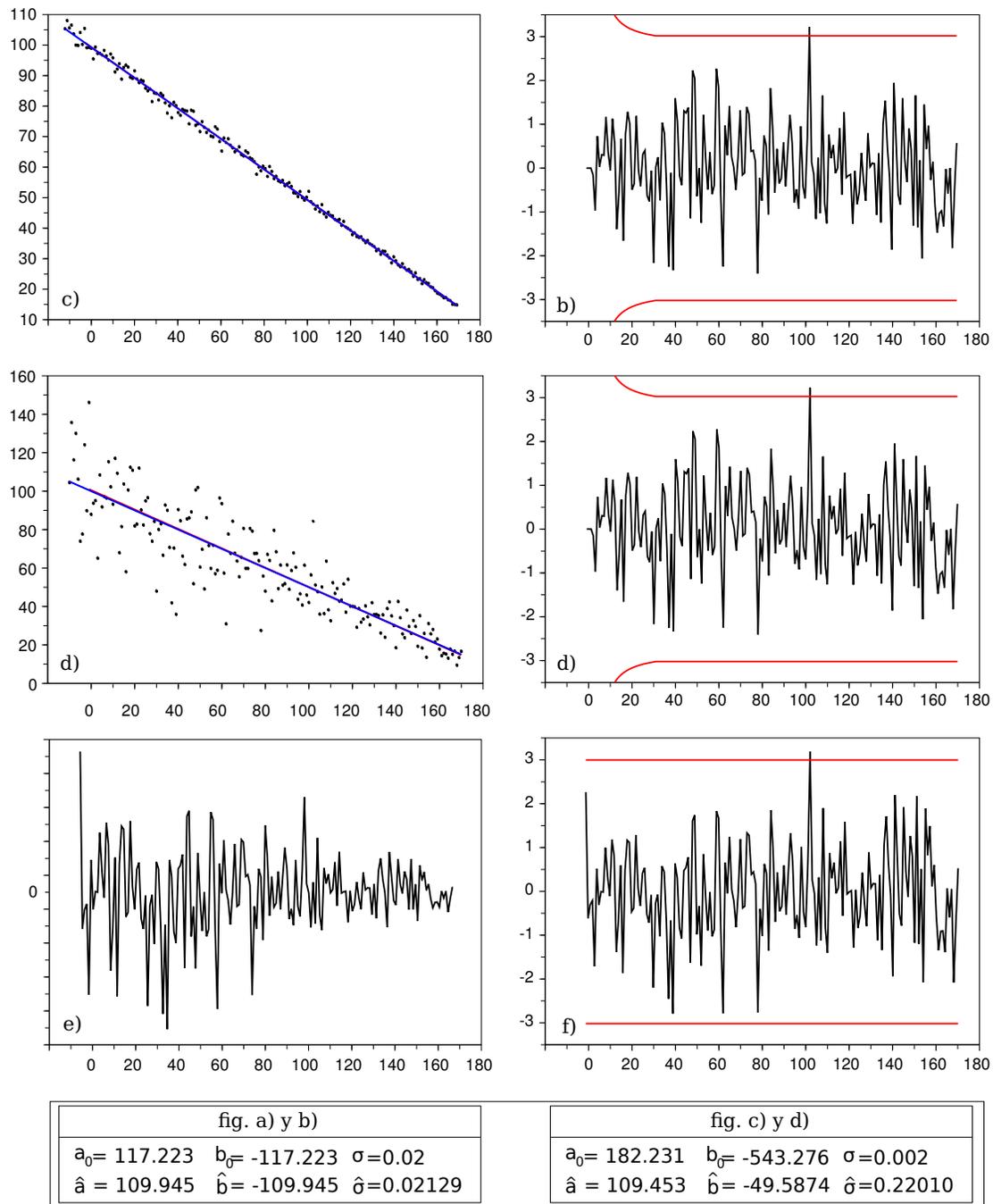


Figura 3.4: Estimación de una línea recta por MCG usando el algoritmo 3.3 para distintas desviaciones típicas, por unidad de y , de las perturbaciones σ . Los verdaderos parámetros de la línea $y = a + bx$, son $a = 110$ y $b = -50$, por lo que el ángulo determinado por b es $\angle b = -88,8542^\circ$. Ver comentario en el texto.

$y = 10 + 50x$	$a_0 = 10,1077$	$b_0 = -49,2679$	$\sigma = 0,002$
	$\hat{a} = 9,9978$	$\hat{b} = 50,0025$	$\hat{\sigma} = 0,00212$
$y = 110 - 50x$	$a_0 = 110,772$	$b_0 = -54,9328$	$\sigma = 0,002$
	$\hat{a} = 109,995$	$\hat{b} = -49,9959$	$\hat{\sigma} = 0,00212$

los comportamientos de los estadísticos de detección son los mismos que los mostrados en las figuras 3.3 b) y d) y 3.4 b) y d), respectivamente.

Pasamos a comentar la figuras citadas y la tabla anterior. En las subfiguras a y, c) aparecen las nubes de puntos, las rectas teóricas en azul y las estimadas en rojo. Bajo estas subfiguras aparece una tabla con los resultados de las estimaciones. Se muestran el valor de σ , los valores iniciales, a_0 y b_0 y las estimaciones finales \hat{a} , \hat{b} y $\hat{\sigma}$. Los errores en las estimaciones de a y b son del orden de σ , a pesar de que el sesgo de la estimación inicial del orden de 50σ . El error relativo de $\hat{\sigma}$ es inferior al que se considera tolerable, 12% para este tamaño de muestra ver el inicio de esta sección 3.4. Puede verse que el aumento del error de las estimaciones es prácticamente igual al aumento de σ , como ocurría en el caso MCO. En las subfiguras b), d) aparecen los valores de los estadísticos para la detección de outliers, t_{exp} , la líneas rojas muestran los valores críticos para un nivel de significación $\alpha = 0,01$. En e) se muestra la serie real de perturbaciones heterocedásticas ε_n para distintos valores de σ las series son versiones escaladas unas de otras, por lo que sólo se muestra una de las gráficas. Esta serie se obtiene a partir del ruido base δ , representado en la subfigura f), entre las líneas rojas se hallarían comprendidos el 99% de los valores normales, fuera de ellas estarían los considerados outliers. Se aprecia como en b) y d) se reproduce aproximadamente el comportamiento observado en f), lo que da una muestra de la eficacia del procedimiento al filtrar el ruido heterocedástico del proceso y detectar outliers, independientemente de la varianza del ruido base y siempre que se conozca el modelo que sigue el ruido real.

3.4.3. Estudio del factor de normalización de ganancia

3.4.3.1. Primeros experimentos

Comenzaremos estudiando el comportamiento del algoritmo con rectas horizontales, $y = a$, y por lo tanto será

$$\sigma_n^2 = \sigma^2 E[y_n]^2 = \sigma^2 a^2.$$

Así podemos ver a σ^2 como el valor que toma la verdadera varianza, σ_n^2 , cuando $y = 1$, o sea, una varianza por unidad o unitaria, y entonces a^2 será la varianza para la recta $y = a$. Este enfoque permite que los resultados que se obtengan pueden ser comparados fácilmente con los obtenidos mediante MCO. De este primer estudio se obtendrán resultados que permitirán abordar el caso general.

Si en una determinada estimación se tuviera que $w_n \simeq 1$, al menos para los primeros elementos de la muestra, se tendría, si n es pequeño, que

$$a_n^p = \frac{1}{w_n^2} + X_n' \Sigma_0 X_n \simeq 1 + X_n' \Sigma_0 X_n = a_n^o$$

y llevando este resultado a la matriz de ganancia del filtro

$$K_n^p = \Sigma_0 \frac{X_n}{a_n^p} \simeq \Sigma_0 \frac{X_n}{a_n^o} = K_n^o$$

De esta manera, al ser las matrices de ganancia similares para las primeras observaciones, el filtro MCP puede responder a la innovación de una manera similar al filtro MCO, ver figura 3.7 a) y b) y tabla 3.1 a). Un ejemplo de este caso se da al estimar la recta $y = 1$ con un error de medida proporcional al dato medido, o sea $\sigma_n = \sigma y_n$ y por tanto $w_n = \frac{1}{\hat{y}_n}$, así se tendrá que $w_n \simeq 1$ para toda la muestra y las estimaciones MCO y MCG serán muy similares, ver tabla 3.1 a).

Sin embargo cambiar el valor de la ordenada en el origen, a , se observó un abanico de comportamientos que no se observó en el caso MCO. Estos comportamientos se pueden ejemplificar mostrando dos casos extremos, mostrados en la tabla 3.1.

En esta tabla se muestran los resultados de las estimaciones de rectas de la forma $y = a$ para varios valores de a , con desviación típica unitaria $\sigma = 0,02$, que equivale a un 2% de la medida, y usando dos series de perturbaciones aleatorias, indicadas por 1 y 2 en la tabla. La primera serie estaba caracterizada por una estimación inicial con error “no pequeño” y la segunda por un error “pequeño”, el error inicial en a para la serie 1 es aproximadamente 4 veces mayor y el de b unas 30. En cada caso las rectas se estimaron con los siguientes pesos: $w_n = \frac{1}{\hat{y}_n}$, la asignación habitual, y $w_n = \frac{\omega}{\hat{y}_n}$ con $\omega = a$, así $w_n \simeq 1$ y estaríamos en el caso equivalente a MCO antes comentado, que se muestra en la subtabla a). A este factor de escala se le ha denominado factor de compensación de ganancia.

Asignación habitual de pesos Usando la asignación habitual de pesos, $\omega = 1$, diferencias palpables entre los resultados obtenidos con ambas series se observa en la subtabla d), la correspondiente a $a = 100$, que produce pesos pequeños. Los errores cometidos en las estimaciones finales son mayores para la serie 1 que para la 2, en especial en \hat{b} y $\hat{\sigma}$, la variación entre las estimaciones iniciales, ver pie de tabla, y las finales es muy pequeña en la serie 2. Al tener pesos pequeños también se tienen ganancias pequeñas sensiblemente menores que las del caso MCO, ya que aproximadamente

$$w_n \ll 1 \implies a_n^p \gg a_n^o \implies K_n^p \ll K_n^o,$$

y por lo tanto el filtro responderá poco a la innovación y buen funcionamiento dependerá mucho de la exactitud de las estimaciones iniciales. Así en la serie 1, que no cuenta con una buena estimación inicial, se obtienen estimaciones finales bastante peores que en el caso MCO, que para a y b serían los que aparecen en la subtabla a) multiplicados por 100.

recta			MCG	MCO
y =	1	\hat{a}	1,0026	1,0026
		\hat{b}	-0,0038	-0,0038
		$\hat{\sigma}$	0,0180	0,0186
		d_t	0,2155	0,2126
	2	\hat{a}	1,0043	1,0045
		\hat{b}	-0,0043	-0,0044
		$\hat{\sigma}$	0,0182	0,0197
		d_t	0,2611	0,2838
a)				

recta			$\omega = 1$	$\omega = 0,1$
y =	1	\hat{a}	0,1001	0,1003
		\hat{b}	-0,0002	-0,0004
		$\hat{\sigma}$	0,0179	0,0180
		d_t	0,1910	0,2155
	0,1	\hat{a}	0,1007	0,1004
		\hat{b}	-0,0006	-0,0004
		$\hat{\sigma}$	0,0175	0,0182
		d_t	0,2321	0,2611
b)				

recta			$\omega = 1$	$\omega = 10$
y =	1	\hat{a}	10,102	10,026
		\hat{b}	-0,1018	-0,0379
		$\hat{\sigma}$	0,0223	0,0180
		d_t	0,5182	0,2155
	10	\hat{a}	9,9029	10,042
		\hat{b}	0,0661	-0,0427
		$\hat{\sigma}$	0,0202	0,0182
		d_t	0,3141	0,2611
c)				

recta			$\omega = 1$	$\omega = 100$
y =	1	\hat{a}	101,78	100,26
		\hat{b}	-6,1504	-0,3789
		$\hat{\sigma}$	0,1112	0,0180
		d_t	1,3017	0,2155
	100	\hat{a}	98,960	100,42
		\hat{b}	0,7062	-0,4274
		$\hat{\sigma}$	0,0206	0,0182
		d_t	0,1754	0,2611
d)				

Tabla 3.1: Estimaciones de varias rectas horizontales con y sin factor de compensación, ω . Los valores iniciales para ambas muestras en el caso de la recta $y = 1$ fueron $a_0 = 1,0401$ y $b_0 = -0,2217$, para la muestra 1 y $a_0 = 0,9895$ y $b_0 = 0,0080$, para la muestra 2. Para obtener los parámetros iniciales de las otras rectas basta multiplicar estos valores por el valor constante de y de cada una.

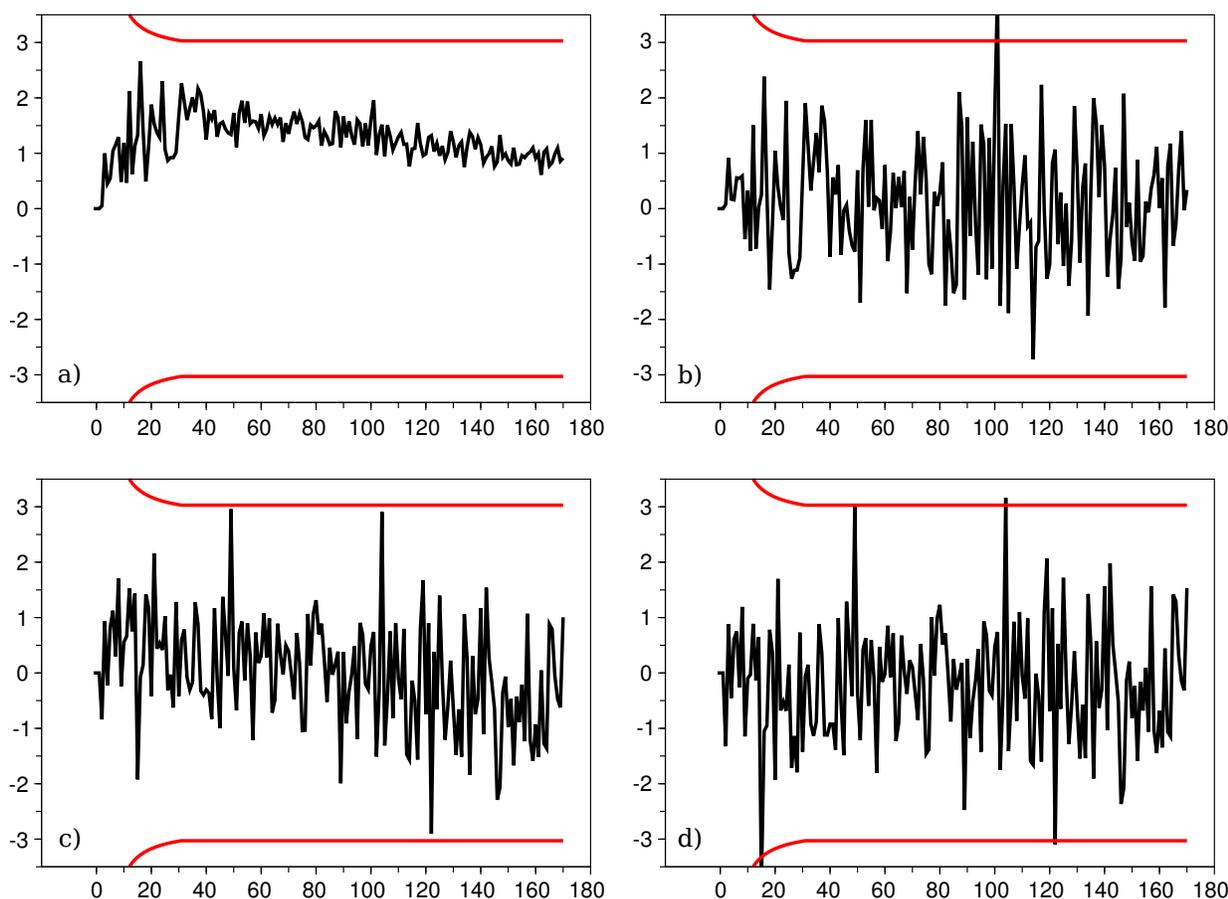


Figura 3.5: Distintas formas de evolución de los estadísticos de contraste. En a) y c) el comportamiento no parece totalmente aleatorio, sin embargo en b) y d) sí.

La serie 2 que cuenta con una buena estimación inicial produce estimaciones finales con errores comparables con los del caso MCO.

Una análisis más detallado de la evolución de los valores de las estimaciones nos muestra otras diferencias entre los resultados obtenidos para ambas series. Recordemos que en el caso MCO la variación del error de las estimaciones finales era igual a la variación de σ , ahora a toma el papel σ como se indicó al inicio de esta subsección, pero la situación es distinta. Para la serie 1 el aumento del error es en todos los casos bastante superior al de a , en especial para la estimación de la pendiente. Sin embargo para la serie 2 se dan casos de variaciones del error mayores o menores que las de a . Mientras que a varía en un factor de 1000, a lo largo de los experimentos de la tabla, el error de \hat{a} lo hace en factores aproximados de 17800 y 1500 y el error de \hat{b} de 30750 y 1200, para las series 1 y 2 respectivamente. Lo que vuelve a mostrarnos los problemas que aquejan a la estimación de la primera serie y la relativa inmunidad de la segunda.

El sesgo en la estimación de la serie 1 para debía producir errores e_n^p , y por lo tanto estadísticos de contraste t_{exp} , con comportamientos anómalos. De hecho la gráfica de los

estadísticos de contraste mostraba una media sensiblemente distinta de 0 y una variabilidad reducida con respecto a la del ruido base, lo que provocaba que en muchos casos no se detectara outliers, ver figura 3.5 a), la serie 2 no mostraba este comportamiento, ver figura 3.5 c), aunque parece observarse cierta tendencia descendente.

El comportamiento de las t_{exp} es un reflejo del de los errores de estimación, ya que son los errores normalizados, ver ecuación 3.10, seguiremos en la discusión refiriéndonos a los estadísticos en lugar de a los errores ya que permiten una mejor comparación al tener sus valores estandarizados. Por otro lado el comportamiento de las t_{exp} debe ser muy similar al del ruido base δ , cuando el filtro funciona correctamente, por lo que se propuso una medida de adecuación del filtro que reflejara este hecho, dando una distancia entre la serie de estadísticos y el ruido base según la siguiente expresión:

$$d_t = \frac{\sum_{n=4}^N |t_{exp,n} - \delta_n|}{N - 3}$$

siendo N el número total de observaciones, no se tienen en cuenta las tres primeras ya que el cálculo del estadístico precisa de más de tres valores. Se podrían definir otras medidas empleando cuadrados en lugar de valores absolutos o el máximo en lugar de la media.

El valor de d_t puede depender de la combinación de varios factores. El más influyente, en especial si la muestra es grande, quizás sea la precisión de las estimaciones finales. En segundo lugar puede situarse la precisión de las estimaciones iniciales que condiciona los valores iniciales de los estadísticos, es al inicio de la estimación donde se pueden producirse las mayores diferencias con el ruido, si la muestra es pequeña este factor es muy determinante. Por último la estabilidad de las estimaciones durante el proceso, una vez pasadas las primeras observaciones el error de las estimaciones debe disminuir paulatinamente, si no podría darse el caso de que teniendo una buena estimación inicial y una buena final, se tuvieran estimaciones intermedias de peor calidad. El experimento consignado en la subtabla c) para la serie 2 muestra este efecto de forma clara: la calidad de sus estimaciones iniciales y finales es muy similar a la que se tiene en la subtabla d) para la misma serie, sin embargo los valores d_t son bastante diferentes, se comprobó que la gráfica de las diferencias que forman d_t mostraba una zona central alejada de cero. Este comportamiento tiene que ver con la forma de la evolución de la ganancia en función del número de observaciones filtradas, como se verá más adelante en esta subsección. La serie 1 produce un comportamiento distinto, con las d_t aumentando o disminuyendo según lo haga a .

Para valores de a inferiores a 1, subtabla b) se tendrá que $w_n > 1 \implies K_n^p > K_n^o$, este aumento de ganancia aumenta la variabilidad de los estadísticos de contraste, especialmente en los primeros momentos de la estimación, pudiendo producir falsos positivos para valores pequeños de a . En este caso no se observan grandes diferencias entre las estimaciones finales para las series 1 y 2.

Asignación de pesos con escala Cuando se procedió a escalar los pesos con $w = a$ los resultados obtenidos seguían la pauta de los comentados para MCO, con variaciones en los errores de las estimaciones finales iguales a la variación de a , además las estimaciones

finales son mejores que para el caso de la asignación habitual, $\omega = 1$. Las distancias d_t tienen para cada serie el mismo valor independientemente de a . En la figura 3.5 b) y d) se observa un comportamiento de los estadísticos de contraste muy similar al del caso MCO, lo que muestra que se está filtrando de forma adecuada el ruido.

3.4.3.2. Método de MonteCarlo

Tras el estudio comparativo de dos casos muy distintos de comportamiento del algoritmo cuando se usa la asignación de pesos habitual, $\omega = 1$, se va a proceder a continuación a intentar obtener el comportamiento de la ganancia en función de la ordenada en el origen, a , y del número de puntos estimados en cada paso, n , y el comportamiento de la distancia d_t en función de a para muestras grandes y para muestras pequeñas. En lugar de estimar rectas de la forma $y = a$ con $\omega = 1$ se puede estimar la recta $y = 1$ con $\omega = a$ para distintos valores de a , así los resultados son más fácilmente comparables. Los resultados son fácilmente generalizables cualquier recta de la forma $y = a'$, entendiendo que en este caso los valores de ω que se usan a continuación no se deben entender en términos absolutos sino como factores de a' , de esta manera los pesos son los mismos que los que se los obtenidos para la recta $y = 1$.

Comportamiento de la ganancia Para estudiar el comportamiento de la ganancia se emplearon 1000 series aleatorias, de 181 valores. Para los valores de ω , ω_i , se dividió el intervalo $[-2, 2]$ en 41 puntos, c_i separados 0,1 unidades, siendo $\omega_i = 10^{c_i}$, o sea, $0,01 \leq \omega \leq 100$. A continuación se realizaron los 41000 procesos de estimación.

La ganancia K_n^p tiene, para las rectas, dos componentes: el primero rige el comportamiento de la estimación de a y el segundo el de la de b , los llamaremos K_a y K_b respectivamente. En la figura 3.6 se muestran los comportamientos de K_a y K_b para $0,5 \leq \omega \leq 10$ y empleando los primeros 25 valores de cada serie. Se observan tres zonas de comportamiento que se aprecian mejor la figura 3.7. Una zona corresponde a los valores centrales de ω , en ella los comportamientos de K_a y K_b son similares al caso MCO. Un ejemplo de la evolución de la ganancia en el caso MCO se puede ver en la figura 3.7 a), funciones monótonas que tienden a 0, en b) tenemos el caso representativo de los valores centrales de ω . En las otras dos zonas, valores pequeños y grandes de ω , los extremos, máximos o mínimos, de las ganancias no se encuentran justo al principio de la estimación. La diferencia entre estas dos zonas está en la escala de la evolución de las ganancias, tanto mayor cuanto mayor sea ω , como puede verse claramente en la figura 3.7 c) a f), y en el momento en el que se produce el extremo: en momentos iniciales de la estimación cuando ω es grande, pudiendo alterar considerablemente los valores de los estadísticos iniciales, y ya avanzada la estimación cuando ω es pequeño. En la zona central se tiene que, aproximadamente, $0,6 \leq \omega \leq 6$.

Comportamiento de d_t Para estudiar el comportamiento de las d_t se emplearon las mismas 1000 series aleatorias que en el caso de la ganancia, de 181 valores. Para los valores de ω , ω_i , se dividió el intervalo $[-2, 2]$ en 41 puntos, c_i separados 0,1 unidades, siendo

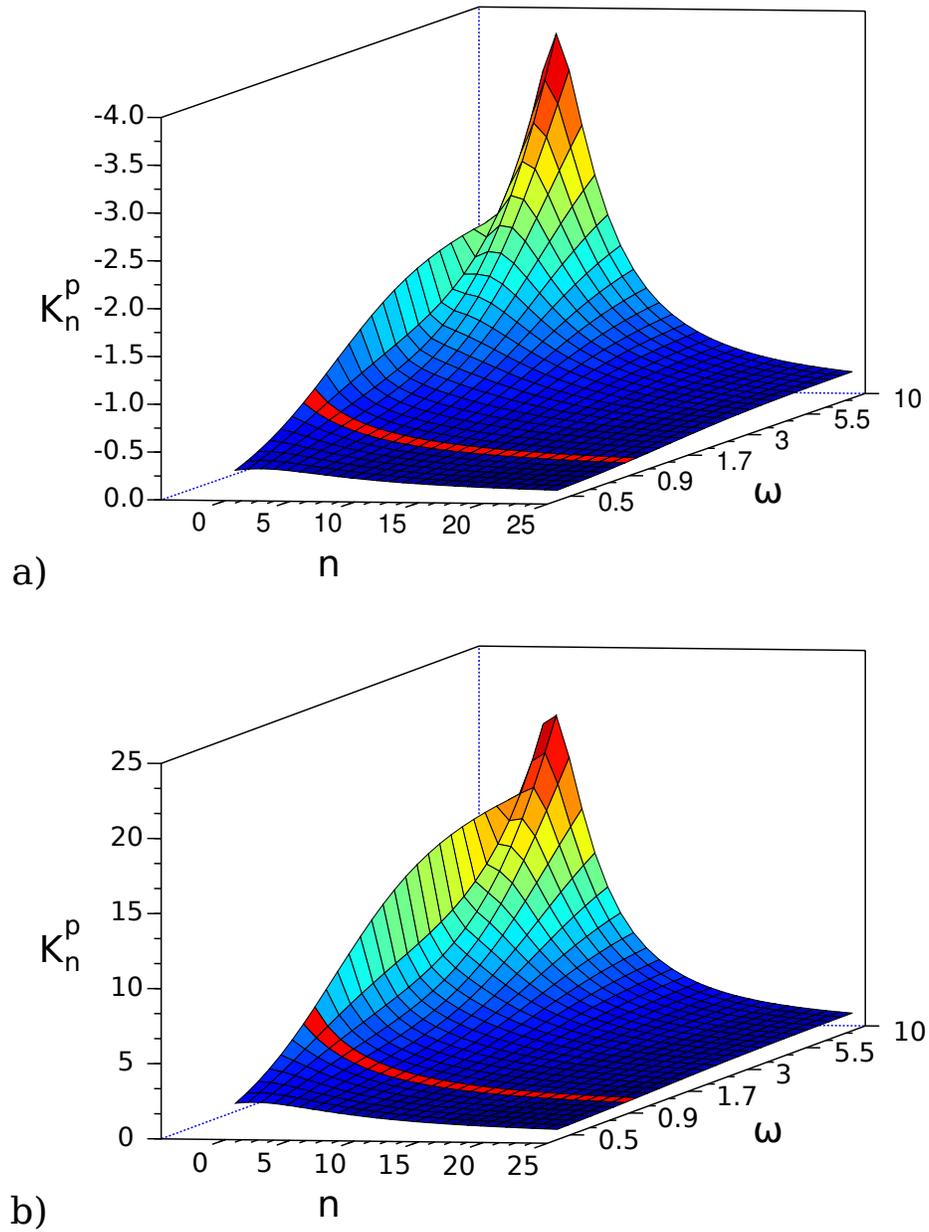


Figura 3.6: Comportamiento de la ganancia en función del número de puntos incluidos en el filtro, n y del factor de compensación ω . En a) para el parámetro a y en b) para b .

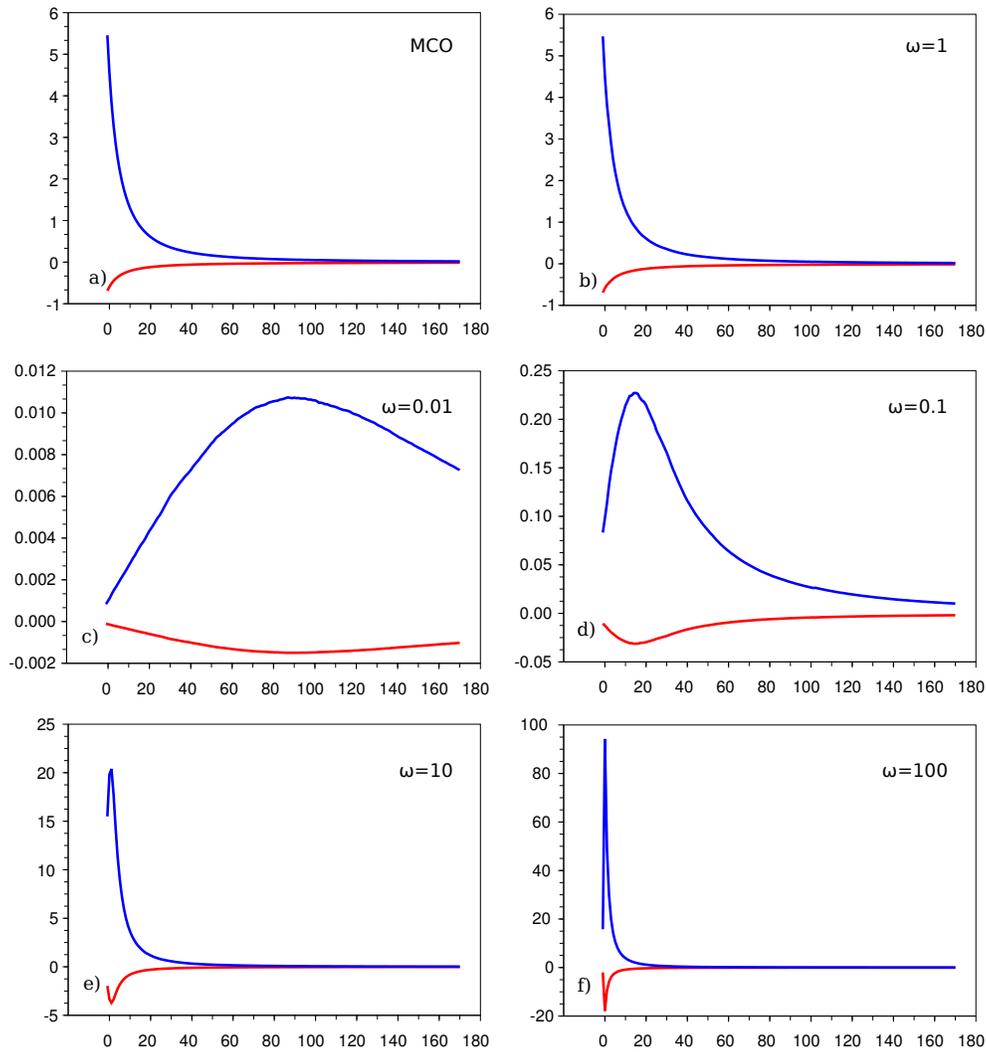


Figura 3.7: Comportamiento de la ganancia para diversos valores del factor de compensación. El comportamiento como MCO a), se produce en $-0,2 \lesssim \log(\omega) \lesssim 0,7$, o sea, $0,6 \lesssim \omega \lesssim 5$ para a y $-0,2 \lesssim \log(\omega) \lesssim 0,85$, o sea, $0,6 \lesssim \omega \lesssim 7$ para b . La línea azul (superior) corresponde a \hat{b} y la roja (inferior) a \hat{a} .

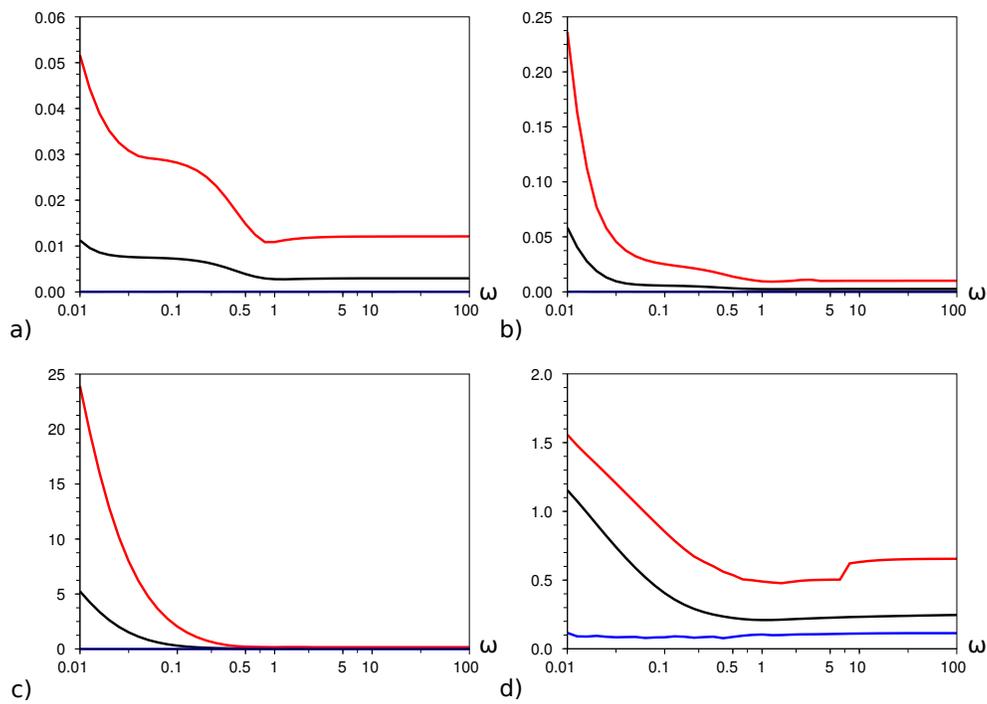


Figura 3.8: Errores y valores de d_t en la estimación de rectas horizontales dependiendo del factor de ganancia. Valores máximos (rojo), medios (negro) y mínimos (azul), para a en a), para b en b), para σ en c) y para d_t en d).

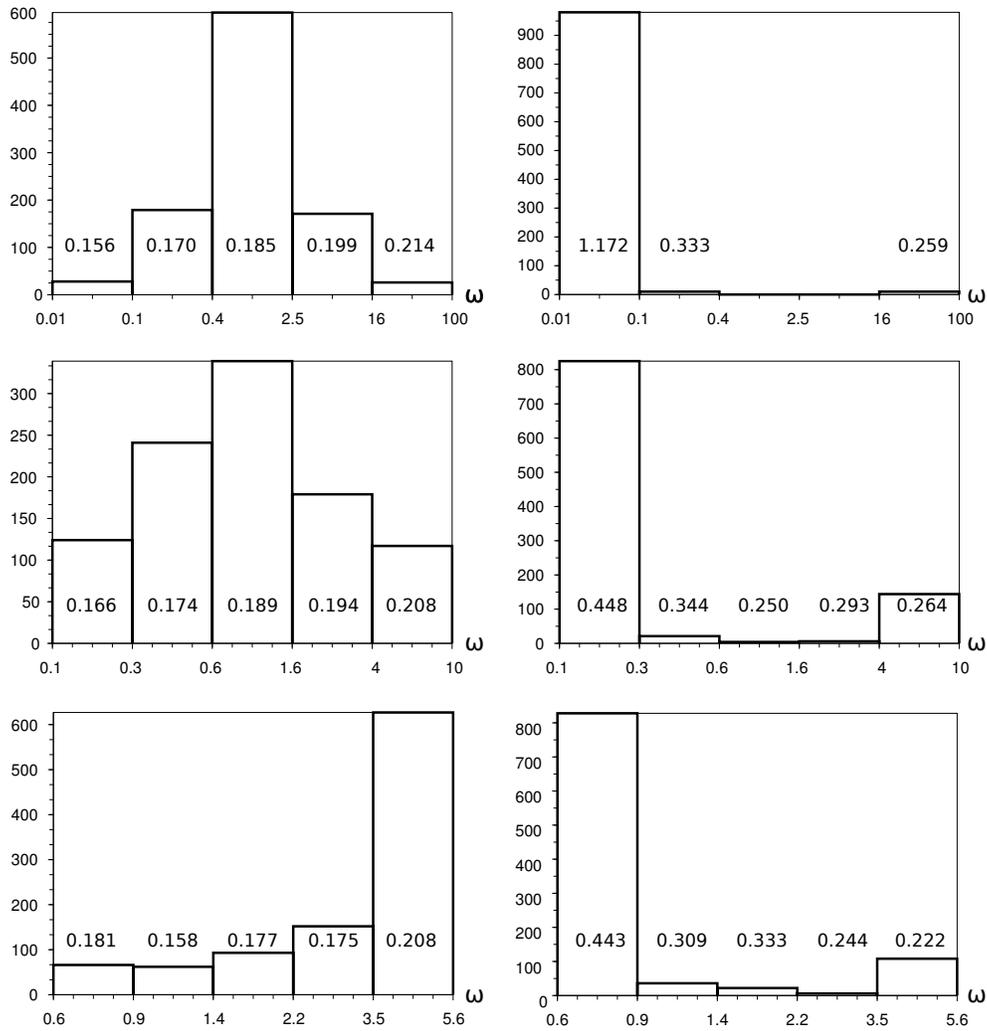


Figura 3.9: Histogramas de d_t con 181 puntos en función de ω , para distintos rangos de ω , al estimar rectas horizontales.

$\omega_i = 10^{c_i}$, o sea, $0,01 \leq \omega \leq 100$. A continuación se realizaron los 41000 procesos de estimación y se calculó el valor de d_t para cada una de ellas.

Para \hat{a} y $\hat{\sigma}$ se ha trabajado con los errores relativos en valor absoluto. En el caso de $\hat{\sigma}$ es irrelevante emplear errores absolutos o relativos ya que siempre se estima la misma cantidad, se han empleado los relativos por considerarlos más informativos a la vista, al presentar ratios. Para \hat{a} y $\hat{\sigma}$ se ha trabajado con los errores relativos en valor absoluto. En el caso de $\hat{\sigma}$ es irrelevante emplear errores absolutos o relativos ya que siempre se estima la misma cantidad, se han empleado los relativos por considerarlos más informativos a la vista, al presentar ratios. Para el caso de \hat{a} se han usado los relativos por dos razones, además de la ya comentada para $\hat{\sigma}$, la primera es que varía a y la segunda es que dado a es el valor de y en $x = 0$, o sea, podemos entender que \hat{a} , se verá sometida a errores variables como le ocurre a y . Sin embargo para \hat{b} usaremos los errores absolutos en valor absoluto ya que por un lado se quiere estudiar el valor $b = 0$, lo que impide el cálculo del error relativo. También se han incluido los valores de las distancias d_t .

Se han calculados las medias, los máximos y los mínimos de los datos agrupándolos según los valores de la pendiente, sus gráficas se encuentran también en la figura 3.8. Como puede observarse en todos los casos el comportamiento de los errores y las d_t es muy similar decrece desde valores relativamente altos y se estabiliza para $\omega \lesssim 1$, gráfica de los máximos de las d_t muestra un ligero, pero abrupto, repunta hacia el valor $\omega \sim 6$. Estos valores de ω concuerdan con los deducidos del estudio de la ganancia.

Para una mayor seguridad y precisión en la estimación de los valores adecuados para ω se procedió a contar, para cada valor de ω , las series cuya estimación que presentaban el mínimo de las d_t en ese valor de ω , de la misma forma se procedió con los máximos. Los resultados se presentan en forma de histogramas en la figura 3.9 a) y b), sobre o dentro de los rectángulos aparecen las medias de las d_t para cada clase.

Los mínimos se distribuyen principalmente por la zona central del histograma, $0,1 \leq \omega \leq 10$, las medias de las distintas clases son similares aunque parece que tienden a aumentar con ω . Los máximos aparecen casi por completo con los valores más pequeños de ω , con una media alta, similar a la que produce el comportamiento observado en la figura 3.5 a). Descartamos las clases primera y quinta como valores adecuados para ω , aunque el efecto de los valores grandes no parece tan influyente como el de los pequeños.

A continuación se procedió con el intervalo $[-1, 1]$, $0,1 \leq \omega \leq 10$, como se había hecho antes con $[-2, 2]$ obteniendo también 41 valores. Los histogramas resultantes se hallan en la 3.9 c) y d). Vuelve a verse una situación similar a la ocurrida con el intervalo $[-2, 2]$, aunque ahora no es tan acusada la preponderancia del intervalo central para los mínimos, y la media en la primera clase para los máximos no es tan elevada como en el caso anterior.

Por último de los datos anteriores se extrajeron los correspondientes a un intervalo similar al sugerido como adecuado por el comportamiento de las ganancias $0,6 \leq \omega \leq 6$. Los histogramas de mínimos y máximos aparecen en la figura 3.9 e) y f). Ahora la situación para los mínimos parece distinta ya que se agolpan en la última clase, aunque este detalle no es muy relevante ya que las medias de cada clase son muy parecidas. En los máximos vuelve a darse la situación descrita anteriormente, por lo que descartaremos la primera clase del conjunto de valores adecuados de ω , Dado que ya no se ven muchas diferencias entre

las medias de los máximos y los mínimos entre las clases restantes se puede recomendar el siguiente conjunto de valores adecuados de ω , el extremo inferior muy próximo a 1 y el superior del orden de 6.

A modo de resumen se puede recomendar que para una recta horizontal de la forma $y = a$ los pesos se definan como $w_i = \frac{\omega}{y_i}$, con $a \leq \omega \leq 6a$.

3.4.3.3. El caso general

A continuación vamos a tratar el caso de rectas no necesariamente horizontales, en las que, por lo tanto, la varianza del error no es constante a lo largo de la recta. La principal duda en este caso está en el comportamiento de la ganancia, ya que los pesos no serán aproximadamente constantes y puede que no se consiguiera la compensación de la ganancia conseguida para las rectas horizontales. Sin embargo dado que la ganancia, en el caso del comportamiento “adecuado”, tiene sus mayores valores al principio disminuyendo rápidamente, se pensó que la condición $w_n \simeq 1 \forall n$ se podía debilitar pidiendo que $w_n \simeq 1$ para los primeros n , con la esperanza que para valores mayores de n el efecto de los pesos no vicie el proceso de estimación. Para asemejarse en lo posible al caso MCO y por razones que se expondrán más adelante, ver subsección , se ha tomado el criterio de de asignar a ω el máximo de los valores absolutos de y en el conjunto de los puntos semilla

$$\omega = \max \{|y_s|\}$$

Dado que los conjuntos semilla tienen pocos puntos el valor $\frac{\max\{|y_s|\}}{\min\{|y_s|\}}$ no se escapa del rango de valores adecuados para ω que, trasladando lo obtenido para las rectas horizontales, deber estar entre 1 y 6 veces el rango de valores iniciales. En la subsección 3.4.2 se comentaron dos experimentos de estimación de recta casi verticales. En esta subsección haremos un tratamiento estadístico empleando un rango amplio de rectas y series aleatorias.

Se generaron 221 rectas con pendientes, b , en el intervalo $[-50, 50]$ valores que corresponden a $\pm 88,85^\circ$, 200 pendientes estaban $[-5, 5]$, $\pm 78,69^\circ$, a tomó valores en $[5, 10]$. Los puntos semillas correspondían a los de ordenada más pequeña de la muestra cuando era $b \geq 0$ y a los de mayor cuando en caso contrario, de esta manera se contemplan los dos casos posibles de posiciones de la semilla. La desviación típica unitaria del ruido fue $\sigma = 0,02$, para otros valores de σ los resultados habrá que escalarlos como se comentó en la subsección 3.4.3.1. Cada recta se estimó usando 100 series de ruido base, se realizaron por lo tanto 22100 estimaciones.

Para \hat{a} y $\hat{\sigma}$ se ha trabajado con los errores relativos en valor absoluto y para \hat{b} los errores absolutos en valor absoluto como se comentó en los experimentos subsección anterior 3.4.3.2. También se han incluido los valores de las distancias d_t . Para todos estos datos se han calculado la media, el máximo, el mínimo y la desviación típica, cuyos resultados aparecen en la tabla 3.2 a) y sus observaciones han sido agrupadas en histogramas, ver figura 3.10.

El histograma de los errores relativos de \hat{a} , figura 3.10 a), muestra una gran asimetría negativa, una alta concentración de la distribución en los valores pequeños, de manera que se puede afirmar que la práctica totalidad de los errores son inferiores al 1 %, el error medio

a)	media	máx.	min.	st. dev.
$E_{rel}(\hat{a})$	0,003007	0,038892	0,000000	0,002705
$E_{abs}(\hat{b})$	0,033308	0,307672	0,000006	0,028995
$E_{rel}(\hat{\sigma})$	0,051360	0,156199	0,000064	0,036464
d_t	0,205149	0,460789	0,075560	0,056386
b)	media	máx.	min.	st. dev.
$E_{rel}(\hat{a})$	0,003067	0,049131	0,000000	0,002994
$E_{abs}(\hat{b})$	0,032734	0,351579	0,000000	0,029410
$E_{rel}(\hat{\sigma})$	0,044935	0,140288	0,000002	0,031754
d_t	0,093839	0,189283	0,007411	0,037065
c)	media	máx.	min.	st. dev.
$E_{rel}(\hat{a})$	0,008583	0,084781	0,000000	0,007195
$E_{abs}(\hat{b})$	0,391905	8,455637	0,000003	0,452074
$E_{rel}(\hat{\sigma})$	0,137983	0,593921	0,000013	0,112978
d_t	0,362671	0,837461	0,098914	0,144830
d)	media	máx.	min.	st. dev.
$E_{rel}(\hat{a})$	0,011889	0,121000	0,000000	0,010404
$E_{abs}(\hat{b})$	0,498301	8,963704	0,000003	0,574265
$E_{rel}(\hat{\sigma})$	0,130142	0,476718	0,001367	0,103987
d_t	0,223451	0,618187	0,038739	0,113183

Tabla 3.2: Medidas de los errores cometidos al estimar 22100 rectas en a) mediante en procedimiento secuencial se propuesto en este trabajo, en b) la estimación directa por MCG, se estima usando todos los datos a la vez. En c) y d) sólo se usaron los primeros 25 puntos en la estimación.

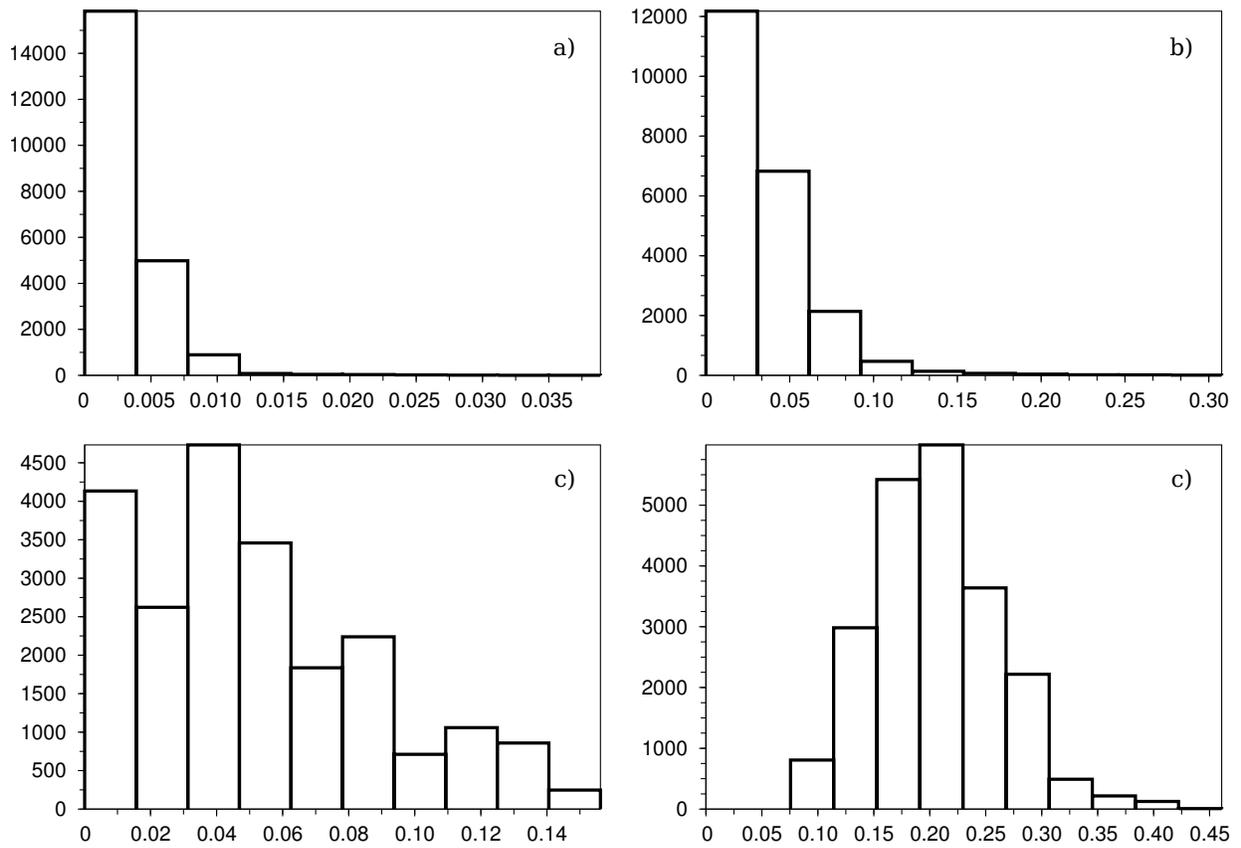


Figura 3.10: Histogramas de los errores para a) a b) b c) σ y valores d_t d) rectas con $b = [-50:5:-5, -4.95:0.05:4.95, 5:5:50]$, 100 muestras $\sigma = 0,02$

es del 0,3 %, errores superiores al 1,2 % sólo aparecen se forma testimonial y las alturas de sus rectángulos casi no se aprecian en el histograma, no aparecen errores superiores al 4 %.

El histograma de los errores relativos de \hat{b} , figura 3.10 b), muestra una gran asimetría negativa, una alta concentración de la distribución en los valores pequeños, de manera que se puede afirmar que la gran mayoría de los errores son inferiores a 0,1 ($\sim 5,7^\circ$), el error medio es del 0,033 ($\sim 1,9^\circ$), errores superiores a 0,15 ($\sim 8,5^\circ$) sólo aparecen en se forma testimonial y las alturas de sus rectángulos casi no se aprecian en el histograma, no aparecen errores superiores a 0,31 ($\sim 17,2^\circ$) .

El histograma de los errores relativos de $\hat{\sigma}$, figura 3.10 c), muestra una asimetría negativa, una cierta concentración de la distribución en los valores pequeños, se puede afirmar que un número considerable de errores son inferiores al 10 %, el error medio es del 5,1 % y no aparecen errores superiores al 15,6 %. Por lo que en general error relativo de $\hat{\sigma}$ es inferior al que se considera tolerable, 12 % para este tamaño de muestra ver el inicio de esta sección 3.4.

El histograma de las distancias d_t , figura 3.10 d), muestra una distribución aproximadamente normal, aunque un poco apuntada, con media 0,205 y desviación típica de 0,056, con un mínimo en 0,076 y una corta cola de valores grandes hasta 0,46.

Por último compararemos estos resultados con los que se obtienen al aplicar el procedimiento tradicional de MCG, ver subsección 2.4.1.2. A la hora de la comparación hay que tener en cuenta que este procedimiento es “off-line”, se estima con todos los datos juntos una vez adquiridos, el procedimiento que se propone en este trabajo es “on-line”, se estima secuencialmente conforme llegan los datos. Los estadísticos de contraste, t_{exp} , han sido calculados “a posteriori”, o sea, usando los estimadores finales de a , b y σ , por lo que el valor de d_t representaría una especie de límite inferior de la distancia entre estadísticos y ruido base.

Los resultados de la estimación se muestran en la tabla 3.2 b). Los resultados obtenidos mediante ambos procedimientos para los errores a , b , y σ son muy similares. Los resultados obtenidos para las d_t en el procedimiento propuesto muestran una media es unas 2 veces mayor que en el de MCG y el máximo unas 2,5. Podemos concluir que la elección del factor de compensación, ω , es acertada ya que hace que el filtro “on-line” tenga la misma precisión en sus estimaciones que el procedimiento “off-line”. También se muestran en la tabla 3.2 c) y d) los resultados cuando sólo se había usado 26 puntos para la estimación, este valor viene sugerido por el gráfico de ganancias, figura 3.6, en el que se observa una relativa estabilización al llegar a este valor. Los resultados vuelven a mostrar la misma situación que cuando se usaban todos los puntos: errores similares en las estimaciones entre el procedimiento propuesto y MCG y una relación entre las d_t similar a la observada anteriormente.

3.5. Experimentos en coordenadas polares

3.5.1. Descripción

Los resultados anteriores en modelos lineales son muy alentadores en cuanto a la eficacia del procedimiento propuesto, pero al pasar al terreno no lineal, en concreto a las coordenadas polares, hay que tener cuidado ya que pueden aparecer problemas, también en los procedimientos “off-line”, de muy diversa índole: estadísticos, comentados en la subsección 2.4.1.3, de determinación de ángulos en el caso de los jacobianos, comentado en la subsección 3.3.2, de rigidez de la matriz hessiana, $J'J$, en los procedimientos “off-line” y el efecto de los pesos, como veremos a continuación.

En esta sección se va a comparar el procedimiento propuesto con el de (LM), que es “off-line”. En la fórmula del jacobiano J , ver subsección 3.3.2, aparecen denominadores que pueden tomar valores cercanos a 0, y con signos distintos, que pueden provocar problemas de rigidez en la matriz hessiana que malogran el resultado de la estimación de cualquier método de aproximación lineal², como es LM. Fijar las determinaciones de los ángulos en el cálculo de J , ayuda a disminuir el número de estimaciones malogradas, como se pudo comprobar. La aplicación del algoritmo LM se hace siguiendo el esquema MCG, o sea, una primera estimación MCONL de la que se obtienen los pesos y una segunda estimación ponderada.

Para fijar ideas recordemos que el modelo a estimar es

$$r = \frac{\rho}{\cos(\varphi - \theta)} + \varepsilon$$

con $\varepsilon \sim N(0, \sigma^2 E^2[r])$, ρ y θ son los parámetros a estimar. Tenemos una desviación típica del error proporcional a la verdadera medida. Como este procedimiento se diseña para ser aplicado a medias obtenidas por sensores de rango, conviene plantearnos con qué rango de errores queremos trabajar. No podemos esperar que en un procedimiento no lineal se mantengan, para todos los valores de σ , las relaciones observadas entre su variación y las de los errores de estimación como se vió que ocurría en el caso lineal. Los errores habituales de los escaners de láser³ usados suelen ser del orden de 1 cm. a 10 m., o sea del 0.1 % si pensamos que los errores son proporcionales a las medidas, lo que equivaldría a una desviación típica unitaria $\sigma = 0,001$. Usaremos en los experimentos el valor $\sigma = 0,002$, para tener en cuenta efectos debidos al material contra el que choca el haz como: rugosidad, color, reflectancia, etc., estos serán los errores usuales que proporcionen las superficies rectas más o menos pulidas, como paredes, muebles, etc. También se estudiará el comportamiento del procedimiento para errores diez veces más grandes $\sigma = 0,02$, que serían los que producirían objetos como setos de jardín, cortinas, paredes derruidas, etc. Así se comprobará si los resultados siguen siendo independientes de σ , salvo el factor de escala, como lo eran en el caso lineal, ver subsección 3.4.2. Como unidad de medida se usará el metro.

²Aproximaciones de mayor orden aumentan el problema con los denominadores.

³La precisión de los sónares usados en robótica es similar cuando el ángulo de incidencia es relativamente pequeño.

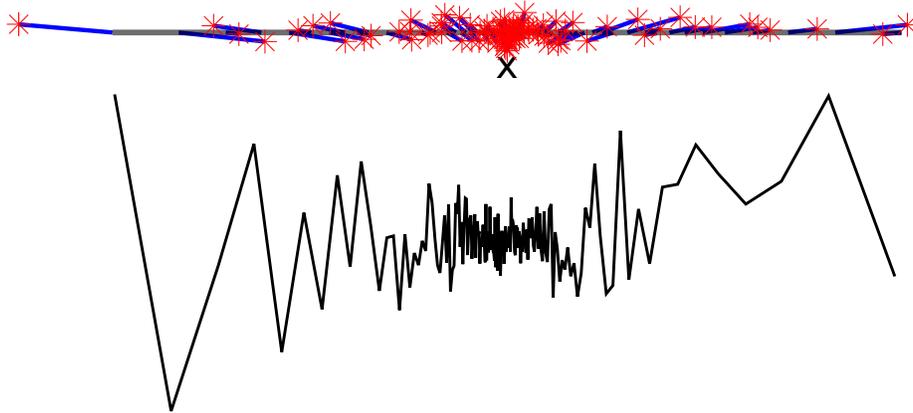


Figura 3.11: Errores radiales. La zona superior muestra una recta y una serie de puntos generados a partir de ella con error proporcional a la distancia al aparato de medida, X, cuyo eje es perpendicular a la recta. Visto en esta representación parece que la mayor dispersión se da en la parte central de la recta, donde realmente debe ser menor, ya que ahí las distancias al origen de medición son menores. Si se observan los segmentos que unen los puntos con la recta, que indican los verdaderos errores en sentido radial, se verá que nos es así, los mayores errores se sitúa en los extremos, en la gráfica inferior están representados dichos errores para una mejor apreciación.

Supongamos que el aparato de medición, representado con un X en la figura 3.11, tiene una apertura de haz de 360° , así se percibe la recta casi por completo y hace la discusión independiente de la dirección del eje central del aparato. En el planteamiento en polares hace independiente el ángulo que forma un segmento rectilíneo con el eje del aparato de medición, ver pie figura 3.1, ya que lo único que puede variar es el ángulo de incidencia del haz, $\varphi - \theta$, al principio y al final del segmento que se quiera estimar. Por esto sin pérdida de generalidad se puede suponer $\theta = 0$.

Podemos reescribir el modelo como:

$$\frac{r}{\rho} = \frac{1}{\cos(\varphi - \theta)} + \frac{\varepsilon}{\rho}$$

o lo que es lo mismo

$$r' = \frac{\rho'}{\cos(\varphi - \theta)} + \varepsilon' \text{ con } \rho' = 1$$

y de esta manera se hace la discusión, y el procedimiento, independiente del valor de ρ , que queda como un factor de escala que actúa sobre su propia estimación y sobre la de la desviación típica de la perturbación sin embargo no afecta a la estimación de θ . El factor de escala se cancela cuando se emplean errores relativos, por lo que los resultados de estos experimentos no cambiarían al variar ρ . Se hicieron pruebas cambiando el valor de ρ hasta un factor de 10^5 sin obtener la más mínima diferencia entre los resultados.

Dado que los resultados no dependen de ρ y que para las muestras usadas se considerarán todos los ángulos posibles de inicialización, salvo los muy próximos a 90° , o lo que es lo mismo todos los posibles ángulos de incidencia, estos experimentos abarcan de forma exhaustiva el espacio paramétrico de ρ y θ , para los tamaños de muestra y valores de σ empleados.

Para estudiar el efecto de la variación del factor de compensación de ganancia, ω , dependiendo de la posición, ángulos de incidencia, de la semilla de inicialización se procedió a generar segmentos de la recta $r = \frac{1}{\cos(\varphi)}$ de una apertura de $88,5^\circ$, compuestos de 181 puntos separados $\sim 0,5^\circ$. Se generaron 100 segmentos, a partir de las correspondientes series de ruido, para cada valor de φ , ahora el ángulo de incidencia, desde $88,5^\circ$ hasta 0° . La diferencia máxima de distancias entre los extremos de un segmento podía llegar a 37,8 m., siendo, por lo tanto, el rango de ω es intervalo $[1, 37,8]$ y también el del factor de variación máximo de σ . Como de costumbre los primeros 9 valores se usaron para la semilla, para la estimación inicial se usó el método de la mediana repetida en coordenadas polares, ver subsección 5.2.

Se repitió a continuación el experimento pero usando series de 35 puntos, 9 de ellos de semillas, abarcando unos 17° , el ángulo de incidencia del comienzo del segmento se hizo variar desde $-88,5^\circ$ hasta $71,8^\circ$. De esta manera podremos estudiar la características del método en muestras más pequeñas, como se hizo en el caso lineal.

Se han calculado los errores en valor absoluto de las estimaciones, relativos para ρ y σ , ya absolutos para θ . Este criterio es el mismo que se siguió en los experimentos lineales, subsección 3.4.3.2, ρ es una distancia al igual que a en el caso lineal, θ es un ángulo y b es la tangente de un ángulo, por lo que les aplicamos el mismo criterio. También se han obtenidos las distancias d_t , calculadas “a posteriori” para LM.

Se ha observado una gran asimetría en las muchas de las distribuciones estudiadas, tanta que los histogramas no resultan visualmente informativos, por lo que se ha preferido añadir a las tablas, ya empleadas de medias, mínimos, máximos y desviaciones típicas, los resultados de las medianas y de los percentiles 95 y 99. Se han incluido también las gráficas que muestran los comportamientos de medias, máximos y mínimos con respecto a valor φ al inicio del segmento.

3.5.2. Discusión de resultados

3.5.2.1. 181 puntos

Primero analizaremos los resultados correspondientes a las muestras de 181 puntos. Se realizaron 18100 estimaciones.

Error “grande” Comenzaremos por el valor “grande” del error unitario $\sigma = 0,02$, 20 cm. a 10 m., así los errores de los valores no-outliers estarían comprendidos en ± 60 cm. Los resultados se muestran en la tabla 3.3 a) y b).

En la estimación de ρ , tabla 3.3 a), el procedimiento propuesto obtiene un error medio del 0,8%, el mediano es 0,2% y el 99% de los errores es inferior al 9,4%. Sin embargo

a)	media	min.	mediana	per. 95	per. 99	máx.	st. dev.
$E_{rel}(\hat{\rho})$	0,00811	0,00000	0,00205	0,02990	0,09445	1,31106	0,03410
$E_{abs}(\hat{\theta})$	0,00922	0,00000	0,00212	0,02092	0,06425	3,03223	0,09326
$E_{rel}(\hat{\sigma})$	0,26427	0,00001	0,05461	0,79585	2,57985	185,253	2,64002
d_t	0,30060	0,06223	0,22163	0,79652	1,07219	2,33870	0,22211
b)	media	min.	mediana	per. 95	per. 99	máx.	st. dev.
$E_{rel}(\hat{\rho})$	0,12952	0,00000	0,00211	0,98451	0,99141	0,99400	0,32511
$E_{abs}(\hat{\theta})$	0,05715	0,00000	0,00181	0,43353	0,70883	4,50188	0,18126
$E_{rel}(\hat{\sigma})$	3,56167	0,00000	0,04314	24,1451	29,1715	231,056	9,89794
d_t	0,22206	0,02810	0,09996	1,05991	1,12011	1,24208	0,32114
c)	media	min.	mediana	per. 95	per. 99	máx.	st. dev.
$E_{rel}(\hat{\rho})$	0,00017	0,00000	0,00014	0,00042	0,00061	0,00151	0,00014
$E_{abs}(\hat{\theta})$	0,00019	0,00000	0,00014	0,00052	0,00075	0,00140	0,00017
$E_{rel}(\hat{\sigma})$	0,04562	0,00000	0,03762	0,11429	0,15544	0,58528	0,03642
d_t	0,20541	0,06429	0,19588	0,32535	0,40816	0,83698	0,06670
d)	media	min.	mediana	per. 95	per. 99	máx.	st. dev.
$E_{rel}(\hat{\rho})$	0,00553	0,00000	0,00014	0,00039	0,00054	0,95212	0,06461
$E_{abs}(\hat{\theta})$	0,00057	0,00000	0,00013	0,00051	0,00084	0,13322	0,00492
$E_{rel}(\hat{\sigma})$	1,63462	0,00000	0,03510	0,10675	0,16568	257,052	18,7024
d_t	0,09007	0,00340	0,07974	0,15221	0,22413	1,13913	0,08624

Tabla 3.3: Medidas de los errores cometidos al estimar 18100 rectas en a) mediante en procedimiento secuencial se propuesto en este trabajo, en b) la estimación directa por MCG, se estima usando todos los datos a la vez. En c) y d) sólo se usaron los primeros 25 puntos en la estimación.

aparecen valores extremos del error indicados por un máximo del 131,1 %, estos valores extremos aparecen en una probabilidad pequeña pero no desdeñable. En las figura 3.12 a) y b), puede verse cómo a partir de aproximadamente 60° empiezan a producirse máximos con valores de tendencia creciente, que a su vez inducen esa tendencia en las medias, haciéndolas variar de aproximadamente el 0,2 % al 2 %.

Comparamos a continuación estos resultados con los obtenidos mediante LM, tabla 3.3 b). Los resultados del procedimiento LM están muy condicionados numerosas estimaciones malogradas. Como puede verse observando los percentiles 95 y 99 con porcentajes de error cercanos al 100 %. En términos de medianas los resultados son muy similares a los del procedimiento propuesto. En las figura 3.12 c) y d), puede verse cómo a partir de aproximadamente 45° empiezan a producirse máximos que pasan bruscamente de aproximadamente 0,5 % al 100 %, esto induce una tendencia creciente en las medias que las hace alcanzar casi ese 100 % debido a la abrumadora presencia de máximos cuando el ángulo de comienzo del segmento se acerca a 0° . El estudio de caso $\sigma = 0,002$ arrojará luz sobre este particular. Si comparamos las medianas de ambos procedimientos encontraremos que son muy semejantes, y pequeñas, lo que lleva a pensar que en ausencia de errores extremos ambos procedimientos podrían tener un resultado similar entre sí y similar también al de caso lineal, comparar con medias para a en tabla 3.2 a) y b).

Los resultados para θ muestran una situación similar a la descrita para ρ , motivada por la presencia de máximos extremos. El procedimiento propuesto obtiene un error medio de 0,009 ($0,53^\circ$), el mediano es 0,0021 ($0,12^\circ$) y el 99 % de los errores es inferior al 0,064 ($3,7^\circ$), sin embargo aparecen valores extremos, como era de esperar. LM muestra un error medio del 0,057 ($3,2^\circ$), el mediano es 0,0018 ($0,10^\circ$), con el 99 % de errores por debajo de 0,7 ($40,6^\circ$) y la presencia de máximos extremos. En las figura 3.12 d) y e), puede verse cómo el procedimiento propuesto a partir de aproximadamente 60° , empieza a producir máximos con valores de tendencia creciente, que a su vez aumenta la cierta tendencia en las medias, haciéndolas variar de aproximadamente 0,002 ($0,11^\circ$) hasta 0,01 ($0,57^\circ$) aunque con la presencia de máximos. El mayor máximo de las medias, no se aprecia en la figura, fue de 0,12 ($6,9^\circ$). LM vuelve a mostrar los problemas comentados a partir de 45° con numerosos máximos extremos que llevan el error de la media hasta unos 0,35 (20°), figura 3.12 f) y g). Las medianas de ambos procedimientos son muy semejantes, pequeñas y aparentemente menores que en el caso lineal, comparar con medias para b en tabla 3.2 a) y b) aplicándoles la arcotangente, que para los valores referidos es $\arctan(b) \simeq b$.

En el caso de σ , el procedimiento propuesto obtiene un error medio del 26 %, el mediano es 5,4 % y el 99 % de los errores es inferior al 275,9 % y hay máximos extremos mucho mayores. En el caso de LM la situación es mucho peor, sólo ya la media tiene un error de 356,2 %. La figuras 3.13 a) y b) comparadas con c) y d) muestran cómo el aumento de la media en LM alcanza aproximadamente el valor 3000 % mientras que el en el procedimiento propuesto se alcanza el 50 % con picos de hasta el 250 %. El comportamiento antes del punto crítico del procedimiento propuesto, 60° , son muy similares para máximos y medias, estas últimas en la línea de las medianas obtenidas: 13,5 %, procedimiento propuesto, 10 % LM, similares a las medias obtenidas en los experimentos lineales, tabla 3.2 a) y b).

El comportamiento de las distancias d_i vuelve a mostrar todo lo comentado hasta ahora,

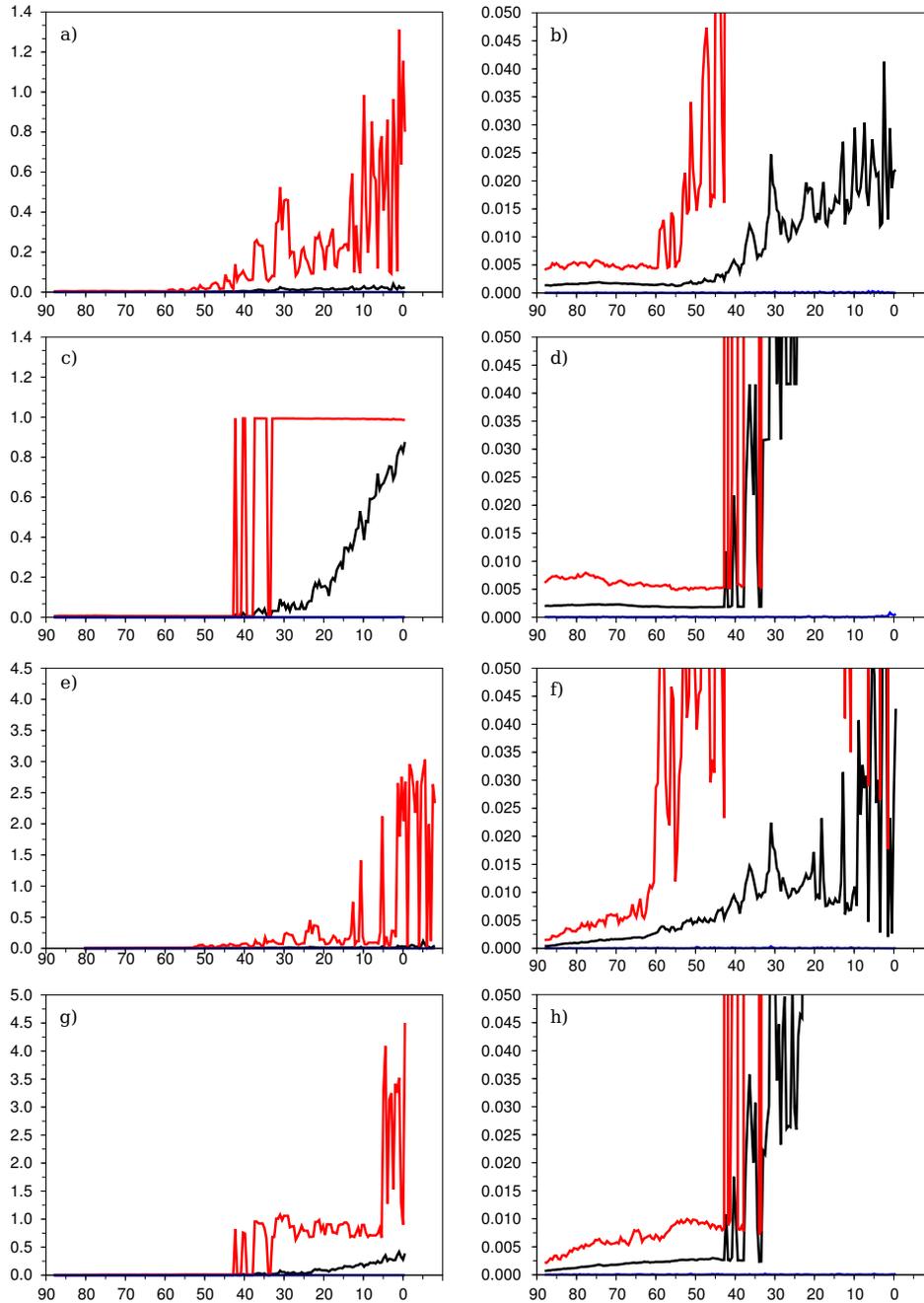


Figura 3.12: Comparación del procedimiento propuesto con LM. Errores con 181 puntos, con error de medición “grande”. Valores máximos (rojo), medios (negro) y mínimos (azul). Ver comentario en el texto.

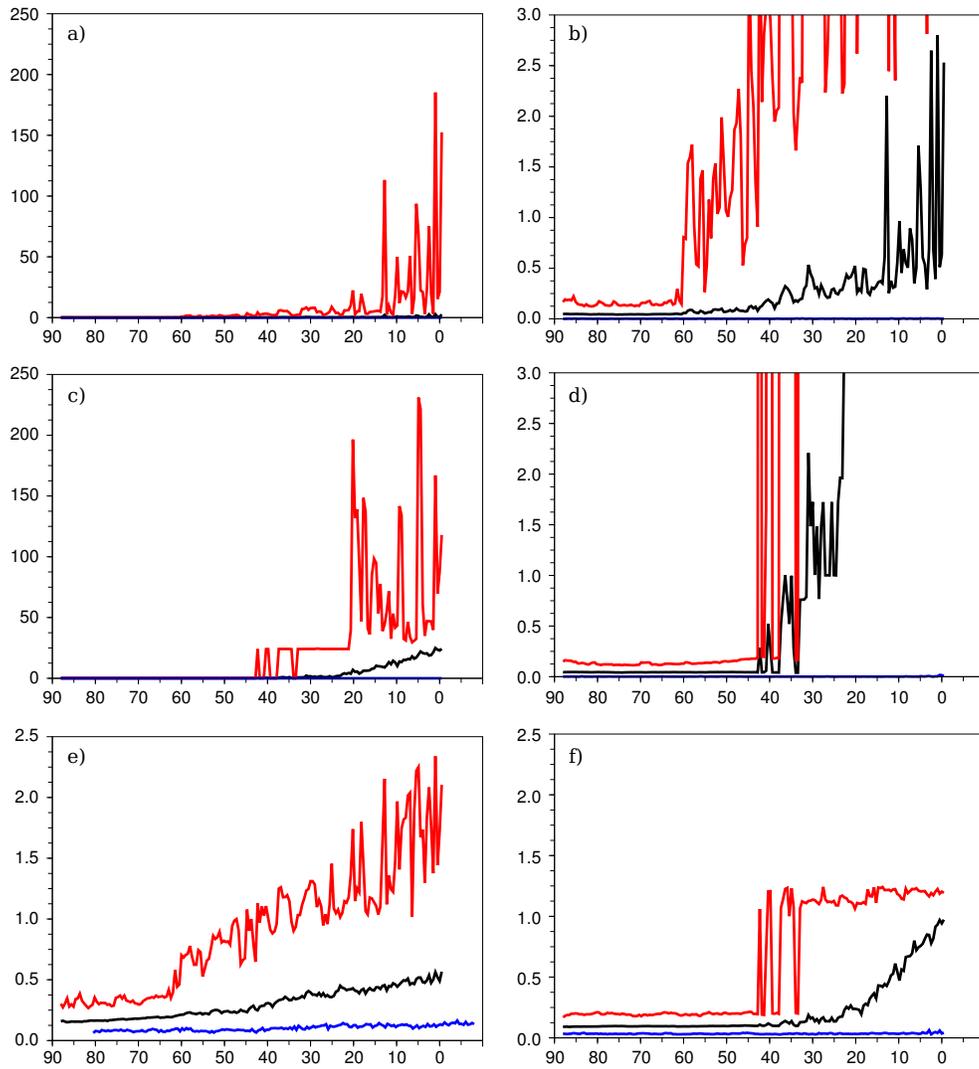


Figura 3.13: Comparación del procedimiento propuesto con LM. Error y d_t con 181 puntos, con error de medición “grande”. Valores máximos (rojo), medios (negro) y mínimos (azul). Ver comentario en el texto.

en esta ocasión los valores producidos por ambos procedimientos son más parecidos. Los valores medios y medianos obtenidos por LM son un poco menores como era de esperar ya que aquí las t_{exp} se calculan “a posteriori”. Los percentiles 95 y 99 son similares y aunque LM da un máximo menor, las gráficas de la 3.13 e) y f) cómo la media aumenta más en LM, lo que indica, como ya sabemos, un mayor número de máximos extremos, en este procedimiento.

En resumen ateniéndonos a la distribución de errores por debajo del percentil 99 para ρ y θ , el procedimiento propuesto alcanza unos resultados que pueden considerarse aceptables, cuando se escanea una superficie recta, pero apreciablemente irregular, abarcando un ángulo de casi 90° . Si nos atenemos a los errores de σ y a d_t convendría pensar mejor en el percentil 95, ya que por encima pudiera estar viéndose afectada la capacidad del método para detectar outliers “impropios”, ver subsección . En un porcentaje inferior al 1% aparecen resultados realmente malos. Si las medias fueran similares a las medianas, con la consiguiente disminución del número de valores grandes y medio grandes, el procedimiento propuesto conseguiría una precisión muy aceptable. En comparación con el procedimiento “off-line” es bastante menos susceptible a proporcionar estimaciones malas, las estimaciones donde los métodos no experimentan problemas son bastante similares.

Error “normal” Pasemos ahora al caso del valor “normal” del error unitario $\sigma = 0,002$, 2 cm. a 10 m. Comentemos en primer lugar que como se observa en las figuras 3.14 y 3.15, salvo las gráficas superiores que corresponden al procedimiento propuesto, el procedimiento LM experimenta problemas cuando el comienzo del segmento corresponde a aproximadamente 5° . Se observó que los problemas se presentaban en los últimos puntos de los segmentos y al suprimir los correspondientes a los 5° finales desaparecieron los problemas. Por esto comparamos LM y el procedimiento propuesto mediante la mediana y los percentiles 95 y 99 y en el caso de las gráficas obviamos el intervalo comprendido entre 5° y 0° , salvo para comparar los errores extremos.

La tabla 3.3 c) muestra los resultados del procedimiento propuesto y d) los de LM. Comparando medianas y percentiles de los errores de las estimaciones de ambos procedimientos se observa que son muy similares o prácticamente iguales, por lo que solo vamos a detallar los correspondientes al procedimiento propuesto.

En la estimación de ρ se obtiene un error mediano es 0,014% y el 99% de los errores es inferior al 0,054%. El error medio 0,017% es prácticamente igual al mediano, indicio de inexistencia de valores anómalos, lo que se confirma al observar un error máximo de 0,15%, y es aquí donde aparece la diferencia con LM que alcanza un error máximo de 95,2%. La figura 3.14 a), b) y c) muestra gráficamente la situación aquí comentada, el error fluctúa ligeramente con máximos cerca de los extremos del segmento y un mínimo en la zona central.

Los resultados para θ muestran, en el procedimiento propuesto, un error mediano de 0,00014 ($0,008^\circ$) y el 99% de los errores es inferior al 0,00061 ($0,035^\circ$). El error medio 0,00017 ($0,01^\circ$) es prácticamente igual al mediano, el error máximo de 0,0014 ($0,08^\circ$) mientras que en LM que alcanza un valor de 0,133 ($7,6^\circ$). La figura 3.14 d), e) y f) muestra

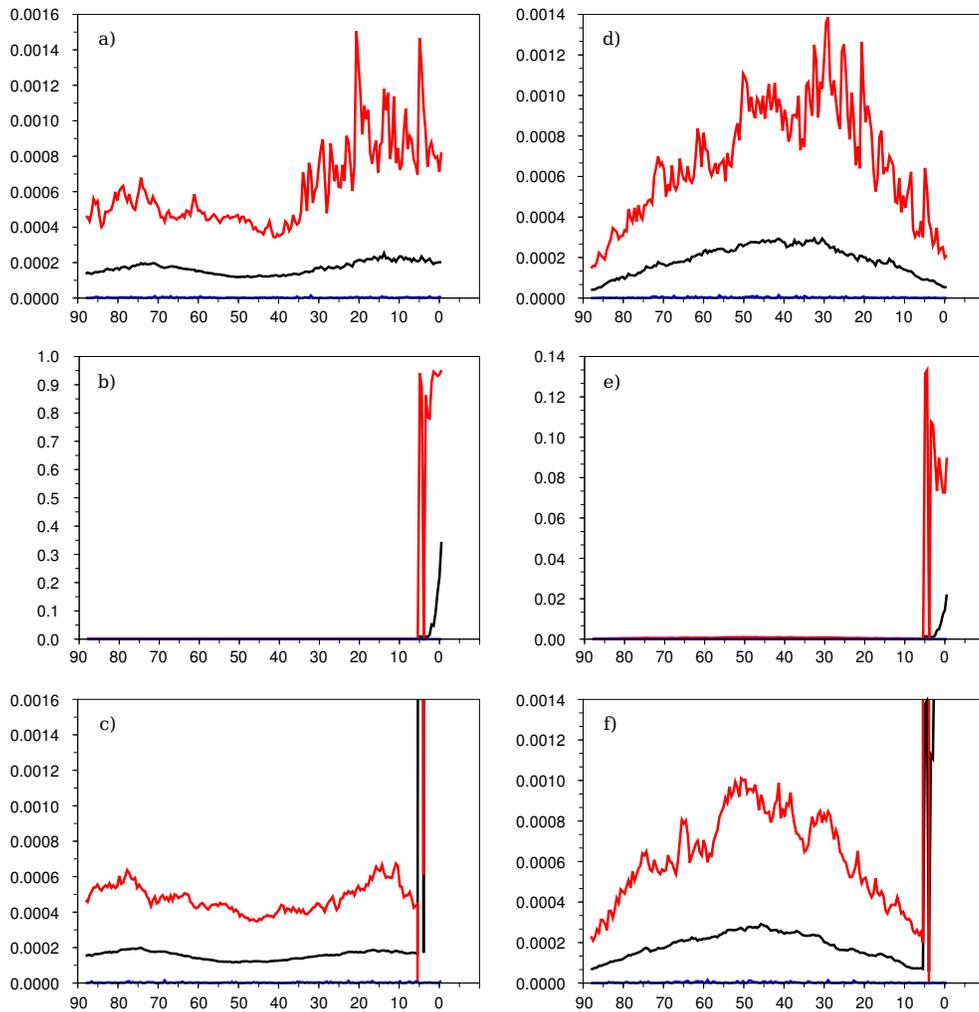


Figura 3.14: Comparación del procedimiento propuesto con LM. Errores con 181 puntos, con error de medición “normal”. Valores máximos (rojo), medios (negro) y mínimos (azul). Ver comentario en el texto.

gráficamente la situación aquí comentada, es curioso observar que los máximos errores de θ se alcanzan cuando el segmento comienza aproximadamente a 45° .

Para σ el procedimiento propuesto comete un error mediano de 3,7% y el percentil 99 de los errores se sitúa 15,5%. Vuelve la media, 4,6%, a ser muy semejante a la mediana. El máximo se sitúa en el 58,5% y para LM es el 25705,2%. La figura 3.15 a), b) y c) muestra gráficamente estos hechos, es de constatar que el máximo error obtenido por el procedimiento propuesto se produce de forma aislada, y en la zona inadecuada para LM, los otros máximos son del orden del 30%. En general se cumple el requisito de error relativo inferior al 12%, comentado al principio de la sección 3.4, aunque dado que estamos en un modelo no lineal el sentido de ese valor puede quedar un tanto desdibujado.

Los valores de las d_t son para el procedimiento propuesto del orden del doble que para LM, que las calcula “a posteriori”. La figura 3.15 d) y e) se observa una ligera tendencia creciente en el procedimiento propuesto frente a la relativa constancia de LM, mientras no se llegue a la zona final.

Los resultados obtenidos cuando se escanea una superficie recta, razonablemente pulida, abarcando un ángulo de casi 90° mediante el procedimiento propuesto se consigue una exactitud muy apreciable. Para ρ y θ es del orden de las centésimas, en porcentaje y grados respectivamente, y como mucho poco más de la décima. Para σ , teniendo en cuenta que admitimos valores mayores, también nos movemos en valores bastante aceptables. Los valores de las d_t nos indican que en algunos casos podría haber problemas en la detección de outliers, no demasiado anómalos.

3.5.2.2. 35 puntos

Se realizaron 32600 estimaciones de rectas.

Error “grande” Comenzaremos por el valor “grande” del error unitario $\sigma = 0,02$, 20 cm. a 10 m. Los resultados se muestran en la tabla 3.4 a) y b). Se ha preferido en las figuras no mostrar todo el rango de variación de los errores, en especial para LM, y sólo mostrar un ranro de valores no muy anómalos.

En la estimación de ρ , tabla 3.4 a), el procedimiento propuesto comete un error medio del 2,1%, el mediano es 1,5% y el 99% de los errores es inferior al 19,2%. Sin embargo aparecen valores extremos del error indicados por un máximo del 138,3%, estos valores extremos aparecen en una probabilidad pequeña pero no desdeñable. En la figura 3.16 a), puede verse cómo a partir de aproximadamente -60° y hasta 60° , en consonancia con lo que se observó en el experimento anterior para este valor de σ , empiezan a producirse máximos con valores de tendencia creciente, que a su vez inducen ligeramente esa tendencia en las medias, se observan unos máximos en las medias aproximadamente a $\pm 45^\circ$ y un mínimo en 0.

Comparamos a continuación estos resultados con los obtenidos mediante LM, tabla 3.4 b). LM experimenta problemas a partir de aproximadamente 45° hasta el final, de modo similar a como en el experimento anterior, que malogran muchas estimaciones. En la figura 3.16 b) puede verse como aumenta la media del error a partir de ese punto.

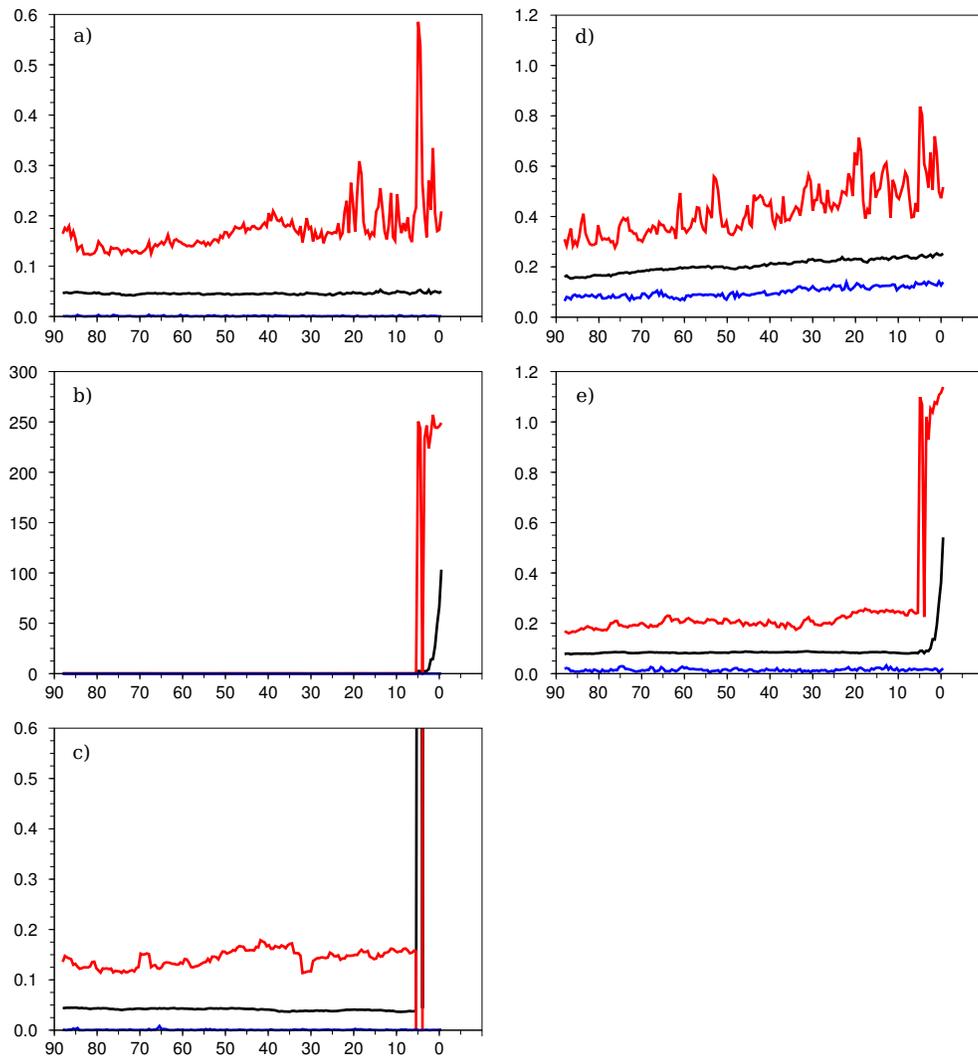


Figura 3.15: Comparación del procedimiento propuesto con LM. Error y d_t con 181 puntos, con error de medición “normal”. Valores máximos (rojo), medios (negro) y mínimos (azul). Ver comentario en el texto.

a)	media	min.	mediana	per. 95	per. 99	máx.	st. dev.
$E_{rel}(\hat{\rho})$	0,02069	0,00000	0,01466	0,09251	0,19276	1,38262	0,04205
$E_{abs}(\hat{\theta})$	0,05234	0,00000	0,02014	0,20599	0,43020	1,25845	0,09154
$E_{rel}(\hat{\sigma})$	0,25437	0,00000	0,13564	0,96315	2,09921	4,72462	0,39431
d_t	0,50940	0,07833	0,42327	1,14125	1,54037	2,46308	0,29948

b)	media	min.	mediana	per. 95	per. 99	máx.	st. dev.
$E_{rel}(\hat{\rho})$	0,02488	0,00000	0,01323	0,05107	0,07855	22,1969	0,18753
$E_{abs}(\hat{\theta})$	0,03436	0,00000	0,01749	0,09145	0,13911	16,0503	0,19949
$E_{rel}(\hat{\sigma})$	1,22840	0,00001	0,09889	0,29874	0,44743	21092,4	120,997
d_t	0,13639	0,01207	0,12513	0,25128	0,35698	0,842067	0,07264

c)	media	min.	mediana	per. 95	per. 99	máx.	st. dev.
$E_{rel}(\hat{\rho})$	0,00133	0,00000	0,00099	0,00370	0,00542	0,02262	0,00125
$E_{abs}(\hat{\theta})$	0,00228	0,00000	0,00128	0,00781	0,01270	0,04735	0,00287
$E_{rel}(\hat{\sigma})$	0,12264	0,00000	0,10365	0,31873	0,43432	1,40147	0,10100
d_t	0,39450	0,08180	0,36389	0,70837	0,90157	1,75043	0,163875

d)	media	min.	mediana	per. 95	per. 99	máx.	st. dev.
$E_{rel}(\hat{\rho})$	0,00171	0,00000	0,00128	0,00484	0,00671	0,01195	0,00152
$E_{abs}(\hat{\theta})$	0,00273	0,00000	0,00169	0,00887	0,01291	0,02262	0,00292
$E_{rel}(\hat{\sigma})$	0,11383	0,00001	0,09654	0,28065	0,36405	0,55887	0,08560
d_t	0,130020	0,00743	0,12247	0,24111	0,29845	0,48287	0,06004

Tabla 3.4: Medidas de los errores cometidos al estimar 18100 rectas en a) mediante en procedimiento secuencial se propuesto en este trabajo, en b) la estimación directa por MCG, se estima usando todos los datos a la vez. En c) y d) sólo se usaron los primeros 26 puntos en la estimación.

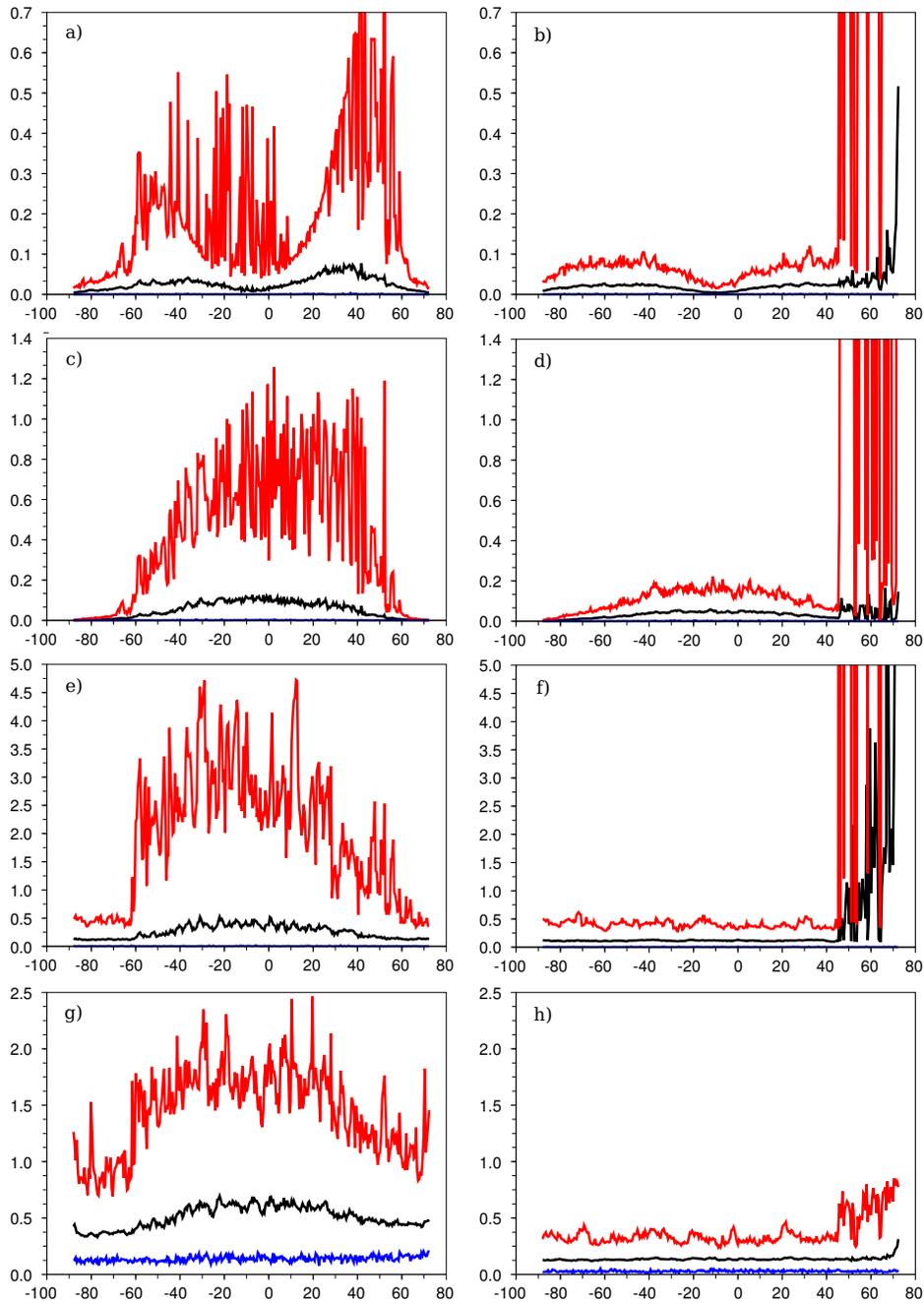


Figura 3.16: Comparación del procedimiento propuesto con LM. Errores con 35 puntos, con error de medición “grande”. Valores máximos (rojo), medios (negro) y mínimos (azul). Ver comentario en el texto.

En términos de medias y medianas los resultados son muy similares a los del procedimiento propuesto aunque en este caso el percentil 99 es 7,9%, sin embargo aparecen máximos más grandes en LM. En la figura 3.16 b) puede verse como si se prescinde de esos valores extremos los resultados de LM parecen ligeramente mejores que los del procedimiento propuesto, sobre todo en la distribución de máximos aunque estos no son muy numerosos ya que no alteran demasiado la distribución de medias en el procedimiento propuesto. La medias de ambos procedimientos son similares al caso lineal, comparar con medias para a en tabla 3.2 c) y d).

Los resultados para θ muestran una situación similar a la descrita para ρ , motivada por la presencia de máximos extremos. El procedimiento propuesto se obtiene un error medio de 0,052 (3°), el mediano es 0,020 ($1,2^\circ$) y el 99% de los errores es inferior al 0,43 ($24,6^\circ$), sin embargo aparecen valores extremos, como era de esperar. LM muestra un error medio del 0,034 (2°), el mediano es 0,0017 ($0,10^\circ$), con el 99% de errores por debajo de 0,14 (8°) y la presencia de máximos extremos. En la figura 3.16 c), puede verse cómo el procedimiento propuesto a partir de aproximadamente 60° , empieza a producir máximos con valores de tendencia creciente hasta llegar a 0° donde comienza a decrecer. LM vuelve a mostrar los problemas comentados a partir de 45° con numerosos máximos extremos que llevan la media del error hasta unos 0,35 (20°), figura 3.12 d). En la figura 3.16 d) puede verse como si se prescinde de esos valores extremos los resultados de LM parecen ligeramente mejores que los del procedimiento propuesto, sobre todo en la distribución de máximos aunque estos no son muy numerosos ya que no alteran demasiado la distribución de medias en el procedimiento propuesto con respecto a LM. Las medias de ambos procedimientos son aparentemente mejores que en el caso lineal, tabla 3.2 c) y d)⁴.

El comparar la figura 3.16 a), b), c) y d) con las de la figura 5.2 a) y b) resulta evidente que el comportamiento del error de ρ y θ es consecuencia directa del comportamiento del error de los valores iniciales ρ_0 y θ_0 proporcionados por la semilla.

En el caso de σ el procedimiento propuesto obtiene un error medio del 25,4%, el mediano es 13,6% y el 99% de los errores es inferior al 209,9% y hay máximos extremos mayores. En el caso de LM la situación es peor debido a los grandes errores finales, sólo ya la media tiene un error de 122,8% y hay máximos muy extremos. En términos de percentiles LM muestra un comportamiento mejor ya que el 99% de los errores es inferior al 44,7%, esto también puede compararse las figuras 3.17 e) y f), fuera de su zona problemática LM parece que ofrece un rendimiento mejor que el método propuesto, en especial en términos de máximos. Las medianas obtenidas mediante ambos procedimientos son similares a las medias obtenidas en los experimentos lineales, tabla 3.2 c) y d).

El comportamiento de las distancias d_t vuelve a mostrar todo lo comentado hasta ahora, aunque los valores obtenidos por LM son menores, como era de esperar ya que aquí las t_{exp} se calculan “a posteriori”. El método propuesto acusa la presencia de máximos, las gráficas de la 3.16 g) y h) muestran la situación.

En resumen el procedimiento propuesto se ve afectado por un problemas de errores relativamente grandes con una frecuencia aproximada del 5%, cuando se escanea una

⁴ $\arctan(b) \simeq b$ si $|b| < 0,5$

superficie recta, pero apreciablemente irregular, abarcando un ángulo de 17° . En términos de medias y medianas el procedimiento propuesto consigue una precisión muy aceptable. En comparación con el procedimiento “off-line” muestra unos errores grandes con más frecuencia que el procedimiento “off-line” pero estos errores no son tan extremos.

Error “normal” Pasemos ahora al caso del valor “normal” del error unitario $\sigma = 0,002$, 2 cm. a 10 m. Comentemos en primer lugar que como se observa en las figura 3.17 b), d), f) y h), LM ya no presenta problemas de errores extremos.

La tabla 3.4 c) muestra los resultados del procedimiento propuesto y d) los de LM. Comparando medias medianas y percentiles de los errores de las estimaciones de ambos procedimientos se observa que son muy similares o prácticamente iguales. Al observar en la figura 3.17 vemos que la única diferencia en las estimaciones se da en la aparición de algunos máximos algo mayores que los de LM en los resultados del procedimiento propuesto. Dada esta igualdad solo vamos a detallar los resultados correspondientes al procedimiento propuesto.

En la estimación de ρ se obtiene un error medio 0,13 %, un error mediano de 0,10 % y el 99 % de los errores es inferior al 0,54 %. La figura 3.17 a) y b), para LM, muestra gráficamente la situación aquí comentada, el error fluctúa ligeramente con máximos cerca de los extremos del segmento y un mínimo en la zona central.

Los resultados para θ muestran, en el procedimiento propuesto, un error medio 0,0023 ($0,13^\circ$), un error mediano de 0,0013 ($0,07^\circ$) y el 99 % de los errores es inferior al 0,013 ($0,7^\circ$). La figura 3.14 c), y d) muestra gráficamente la situación aquí comentada.

Para σ el procedimiento propuesto comete un error medio del 12,2 %, un error mediano de 10,4 % y el percentil 99 de los errores se sitúa 43,4 % . El máximo se sitúa en el 140,1 % aunque estos valores grandes son raros y el comportamiento del error es similar al de LM, como puede verse en la figura 3.15 e) y f). El requisito de error relativo inferior al 29 %, ver sección 3.4, es cumplido aproximadamente por el percentil 95.

Los valores de las d_t muestran la diferencia entre un procedimiento “off-line” con unas d_t calculadas “a posteriori” cuando no hay muchos puntos en el segmento y el procedimiento “on-line”. Basta observar la figura 3.15 e) y f) y ver cómo los valores medios del procedimiento propuesto son semejantes a los máximos de LM, esto puede estar comprometiendo la capacidad del procedimiento propuesto para detectar outliers. Todo dependerá de cuan anómalo sea el valor, postergamos esta cuestión hasta que se defina qué clase de outliers se desea encontrar en el capítulo 5.

Los resultados obtenidos cuando se escanea una superficie recta, razonablemente pulida, abarcando un ángulo de 17° mediante el procedimiento propuesto se consigue una exactitud muy apreciable. Para ρ y θ es del orden de las décimas, en porcentaje y grados respectivamente, y como mucho algo superior a la unidad. Para σ , teniendo en cuenta que admitimos valores mayores y la no linealidad del modelo, también nos movemos en valores bastante aceptables en una gran mayoría de casos. Los valores de las d_t nos indican que podría haber problemas en la detección de outliers, no demasiado anómalos.

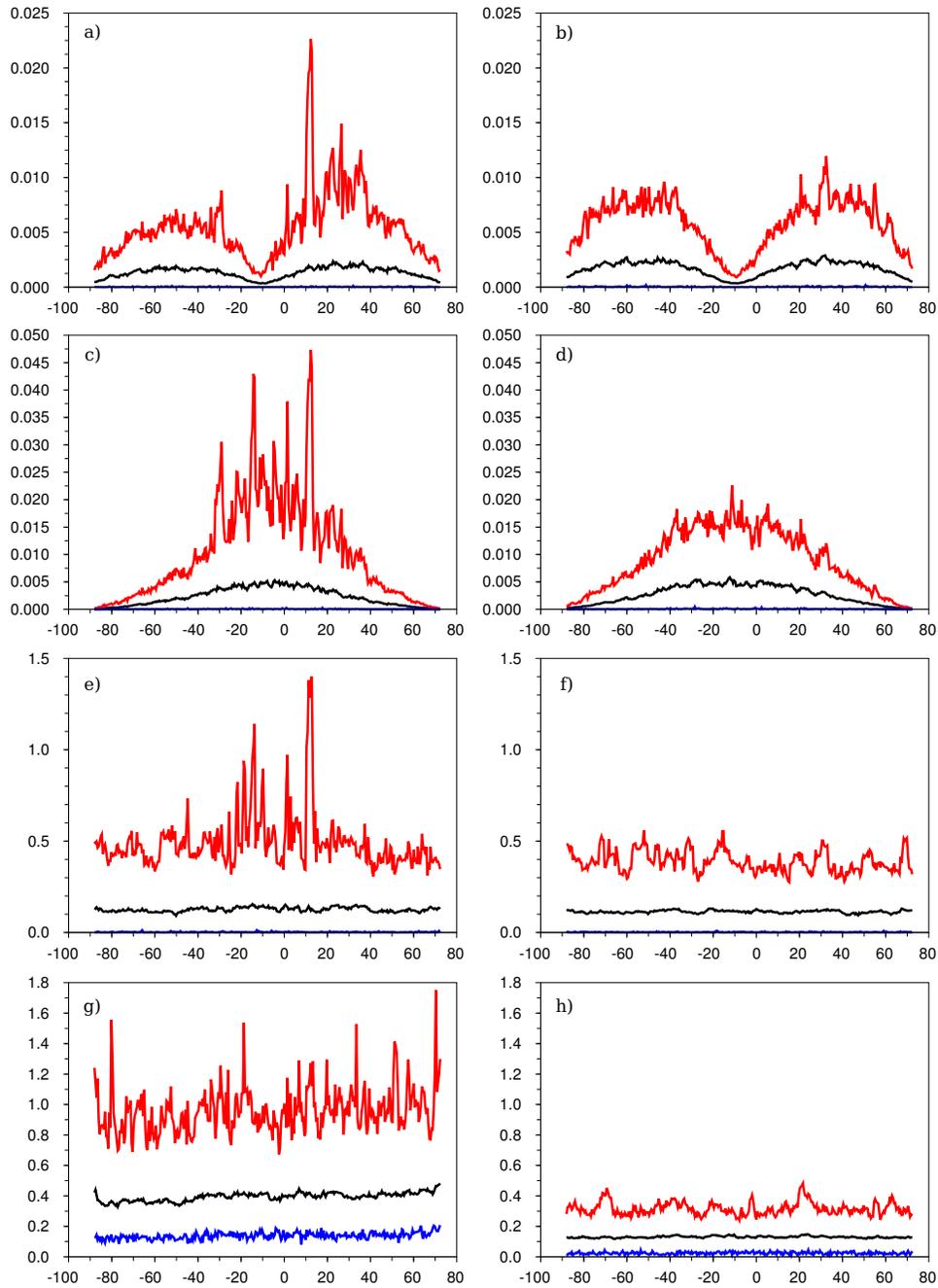


Figura 3.17: Comparación del procedimiento propuesto con LM. Errores con 35 puntos, con error de medición “normal”. Valores máximos (rojo), medios (negro) y mínimos (azul). Ver comentario en el texto.

a)	media	min.	mediana	per. 95	per. 99	máx.	st. dev.
$E_{rel}(\hat{\rho})$	0,02245	0,00000	0,01264	0,07661	0,16253	0,57597	0,03212
$E_{abs}(\hat{\theta})$	0,04784	0,00000	0,01726	0,20297	0,34103	0,69837	0,07336
$E_{rel}(\hat{\sigma})$	0,26877	0,00000	0,15082	0,98377	1,81953	4,41193	0,36331
d_t	0,53888	0,06836	0,46717	1,10323	1,41893	1,66819	0,15808

b)	media	min.	mediana	per. 95	per. 99	máx.	st. dev.
$E_{rel}(\hat{\rho})$	0,00115	0,00000	0,00087	0,00320	0,00443	0,00820	0,00101
$E_{abs}(\hat{\theta})$	0,00183	0,00000	0,00109	0,00603	0,00897	0,02100	0,00204
$E_{rel}(\hat{\sigma})$	0,13201	0,00001	0,10791	0,33461	0,48425	1,52994	0,10952
d_t	0,40589	0,09297	0,37917	0,70564	0,89220	0,842067	0,07264

Tabla 3.5: Medidas de los errores cometidos al situar la semilla aproximadamente en el centro del segmento. Se realizaron 32600 estimaciones de rectas de 35 puntos, 9 de semilla, rectas en a) con $\sigma = 0,02$ y en b) con $\sigma = 0,002$

3.5.3. El procedimiento bilateral

Hasta ahora la semilla se ha situado al comienzo del segmento pero esa información sobre la posición de la semilla no estará disponible en los casos reales, por lo que se debería poder realizar la introducción de puntos sucesivamente a cada lado de la semilla hasta que se de una condición, basada en outliers, que permita decidir si ha llegado a cada extremo del segmento dentro del scan completo del láser. Del problema de la detección de los extremos del segmento nos ocuparemos en el capítulo 5, por ahora basta saber que está basada en la detección de outliers. El introducir los puntos sucesivamente a cada lado de la semilla no obliga a replantear el procedimiento propuesto, ya que se siguen introduciendo puntos equiespaciados. Así se cumple la condición comentada al principio de la sección 3.3, de que los valores de la variable independiente estén equiespaciados, por otro lado no hay nada que obligue a que el nuevo punto que se introduce en una regresión tenga que ser mayor o menor que los que ya han sido estimados.

Para comprobar el rendimiento de esta variación del método se repitieron los experimentos con 35 puntos realizados en la subsección anterior, pero colocando la semilla en el centro del segmento, los resultados pueden verse en la tabla 3.5. La única diferencia apreciable con respecto a la tabla 3.4 a) y c) es una reducción de aproximadamente un 50 % en los valores máximos del error en la estimación de los parámetros ρ y θ para ambos valores de σ , en el caso de $\sigma = 0,002$ estos máximos son muy similares a los obtenidos por LM, ver tabla 3.4 d). Esta reducción del error máximo hace preferible que la semilla esté aproximadamente centrada en el segmento, antes que en un extremo, y muestra el carácter equilibrador del procedimiento bilateral.

3.6. Demostraciones

3.6.1. MCP mediante filtro de Kalman

En las subsección 2.4.1.2 se comentó el procedimiento en dos etapas para la estimación MCP, una primera etapa por MCO y una segunda etapa en la que se ponderan los datos y a partir de los datos transformados se vuelve a realizar MCO. Sin embargo el procedimiento expuesto en la subsección 3.2.3 difiere de lo comentado en que no transforma los datos sino que sólo cambia la expresión del elemento a_i del algoritmo 3.2, ganando en eficiencia. Se va a probar que las segundas etapas de ambos procedimientos son equivalentes y dado que para el primero, ya que emplea sólo MCO, se conocen las distribuciones estadísticas necesarias para la construcción del test de detección de outliers, se podrá definir dicho test en los términos del segundo procedimiento. Al primer procedimiento lo denominaremos modelo transformado y al segundo modelo no-transformado.

3.6.1.1. Relaciones entre estimaciones

Primero se analizan las relaciones entre los elementos que componen el algoritmo del filtro de Kalman cuando se realiza la segunda etapa. La segunda etapa del modelo transformado, usamos un asterísco como superíndice, se estima por MCO, ver subsección 2.4.1.2, por lo que se le aplica al algoritmo 3.2. La segunda etapa del modelo no-transformado, usamos una p como superíndice, se estima según la segunda mitad del algoritmo 3.3.

Las relaciones se demuestran por inducción, para $n = 1$ se tendrá

$$\left. \begin{aligned} a_1^* &= 1 + X_1^* \Sigma_0 X_1 = 1 + (w_1 X_1)' \Sigma_0 (w_1 X_1) = 1 + w_1^2 X_1' \Sigma_0 X_1 \\ a_1^p &= \frac{1}{w_1^2} + X_1' \Sigma_0 X_1 \end{aligned} \right\} \implies a_1^* = w_1^2 a_1^p \quad (3.16)$$

$$\left. \begin{aligned} K_1^* &= \Sigma_0 \frac{X_1^*}{a_1^*} = \Sigma_0 \frac{w_1 X_1}{a_1^*} = \Sigma_0 \frac{X_1}{w_1 a_1^p} \\ K_1^p &= \Sigma_0 \frac{X_1}{a_1^p} \end{aligned} \right\} \implies K_1^* = \frac{K_1^p}{w_1} \quad (3.17)$$

$$\left. \begin{aligned} \Sigma_1^* &= \Sigma_0 - K_1^* X_1^* \Sigma_0 = \Sigma_0 - K_1^* (w_1 X_1) \Sigma_0 = \Sigma_0 - K_1^p X_1 \Sigma_0 \\ \Sigma_1^p &= \Sigma_0 - K_1^p X_1 \Sigma_0 \end{aligned} \right\} \implies \Sigma_1^* = \Sigma_1^p \quad (3.18)$$

$$\left. \begin{aligned} e_1^* &= y_1^* - X_1^* \beta_0 = w_1 y_1 - (w_1 X_1)' \beta_0 = w_1 (y_1 - X_1' \beta_0) \\ e_1^p &= y - X' \beta_0 \end{aligned} \right\} \implies e_1^* = w_1 e_1^p \quad (3.19)$$

$$\left. \begin{aligned} \beta_1^* &= \beta_0 + K_1^* e_1^* = \beta_0 + K_1^p e_1^p \\ \beta_1^p &= \beta_0 + K_1^p e_1^p \end{aligned} \right\} \implies \beta_1^* = \beta_1^p \quad (3.20)$$

Por lo que tras la introducción de la primera observación los resultados obtenidos para los estimadores y las estimaciones de sus matrices de varianzas son los mismos para ambos métodos. Supongamos las relaciones obtenidas en las ecuaciones (3.16) a (3.20) son válidas

para la observación n -ésima. Demostrarlas para la observación $n+1$ es trivial, basta cambiar el subíndice 0 por n y el 1 por $n+1$ en las demostraciones anteriores dado que, por hipótesis de inducción, los valores de los estimadores y las estimaciones obtenidas en observación n -ésima son los mismos para ambos métodos, como ocurre en la primera observación con los valores iniciales. Luego podemos afirmar que $\forall n$ se verifica

$$\begin{aligned} a_n^* &= w_n^2 a_n^p \\ K_n^* &= \frac{K_n^p}{w_n} \\ e_n^* &= w_n e_n^p \\ \beta_n^* &= \beta_n^p \\ \Sigma_n^* &= \Sigma_n^p \end{aligned} \tag{3.21}$$

o sea, las estimaciones producidas por ambos métodos son las mismas.

3.6.1.2. Test de detección de outliers

Las SCR de los modelos transformado y no-transformado deben coincidir, ya que las estimaciones producidas por ambos métodos son iguales. La expresión SCR del modelo transformado, que se calcula usando los estimadores obtenidos con la n -ésima observación, será

$$SCR_n^* = \sum_{i=1}^n \left(\overbrace{y_i^* - X_i^{*T} \beta_n^*}^{e_{i/n}^*} \right)^2$$

y aplicando la ecuación (3.6) queda

$$SCR_n^* = SCR_{n-1}^* + \frac{(e_n^*)^2}{a_n^*} = SCR_{n-1}^* + \frac{(e_n^p)^2}{a_n^p}$$

con $SCR_0^* = 0$. Así si en el modelo no-transformado definimos

$$SCR_n^p = SCR_{n-1}^p + \frac{(e_n^p)^2}{a_n^p} \tag{3.22}$$

con $SCR_0^p = 0$, se tendrá, al ser los valores iniciales $SCR_0^* = SCR_0^p$, que

$$SCR_n^* = SCR_n^p \quad \forall n$$

y por lo tanto la ecuación (3.22) da la forma de calcular secuencialmente la SCR del modelo no-transformado.

Las estimaciones de las varianzas serían equivalentes

$$(\hat{\sigma}_n^*)^2 = \frac{SCR_n^*}{n-k} = \frac{SCR_n^p}{n-k} = (\hat{\sigma}_n^p)^2$$

El estadístico del test de detección de outliers usando el modelo transformado, teniendo en cuenta las relaciones de la ecuación 3.21, sería

$$t_{exp} = \frac{e_{n/n-1}^*}{\hat{\sigma}_{n-1}^* \sqrt{a_n^*}} = \frac{e_{n/n-1}^p}{\hat{\sigma}_{n-1}^p \sqrt{a_n^p}} \sim t(n-1-k)$$

quedando el estadístico expresado en términos del modelo no-transformado.

3.6.1.3. Normalización de ganancia

El cálculo de los estimadores β_n^p y Σ_n^p no se ve modificado por la presencia del factor ω , que se puede obviar englobándolo en la expresión de w_n . Es a la hora de estimar la varianza de las perturbaciones donde se nota su efecto, como veremos a continuación. Dado que el cálculo de la SCR es el mismo en el modelo transformado que en el no-transformado, hacemos el estudio en el transformado, donde resulta más cómodo el razonamiento.

Como se vió en la subsección 2.4.1.2 la varianza del término de perturbación es $\text{var}[u_n] = \sigma_n^2 = \sigma^2/w_n^2$. Sin escalamiento el valor de u_n se pondera mediante w_n , quedando la varianza $\text{var}[u_n^*] = \text{var}[w_n u_n] = w_n^2 \sigma_n^2 = \sigma^2$. Con el escalamiento se tendría $\text{var}[u_n^*] = \text{var}[\omega w_n u_n] = \omega^2 w_n^2 \sigma_n^2 = \omega^2 \sigma^2 = \sigma_\omega^2$, luego la ecuación (3.9) estima en realidad σ_ω^2 , la verdadera estimación de la varianza de las perturbaciones será

$$(\hat{\sigma}_n^*)^2 = (\hat{\sigma}_n^p)^2 = \frac{\hat{\sigma}_\omega^2}{\omega^2}$$

Esto también puede verse, en el modelo transformado, observando que al multiplicar los valores de las observaciones por ω , los errores resultan también multiplicados por ω y por lo tanto la SCR queda multiplicada por ω^2 , de ahí que sea necesario dividir por esta cantidad para obtener la verdadera SCR y el verdadero estimador de la varianza de las perturbaciones. Esto nos lleva a redefinir la ecuación (3.8) como

$$SCR_n^p = SCR_{n-1}^p + \frac{(e_n^p)^2}{\omega^2 a_n^p} \quad (3.23)$$

usando esta expresión ya no es necesario tener en cuenta el factor de escala al estimar de la varianza de las perturbaciones.

Capítulo 4

Clustering mediante scale-space

Encontrar semillas en posiciones adecuadas es crítico para obtener una buena estimación mediante los filtros de Kalman descritos en el capítulo 3. En los experimentos dicho capítulo se partía siempre de semillas pertenecientes a un segmento determinado, pero en un caso real habrá varios segmentos y se deberían conseguir semillas para todos y cada uno de ellos. Si se tienen semillas a caballo entre dos segmentos conducirán a estimaciones de segmentos inexistentes. Si algún procedimiento es capaz de dar de forma aproximada las posiciones iniciales y finales de los posibles segmentos, se podrían encontrar zonas “seguras” para la elección de las semillas. La zona que consideramos más segura para la elección de la semilla es la zona central del posible segmento.

Para encontrar buenas semillas es necesario un procedimiento de clustering que proporcione de forma aproximada las zonas seguras. Para esto se va adaptar el procedimiento scale-space para aproximadamente detectar los bordes de los segmentos que componen el scan de láser en su representación en forma polar. Se ha adoptado esta técnica en un intento que desarrollar un método en el que no existan umbrales críticos que el usuario deba elegir *ad hoc*, como distancia máxima entre dos puntos para considerar que pertenecen a un mismo segmento o nivel de resolución a grano grueso al que se consideran las características distintivas. Esto no quita que en el método no exista algún parámetro ajustable pero su efecto no es tan crítico como en otros métodos.

La representación del mundo en coordenadas polares, ver figura 4.1, tiene ventajas sobre las rectangulares a la hora de la aplicación de técnicas de scale-space. En coordenadas polares a cada ángulo φ le corresponde a lo más un valor r , o sea, la descripción del mundo como $r = f(\varphi)$, se realiza mediante una función, en el estricto sentido matemático del término. Sin embargo en coordenadas rectangulares a cada x le pueden corresponder varios y , o viceversa, por lo que en la descripción $y = g(x)$ la correspondencia g no tiene porqué ser una función en sentido matemático estricto. Esto implica que en coordenadas polares se puede realizar una ordenación total de los puntos según el ángulo y por lo tanto la descripción $r = f(\varphi)$ es unidimensional, mientras que $y = g(x)$ no es estrictamente unidimensional. Por un lado la complejidad computacional al tratar una señal 1D siempre será menor que la de una 2D y por otro hay propiedades muy deseables que cumplen las señales 1D dentro del contexto de scale-space que no siempre se verifican en dimensiones

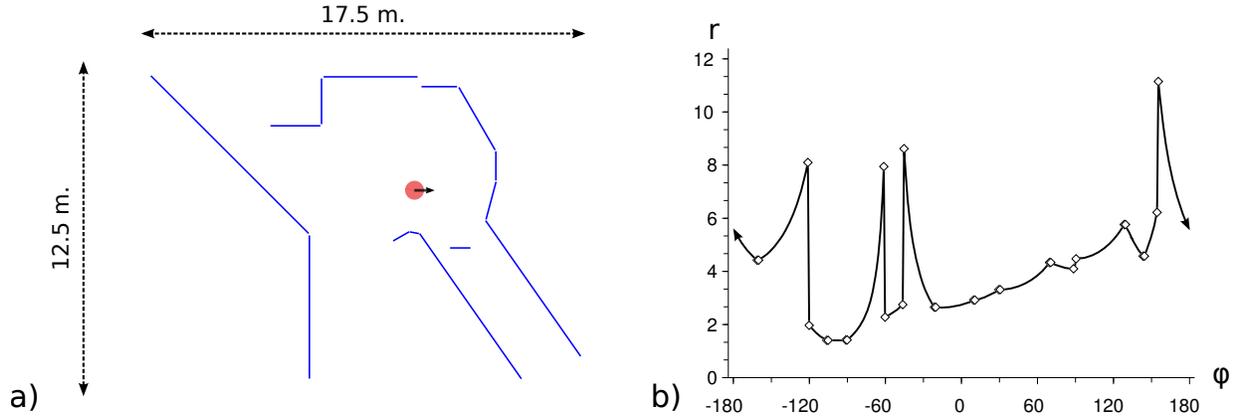


Figura 4.1: Mundo en coordenadas rectangulares a) y polares b). El círculo rojo representa al robot, que toma medidas en 360° con resolución angular de 1, la flecha situada sobre él indica la dirección $\varphi = 0$. Los rombos indican los extremos de los segmentos, la puntas de flecha indican que los extremos de la curva se unen circularmente.

superiores.

En este trabajo se ha seguido la orientación de Witkin (1984) para la obtención de características estables, junto con las aportaciones de Lindeberg para el tratamiento de señales discretas, adaptándola al tipo de señal que nos ocupa. Por lo que vamos primero a estudiar las características de las medidas de rango en su representación en forma polar en relación con la técnica de Witkins. Para ello estudiaremos una señal ideal, sin ruido, y a continuación estudiaremos el efecto del ruido y la técnica desarrollada para filtrarlo. A continuación se explicará la técnica de árbol de intervalos para obtener las características estables, aquellas que persisten a lo largo de un rango más amplio de escalas.

4.1. Curvas de nivel

La técnica de scale-space, para señales 1D $r = f(\varphi)$, consiste en convolucionar de la señal con núcleos gaussianos de distinta varianza σ_s^2 para generar una superficie 2D $F(\varphi, \sigma_s^2)$, por el momento consideraremos el parámetro de escala σ_s^2 y al ángulo φ como continuos. En esta superficie las separaciones entre características de la señal, bordes de segmentos en nuestro caso, aparecen como ceros de las primeras derivadas de F , $\frac{\partial^k F}{\partial \varphi^k}$, máximos, mínimos y puntos de inflexión. Si observamos la figura 4.1 b) veremos que los bordes de los segmentos aparecen como dichos máximos, mínimos y puntos de inflexión, por lo que se van a usar la primera y la segunda derivada para la extracción de los bordes de los segmentos. La figura 4.2 a) y b) muestra las curvas de nivel cero para las derivadas primera y segunda de la superficie $F(\varphi, \sigma_s^2)$ generada a partir de las medidas de rango, $f(\varphi)$, mostradas en la figura 4.1 b). La figura 4.2 c) nos muestra la superposición de las curvas de nivel cero de ambas derivadas y $f(\varphi)$, todos los extremos de los segmentos, salvo

uno, se hallan muy aproximadamente sobre un pie de alguna curva que recorre amplio rango de escalas, lo que resulta muy prometedor para nuestros objetivos. Comentaremos con más detalle la figura c) más adelante, pero antes vamos a describir el procedimiento para obtener las curvas de nivel de una forma factible, lo que implica la discretización de la convolución, del parámetro de escala σ_s^2 y del ángulo φ .

4.1.1. Discretización

El ángulo φ ya viene discretizado cuando se toman medidas con un aparato real, indicaremos mediante $\varphi_d \in \mathbb{Z}$ al ángulo discretizado, se puede suponer que toma valores según el orden de lectura de las medidas. En la implementación y mientras se efectúan interpolaciones con los valores de φ_d , conviene tratarlos como números reales para evitar errores de redondeo, teniendo siempre presente las situaciones en las que deben comportarse como enteros, y redondear definitivamente al final del proceso completo.

El resto de la discretización del procedimiento plantea tres cuestiones. La primera es la discretización de los núcleos gaussianos, la segunda la discretización de la derivada y la tercera la elección del rango y valores de σ_s^2 , hasta que llegue el momento de discutir la tercera se puede suponer que el parámetro de escala es continuo.

4.1.1.1. Discretización de la convolución

En la sección 2.2.3.2 se comentó la cuestión de la discretización del producto de convolución que define la representación scale-space de una señal, para trasladar la notación de aquella sección a la de esta ha de cambiarse en aquellas fórmulas x_n por φ_d y t por σ_s^2 . Allí comentó la cuestión de la elección de la semi-amplitud, M , de los núcleos, cuya expresión quedaría $M = [c\sigma_s] + 1$, o $M = [c\sigma_s]$ si $c\sigma_s$ es entero, con c entre 3 y 6, $[.]$ denota la parte entera.

Las pruebas realizadas han mostrado que el efecto que tiene el valor de M es apreciable visualmente al observar las curvas de nivel cuando se utilizan valores más pequeños de lo apropiado. En este caso aparecían, especialmente en las escalas superiores, pequeños picos que rompen la suavidad de las curvas. Estos picos pueden inducir la aparición de extremos locales que no aparecían en escalas inferiores en aquellas curvas que, sin los picos y cuando se discretice la escala, tuvieran su máximo un poco por debajo de algún nivel de la escala. Pruebas visuales realizadas para las situaciones probadas en el capítulo mostraron que, aunque para las curvas de menor rango bastaba $M = [3\sigma_s] + 1$, era necesario llegar a aproximadamente $M = [5\sigma_s] + 1$ obtener curvas suaves en las escala superiores. Dado que la diferencia entre uno y otro valor no resulta computacionalmente costosa, como se verá más adelante el número de niveles de escala es reducido, no se ha considerado necesario calcular el valor de N apropiado para cada valor de σ_s , sino que se va a usar $M = [5\sigma_s] + 1$.

En el caso de que las mediciones abarquen 360° es necesario insertar valores al comienzo y al final de $r = f(\varphi)$ de manera que simule una lista circular. El número mínimo de valores que es necesario insertar en ambas posiciones es la semi-amplitud del mayor núcleo que se vaya a usar.

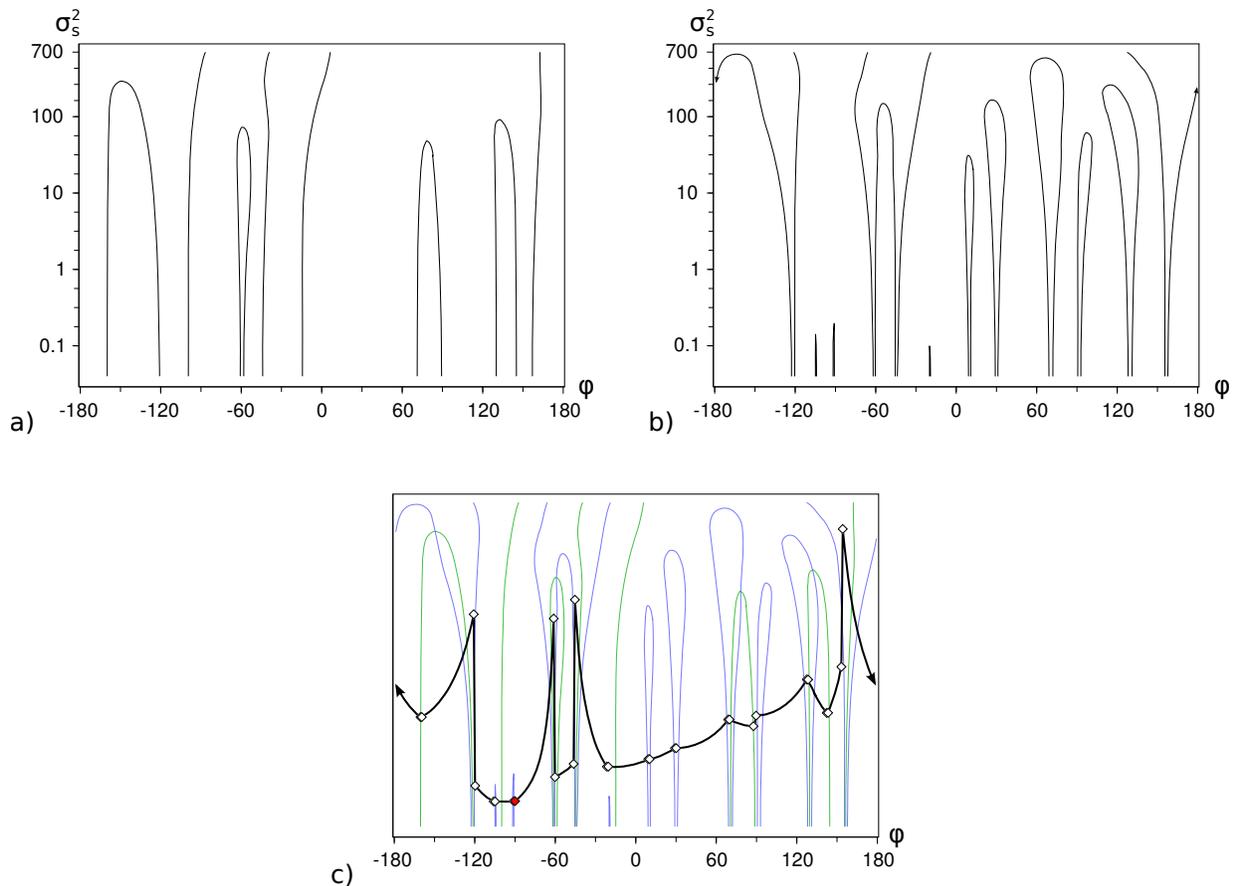


Figura 4.2: Curvas de nivel para los ceros de las derivadas primera a) y segunda b) de la superficie $F(\varphi, \sigma_s^2)$, generada por la convolución de las medidas de rango mostradas en la figura 4.1 b) y núcleos gaussianos de varianza σ_s^2 . c) superposición de las curvas de nivel y las medidas de rango, todos los extremos de los segmentos, salvo el marcado en rojo (-90°), se proyectan muy aproximadamente sobre un pie de alguna curva que recorre amplio rango de escalas. Dado que se abarcan 360° la vista correcta sería sobre un cilindro de manera que el lado izquierdo de la figura se continúa por el derecho.

4.1.1.2. Coste computacional de cálculo de diferencias

La discretización de las derivadas, $\frac{\partial^j F}{\partial \varphi^j}$, se realiza, como es habitual, mediante las diferencias del orden correspondiente,

$$D^j F(\varphi_d) = \sum_{i=0}^j (-1)^{j-i} \binom{j}{i} F(\varphi_d + i)$$

La única cuestión es cómo realizarlas ya que la propiedad sobre las diferencias aplicadas a una convolución discreta establece que:

$$F = T * f \implies D^j F = D^j T * f = T * D^j f$$

La elección de si aplicar las diferencias a F , a los núcleos o a f vendrá determinada por cuestiones de implementación.

La forma habitual y eficiente de realizar productos de convolución es mediante transformadas de Fourier usando el algoritmo FFT (), cuya complejidad espacial es $O(n \log n)$, siendo n el número de valores usados en la transformación. Para fijar ideas supongamos que queremos convolucionar dos señales discretas T y f de tamaños k y l , respectivamente. Primero se calcula la primera potencia entera de 2, n , mayor o igual que la suma de los tamaños de las señales a convolucionar, o sea, $n = 2^{\lceil \log_2(k+l) \rceil + 1}$ o $n = k + l$ si $k + l$ es potencia entera de dos. A continuación se añaden ceros al final de cada señal hasta que su tamaño sea n . Se realizan las transformadas directas de Fourier de ambas señales y se multiplican uno a uno sus valores, que son números complejos. Para terminar se realiza la transformada inversa de Fourier de los productos resultantes y se descartan los $n - k - l$ resultados finales correspondientes a los ceros de relleno. Una transformada de rápida involucra $n \log_2 n$ sumas de complejos, dos sumas reales, y $\frac{1}{2} n \log_2 n$ productos de un real por un complejo, dos productos reales, por lo que el número total de operaciones con números reales, es de $3n \log_2 n$. Para realizar una convolución se necesitan tres transformadas, si las de los núcleos o sus diferencias se precalculan, ya que son constantes, solo se necesitan dos. Por lo que el número de operaciones para la convolución sería $2(3n \log_2 n) = 6n \log_2 n$, por las dos transformadas, más $3n$, por los n productos de complejos, dos productos reales y una diferencia real, en total, $6n \log_2 n + 3n$.

Tengamos en cuenta que se quiere realizar la primera y segunda diferencia. Si empleamos la fórmula $D^j F = D^j T * f$, y se han precalculado transformadas relacionadas con los núcleos, se realizarán menos transformadas que si se emplea así $D^j F = T * D^j f$, ya que de la primera forma solo se calcula la transformada directa de f y de la segunda hay que calcular las de las dos diferencias de f . Empleando la primera forma de la expresión tendremos dos parejas de señales a convolucionar, lo que requiere $2(6n \log_2 n + 3n)$ operaciones. Si por el contrario se aplican diferencias tras la convolución a F , se realizarán una convolución, $6n \log_2 n + 3n$, y dos diferencias de F , $2(3(k+l) - 5)$, en total, $6n \log_2 n + 3n + 2(3(k+l) - 5)$ operaciones. La diferencia en el número de operaciones entre diferenciar antes o después de la, o las, convoluciones será $6n \log_2 n + 3n - 2(3(k+l) - 5)$. Dado que $n \geq k + l$, el caso más favorable para la opción de diferenciar antes será $n = k + l$, en este caso la diferencia

anterior se puede poner $6n \log_2 n - 3n + 10$, que es mayor que 0 para $n \geq 1$. Por lo tanto es preferible la opción de diferenciar después de convolucionar en lugar de hacerlo antes.

4.1.1.3. Niveles de la escala

La forma habitual de tratar los niveles del parámetro de escala σ_s^2 es distribuirlos de forma exponencial, de esta manera se mantiene constante el cambio de información entre niveles sucesivos Mikolajczyk and Schmid (2001), por lo que se elegirán los valores de σ_s^2 de manera que sus logaritmos estén equiespaciados en el rango que se seleccione (octavas). La cuestión de la discretización del parámetro de escala y la elección de los límites de su rango plantea problemas distintos a los comentados anteriormente. Aquí no hay una teoría matemática que muestre el procedimiento óptimo, la solución aquí propuesta está enfocada al tipo de señales producidas por los sensores de rango con suficiente número de mediciones.

El máximo valor de σ_s^2 depende, en la implementación desarrollada, del valor máximo para el tipo de valor que represente a los números reales en la máquina y lenguaje que se use, lo que está íntimamente relacionado con mayor orden con el que se pueden realizar los cálculos de la función de Bessel para generar los núcleos. Por ejemplo con el programa de cálculo numérico Scilab se pueden usar, para los rangos de los núcleos resultantes, hasta el grado 700 y en la implementación en C++ de las librerías Boost hasta 713. En este trabajo hemos tomado el valor de $\sigma_s^2 = 700$, que equivale a una semi-amplitud de núcleo $M = 133$. Este valor muestra, para la resolución angular de 1° , curvas que alcanzan a niveles medios y altos aparte de las que llegan al nivel superior, como en las figuras de los artículos relacionados con el espacio de escalas, lo que ha sido un factor a tener en cuenta a la hora de elegir este nivel máximo, al menos de momento hasta los últimos experimentos.

La cuestiones de si este valor de M se puede reducir, o si es suficiente si el número de mediciones es elevado, o sea, si la resolución angular de las medidas es menor, la aparcamos para cuando cuando hayamos expuesto el conjunto de procedimientos que se describen en este capítulo, ver subsección 4.3.2.2.

Las curvas de la figura 4.2 muestran de manera bastante clara donde se hallan los extremos de los segmentos, pero la situación se complica si se añade ruido. Para comprobarlo observemos la figura 4.3, en la que aparecen las curvas de nivel que corresponden al mundo representado en la figura 4.1 añadiendo ruido a la mediciones. Se han usado los mismos valores para la desviación típica del error de la medición a distancia unidad, σ , que los usados en los experimentos del capítulo 3, $\sigma = 0,002$ equivalente a 2 cm. a 10 m. y $\sigma = 0,02$ equivalente a 20 cm. a 10 m. Al comparar esta figura con la 4.2 vemos que han aparecido curvas estrechas que no alcanzan los niveles superiores, su número se incrementa al aumentar el orden de la diferencia y al aumentar σ . La aparición de estas curvas plantea el problema de dónde fijar el límite inferior de la escala. El rango de valores para σ_s^2 es teóricamente $[0, +\infty)$ y como se trabaja en escala logarítmica el intervalo se convierte en $(-\infty, +\infty)$, lo que complica más la elección del límite inferior. El problema está en que si los niveles descienden bastante en la escala, las curvas generadas por el ruido ocupan una proporción mayor de niveles, de manera que también aumenta la proporción de su longi-

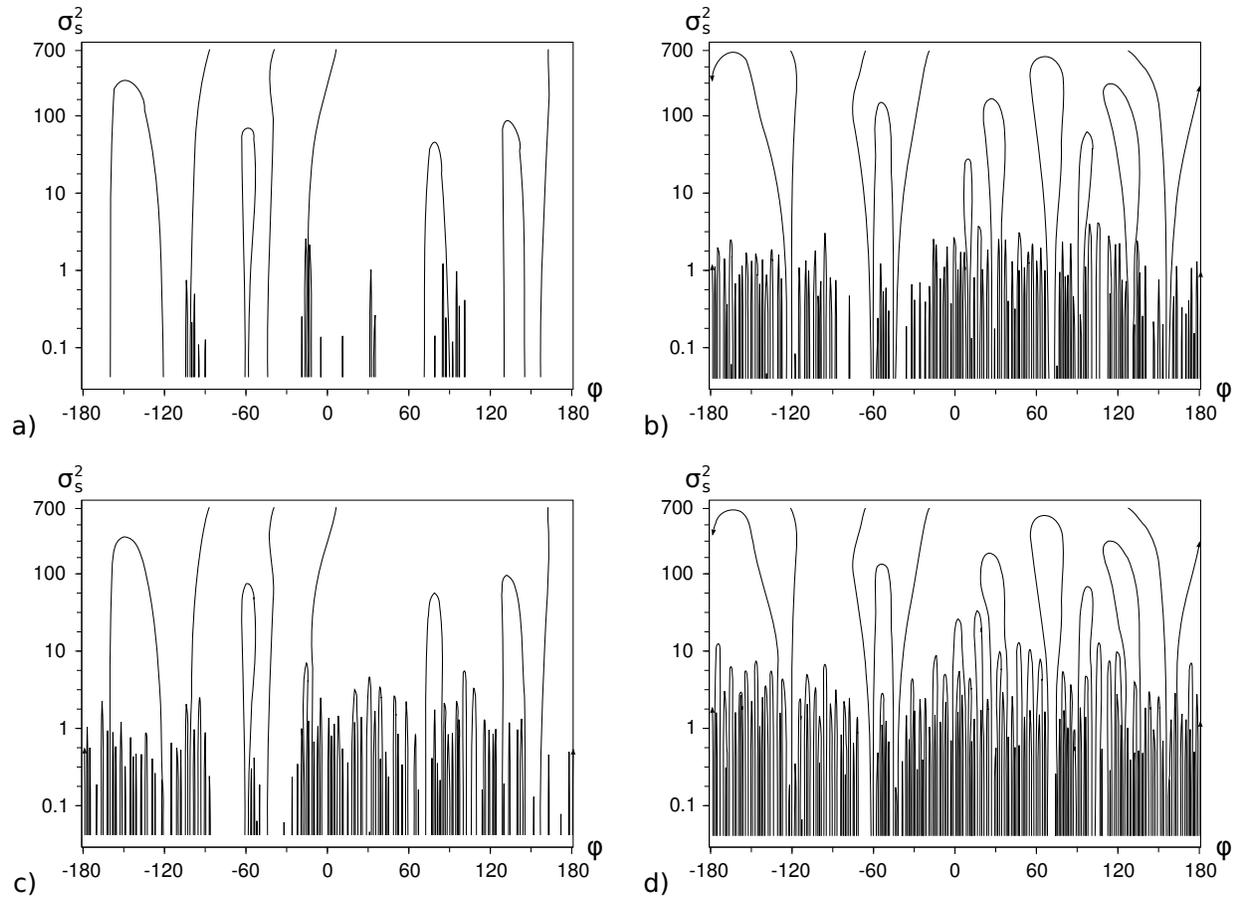


Figura 4.3: Curvas de nivel correspondientes a las de la figura 4.2 pero ahora las mediciones han sido generadas con desviación típica del ruido proporcional a la distancia. En a) se muestran las curvas de la primera derivada y en b) las de la segunda con una desviación equivalente a 2 cm. a 10 m., en c) y d) esas mismas derivadas con una desviación equivalente a 20 cm. a 10 m.

tud con respecto a la de las curvas superiores, las que se supone que indican realmente las características estable. Entonces, si buscamos un criterio de estabilidad de características como aquellas que persisten a lo largo de un rango más amplio de escalas, las curvas debidas al ruido impedirán que se consideren como estables las curvas superiores, obteniéndose una excesiva fragmentación de las características.

Lindeberg deja el problema abierto dependiendo del contexto de aplicación. La solución propuesta en esta tesis consiste en definir dos niveles inferiores: el nivel que se usa para situar las características con precisión, tiene carácter absoluto, y el que se usa para calcular el rango que recorren las curvas, de manera que las curvas debidas al ruido no tengan mucha altura, y que tiene carácter adaptativo, el primero es obviamente menor que el segundo. Para situar el primer nivel, el absoluto, pensemos que la mínima amplitud de núcleo que suaviza es tres y que si el peso otorgado a los extremos de un núcleo de amplitud tres es pequeño, el núcleo tendrá un comportamiento muy similar a uno de amplitud uno, que no suaviza. Por lo que pensamos que un valor mínimo absoluto de σ_s^2 razonable podría ser el mayor que proporciona un núcleo de semi-amplitud uno, con $\sigma_s = 1/5$ se tendría $M = \lceil 5\sigma_s \rceil = 1$, ya que $5\sigma_s$ es entero, los valores del núcleo para esta desviación son $(0,208 \ 0,466 \ 0,208)$.

Dado que hemos fijado unos valores para el máximos absoluto y el mínimo absoluto queda, para calcular las curvas, la cuestión del número de niveles L . En este aspecto hay que establecer un compromiso entre la velocidad de ejecución y la precisión en el cálculo de las alturas de las curvas discretizadas. Como los niveles se usan realmente en escala logarítmica, el intervalo a particionar es $[\log \frac{1}{5^2}, \log 700] = [-3,22, 6,55]$ cuya amplitud es 9,77. En la mayoría experimentos realizados se usaron $L = 24$ niveles lo que da una distancia entre niveles de aproximadamente 0.4, que nos parece suficiente para discretizar sin perder mucha información sobre la altura de las curvas, vista la figura 4.3, y es un número reducido que favorece la eficiencia. Un número similar o mayor no afectaría de forma muy apreciable a las relaciones entre las alturas de las curvas discretizadas, aunque podría ocurrir que alguna característica aparezca fragmentada usando un número dado de niveles, pero no aparezca la fragmentación usando un número de niveles distinto o dos características unidas o separadas. Sin embargo los procedimientos que se expondrán en el capítulo 5 reducen en gran medida estas posibles diferencias. La figura 4.4 se aprecian los 24 niveles y las curvas de nivel discretizadas correspondientes a las curvas continuas de la figura 4.3.

4.1.2. Trazado

Aunque existen librerías que proporcionan la posibilidad de calcular curvas de nivel, suelen estar orientadas a la consecución de curvas suaves usando un gran número de puntos interpolados. Para nuestro objetivo solo necesitamos los puntos donde las curvas interseccionan la malla delimitada por los ángulos discretizados y los niveles de escala, no un resultado estético, y además debido a la propiedad de descarte de variaciones toda curva debe comenzar en el primer nivel, de modo que sólo es necesario visitar las celdas de la malla correspondientes al primer nivel y las que contengan curvas en el resto de niveles. De

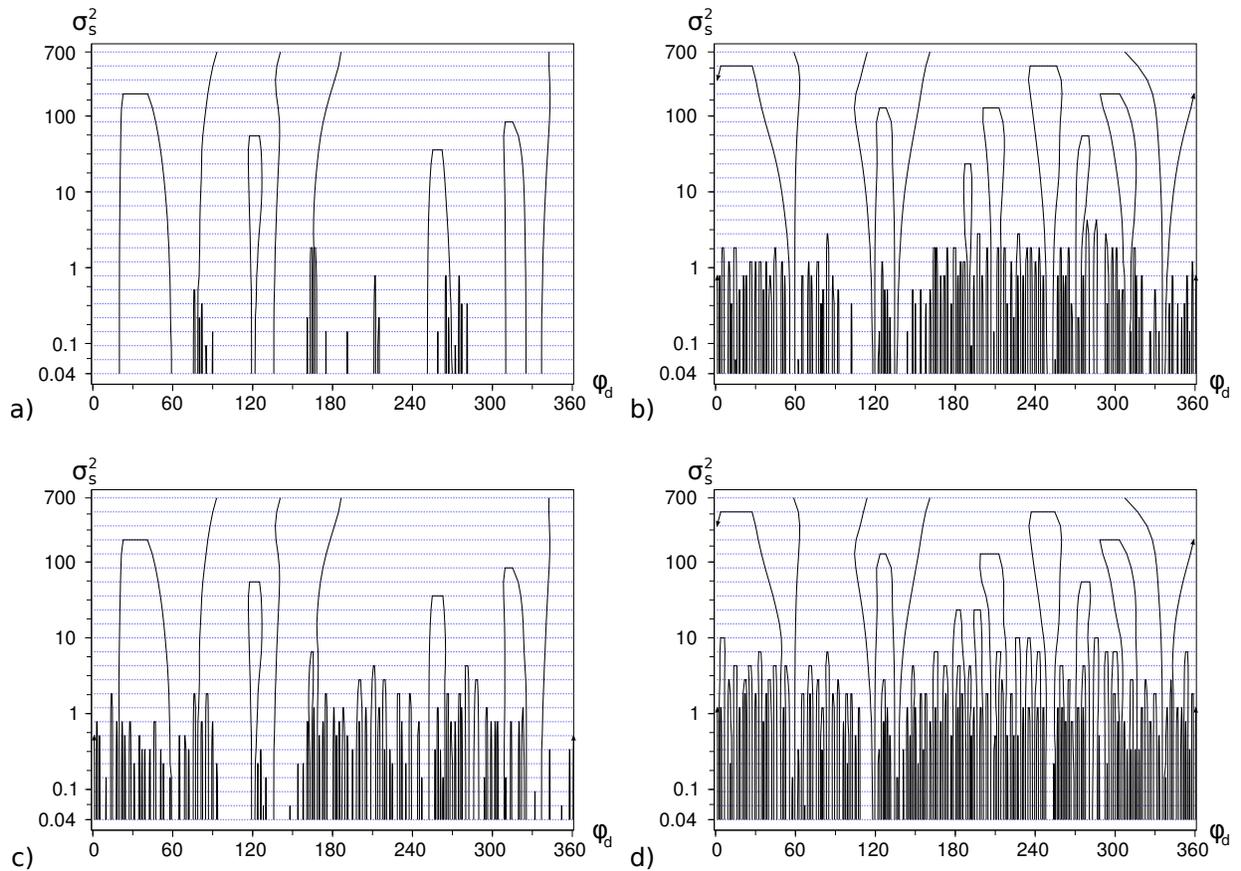


Figura 4.4: Curvas de nivel discretizadas correspondientes a las curvas continuas de la figura 4.3. Los niveles de discretización del parámetro de escala, σ_s^2 , aparecen como líneas discontinuas, los valores del ángulo discretizado, φ_d , corresponden al orden de las mediciones o al número del sensor. En a) se muestran las curvas de la primera diferencia y en b) las de la segunda con una desviación equivalente a 2 cm. a 10 m., en c) y d) esas mismas diferencias con una desviación equivalente a 20 cm. a 10 m.

esta manera el número de celdillas visitadas, que da la complejidad temporal del algoritmo, suele ser inferior a un tercio del número de casillas total, para los valores estudiados del ruido. Por estas razones se implementó el método tradicional de obtención de curvas de nivel de manera que aprovechara estas dos circunstancias, el método tradicional básicamente estudia los signos de la superficie en vértices de cada celdilla y si hay un cambio de signo se realiza una interpolación entre los valores de la superficie para obtener el punto de la curva de nivel.

Cuando las mediciones abarcan 360° , se añade una columna al final de la malla que correspondería al espacio entre el último, n , y el primer sensor, 1. En la discretización final, un ángulo situado en la última columna se asocia con el primer sensor si su redondeo es $n + 1$ y con el último si es n .

4.1.3. Filtrado del ruido

4.1.3.1. Área y altura de las curvas de nivel

Se comentó más arriba la utilización de otro valor mínimo con carácter adaptativo, este mínimo debe adaptarse al ruido presente, por lo que procedimos a buscar una caracterización del ruido. Como se comentó anteriormente el ruido suele generar curvas estrechas no muy altas, por lo que se va a usar el producto de la altura por la anchura en la base, una aproximación del área bajo la curva, para caracterizarlo. Se diseñaron dos experimentos que simulaban perturbaciones como se comenta a continuación. En una mitad de un segmento sin ruido que abarcaba 177° con resolución angular de 1° , se alteraron algunos valores equiespaciados de r , en una cantidad predeterminada que era proporcional en a esos valores de r que notaremos \tilde{r}_i . Los factores de proporcionalidad, κ_j , se aplican de la siguiente manera $\tilde{r}_i = r_i(1 \pm \kappa_j)$ y se usaron 41 valores que abarcaban desde 0.0001 hasta 0.16, de manera que se simulan desviaciones típicas del error de la medición, σ , desde en torno a la diezmilésima de hasta aproximadamente $\frac{0,16}{3} \sim 0,05$, si entendemos que los valores típicos del error se hallan entre $\pm 3\sigma$. Estas desviaciones traducidas a centímetros a 10 m. vienen a corresponder a un rango de aproximadamente 0,1 cm. a 50 cm. En el primer experimento se alteraron los valores de la siguiente forma $\tilde{r}_i = r_i(1 + \kappa_j)$, escogidos con ángulos de incidencia desde 85° hasta 0° separados 5° . En el segundo experimento $\tilde{r}_i = r_i(1 + \kappa_j)$ y $\tilde{r}_{i+1} = r_{i+1}(1 - \kappa_j)$, con los \tilde{r}_i espaciados como en el primer experimento. A fin de que el efecto de la discretización de niveles de escala fuera despreciable se usaron 100 niveles con $\sigma_s^2 \in [0,001, 700]$, el núcleo correspondiente a $\sigma_s^2 = 0,001$ es $(0,0005 \ 0,999 \ 0,0005)$, prácticamente un núcleo de un solo valor.

También se realizaron estos experimentos con los niveles comentados en la subsección 4.1.1.3 a fin de comprobar el efecto de la discretización. Los resultados en forma de histogramas del área se muestran en la figura 4.5. Resulta llamativo la forma del histograma para la primera diferencia cuando se generan alterando dos valores de r consecutivos, figura 4.5 d) y g), se aprecian como dos zonas separadas en el histograma. La zona de la derecha, correspondiente a las mayores áreas, es originada por los mayores valores de κ_j cuando el ángulo de incidencia del haz es próximo a 0. Estos valores extremos se han usado en el

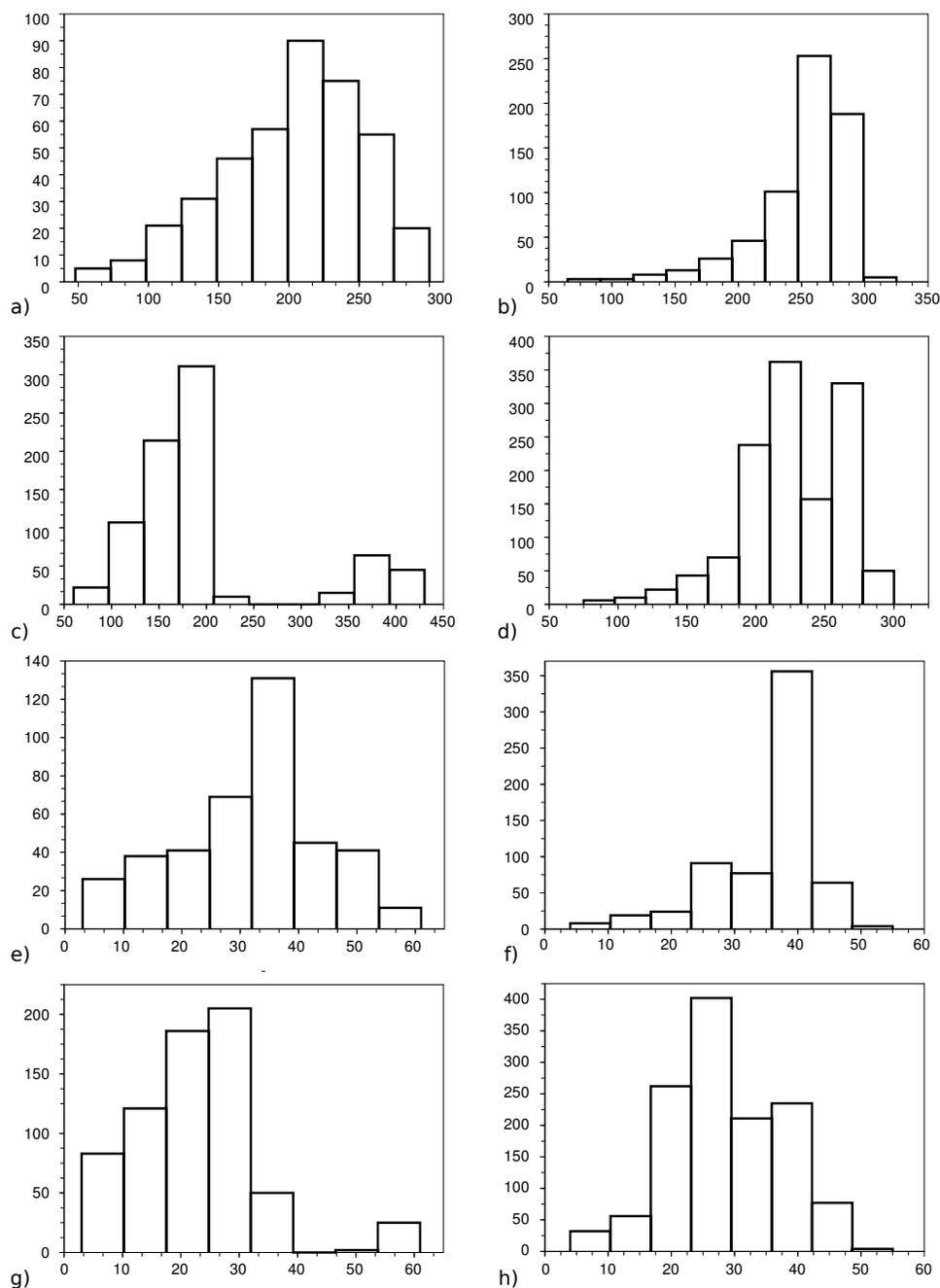


Figura 4.5: Histogramas correspondientes a las áreas obtenidas para las curvas de nivel originadas por las perturbaciones “artificiales”. En a), b), c) y d) se usaron 100 niveles de escala desde 0.001 a 700 y en las restantes 24 niveles, de 0.04 a 700. En a), b), e) y f) las perturbaciones artificiales eran de la forma $\tilde{r}_i = r_i(1 + \kappa_j)$ equiespaciados y en las restantes $\tilde{r}_i = r_i(1 + \kappa_j)$ y $\tilde{r}_{i+1} = r_{i+1}(1 - \kappa_j)$. En a), c), e) y g) las curvas corresponden a la primera diferencia y en el resto a la segunda.

experimento para comprobar los límites del procedimiento y están más allá de los casos de aplicación, por lo que no van a ser tenidos en cuenta en la discusión siguiente. En cuanto al resto de casos de orden de diferencias y tipo de perturbación se observan semejanzas entre los histogramas correspondientes a las dos conjuntos de niveles escalas usados, de modo que la elección del conjunto de niveles más reducido para las aplicaciones prácticas no parece que afecte demasiado a los resultados.

En cuanto a los límites del rango de áreas, parece que es semejante en todos los casos, para un determinado número de niveles, independientemente de tipo de perturbación y el orden de la diferencia. Este hecho no va a permitir introducir una cota superior C , tentativa, para las posibles áreas ocasionadas por perturbaciones “reales”. Esta cota superior no puede ser condición necesaria y suficiente para caracterizar plenamente las curvas generadas por el ruido, ya que puede haber curvas generadas por las verdaderas características del entorno que no superen dicha cota y viceversa. Además las perturbaciones en un mundo real pueden ser mucho menores que las aquí usadas especialmente en interiores. Por ello vamos a diseñar un procedimiento que partiendo de esta primera caracterización se adapte al ruido presente en cada entorno.

Recordemos que se comentó que la característica determinante de las curvas para definir las características estables es su altura, por lo que vamos a refinar la cota tentativa en función de las alturas de las curvas. En la figura 4.6 se muestran los histogramas las alturas de correspondientes a los casos representados en la figura 4.5 e), f), g) y h), $L = 24$ niveles, cuando el área de la curva es menor que $C = 50$. Este valor se ha elegido a partir de los rangos de áreas observados en los histogramas de la figura 4.5 e), f), g) y h), incluso se podría llegar a 60, pero se ha preferido usar un valor algo más “conservador” para intentar evitar en lo posible filtrar curvas debidas a características y no al ruido. Otro hecho que parece observarse en los histogramas de la figura 4.6 es una aproximación a la distribución normal en las alturas, lo que nos va a permitir refinar la cota tentativa con el apoyo de la Estadística. Sea \bar{h}_C la media de las alturas de las curvas cuya área es menor que C y s_C su desviación típica, el filtro consiste en eliminar aquellas curvas, del conjunto total, cuya altura es menor que $\bar{h}_C + ks_C$, donde k es una constante que determina aproximadamente la probabilidad de que una curva sea filtrada. Valores de k posibles son los que se usan habitualmente en este tipo de acotaciones basadas en la normal, o sea, en el rango de 2 a 3, aproximadamente, y pudieran ser distintos para cada diferencia.

Con el objetivo de comprobar si estos resultados son trasladables a mundos simulados, donde hay varios segmentos y el ruido es aleatorio, se realizó el siguiente experimento. Para el mundo de la figura 4.1, muestrado con resolución angular de 1° , se usaron 22 valores para la desviación del error de medición, σ , tomados en el intervalo $[0,0005, 0,02]$ y se generaron 240 series de perturbaciones aleatorias, 20 por cada valor de σ . Se hallaron las alturas de las curvas de nivel debidas al ruido y las que tenían un área menor que $C = 50$, los histogramas se hallan en la figura 4.7. Se puede apreciar una gran similitud entre los histogramas correspondientes a cada diferencia, aunque en las seleccionadas por áreas hay una cola superior ligeramente más alargada. También se puede apreciar una más acusada normalidad en estos histograma que en los producidos por perturbaciones “artificiales”, figura 4.6. Estos dos hechos muestran la capacidad del procedimiento diseñado

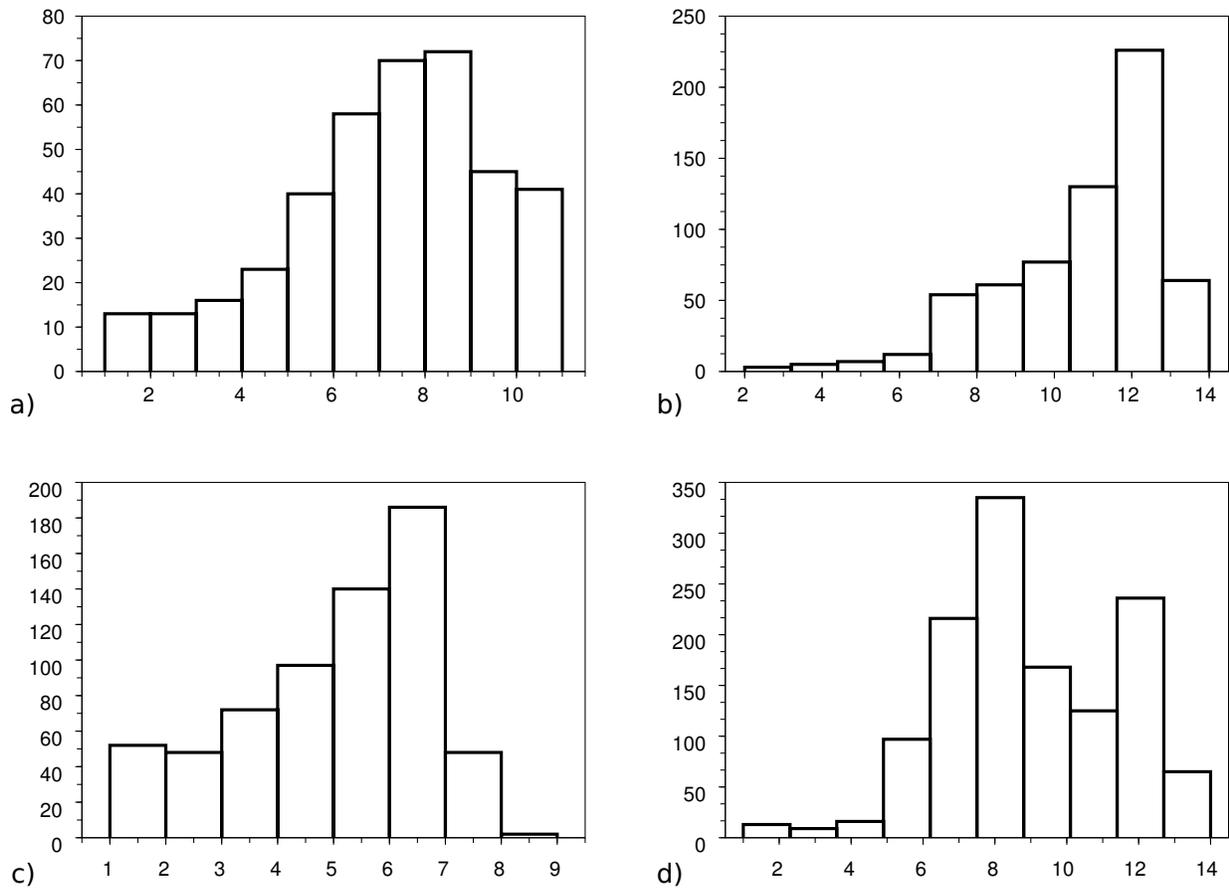


Figura 4.6: Histogramas de las alturas, obtenidas para las curvas de nivel originadas por las perturbaciones “artificiales”, cuando el área de las curvas es menor que 50. Corresponden al caso de 24 niveles de escala, figura 4.5 e), f), g) y h).

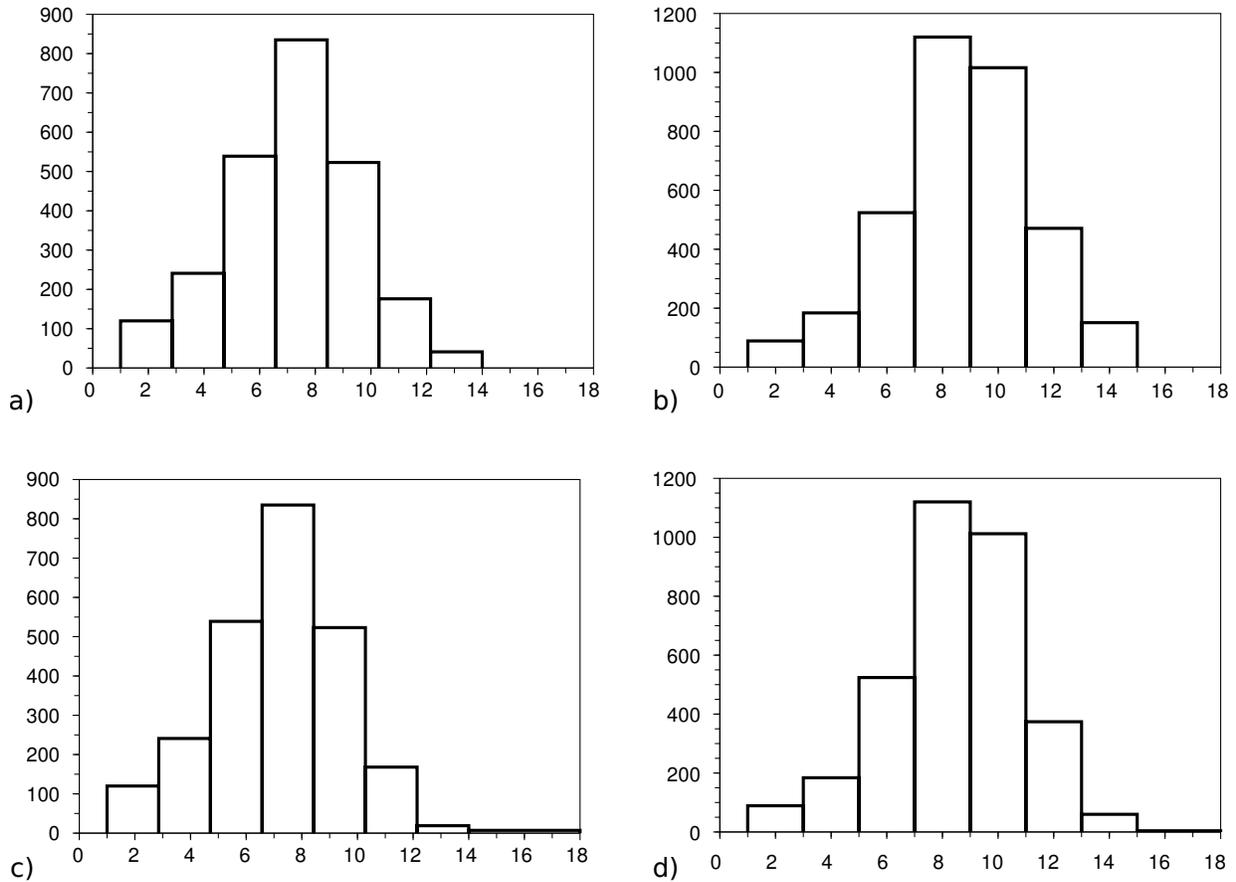


Figura 4.7: Histogramas correspondientes a las alturas obtenidas para las curvas de nivel, generadas a partir del mundo de la figura 4.1, originadas exclusivamente por las perturbaciones, en a) diferencia primera y en b) segunda. Y los correspondientes a las alturas de las curvas cuya área es menor que $C = 50$, c) diferencia primera y d) segunda.

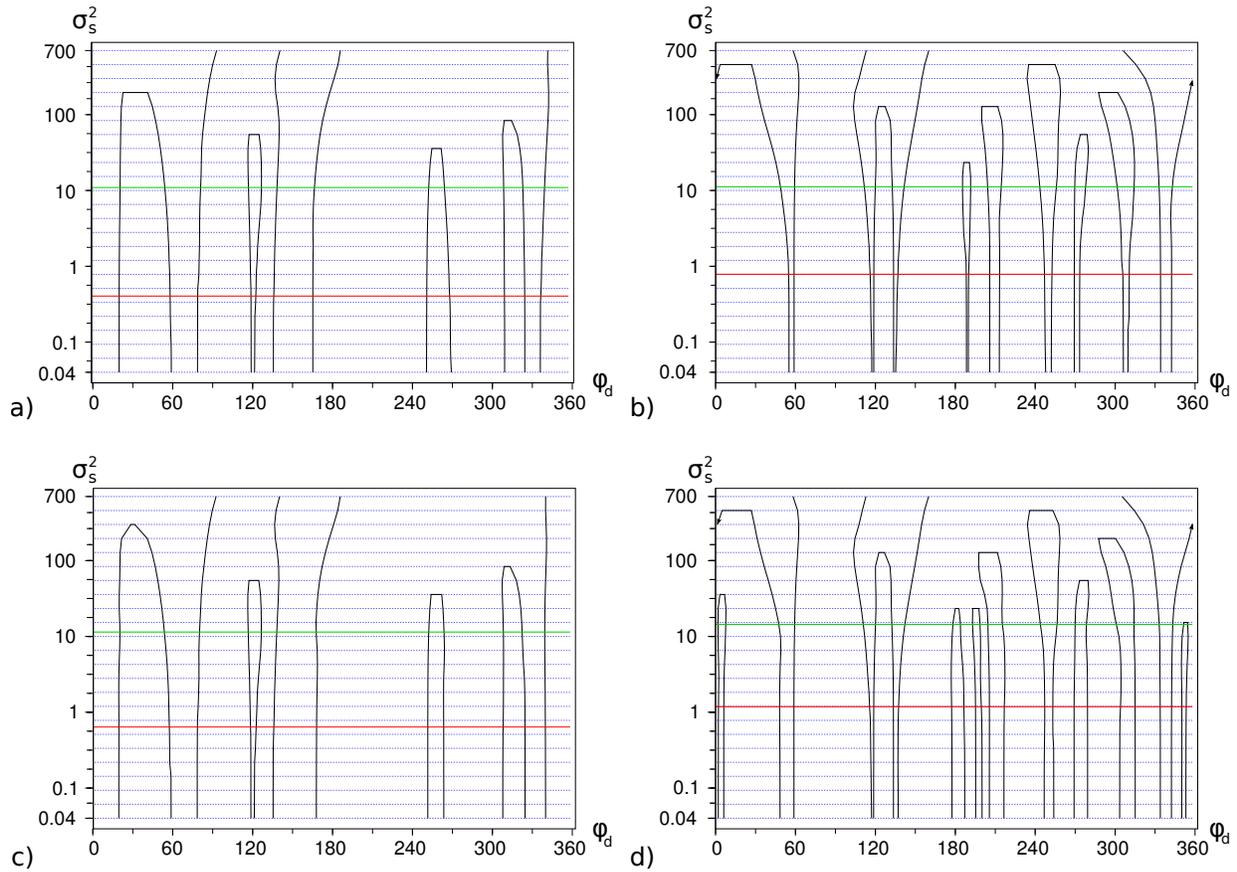


Figura 4.8: Curvas de nivel filtradas. En a) de la primera diferencia y en b) de la segunda para $\sigma = 0,002$ (2 cm. a 10 m.). En c) de la primera diferencia y en d) de la segunda para $\sigma = 0,02$ (20 cm. a 10 m.). La línea continua horizontal roja indica el valor de \bar{h}_C y la verde más arriba el de $\bar{h}_C + 2,5s_C$.

para reconocer el ruido y, por lo tanto para poder filtrarlo. En la figura 4.8 se muestra el resultado de aplicar el procedimiento propuesto al conjunto de curvas de la figura 4.4 con valores $C = 50$ y $k = 2,5$ y desviaciones del error de medición, $\sigma = 0,002$ y $\sigma = 0,02$. Al comparar estas curvas con las curvas de la figura 4.2 se observa que para la primera diferencia el filtro ha eliminado todas las curvas debidas al ruido para los dos valores del error y también para la diferencia segunda para el valor $\sigma = 0,002$. Cuando se aplica la diferencia segunda con $\sigma = 0,02$ no se logran eliminar tres de las curvas creadas por el ruido. El número de curvas creadas por el ruido se aprecia en las figuras 4.3 o 4.4.

El número de total curvas debe ser siempre menor o igual que el de sensores n , puede razonarse que como máximo se puede pensar una situación en la que de cada punto discreto salga y acabe una curva, o se puede comprobar experimentalmente, ver subsección 4.1.3.2. Dado que la anchura y la altura se pueden calcular durante la fase de trazado, habrá como mucho que calcular la media y desviación típica de n valores y realizar n comparaciones para filtrar las curvas, luego la complejidad temporal del filtro es $O(n)$.

4.1.3.2. Número de curvas de nivel en función del ruido

Se mencionó, al comentar la figura 4.3, que al aumentar el ruido también aumentaba en número de curvas producidas por él. Por completitud del estudio de caracterización del ruido, aunque no vamos a usarlo a efectos prácticos, se procedió a estudiar el número, más bien la proporción, de curvas producidas por el ruido. Usando los datos obtenidos al realizar el experimento final de la subsección anterior 4.1.3.1 se procedió a representar gráficamente los resultados obtenidos y se encontró que las medias de las proporciones, p , obtenidas para cada valor de σ y los propios valores de σ se ajustaban bastante bien a una función de la forma:

$$p = \frac{2a}{\pi} \arctan(b\sigma)$$

el factor $\frac{2}{\pi}$ normaliza la arcotangente de manera de el parámetro a representa la proporción máxima, el parámetro b afecta a la pendiente media del tramo cuasi-lineal central de la función.

Para la segunda diferencia no se encontró una diferencia apreciable en las proporciones, el tramo de valores de σ estudiado aparecía en la parte cuasi-horizontal superior de la arcotangente. En cambio para la primera diferencia sí se observaba claramente el comportamiento debido a los dos tramos. Los coeficientes de determinación, R^2 , obtenidos fueron 0.99 cuando se emplearon todos los valores de σ y de 0.95 cuando se empleó la mitad inferior del rango de σ , que correspondía aproximadamente a la zona cuasi-lineal central. El problema que encontramos para extraer alguna consecuencia práctica de este hecho fue que al cambiar de mundo simulado, el valor estimado para el parámetro b variaba considerablemente, aunque se seguían manteniendo los R^2 elevados.

4.2. Árboles de intervalos

4.2.1. Rectángulos estables

En la subsección 2.2.3.3 se describió la técnica de árbol de intervalos para determinar las escalas naturales, aquí se usará como medida de estabilidad la usada por Witkin (1984), la altura del rectángulo. Como se comentó en la subsección 4.1.1.3 el problema en este planteamiento aparece al intentar fijar el mínimo nivel de escala, un mínimo muy pequeño hará que los rectángulos correspondientes a las curvas debidas al ruido tengan una altura mayor que los correspondientes a las verdaderas características, produciéndose una excesiva fragmentación. El filtrado de ruido propuesto en la subsección 4.1.3 elimina la mayor parte de las curvas debidas al ruido pero no puede hacer nada con el problema de las alturas de aquellas curvas que hayan pasado el filtro. Para atajar este problema se propone usar un nivel de escala mínimo adaptativo a la hora de crear el árbol de intervalos, este nivel debe tener en cuenta el ruido presente en las mediciones, por lo que se pensó que la media \bar{h}_C usada para definir el filtro del ruido podía constituir un valor adecuado para el mínimo adaptativo.

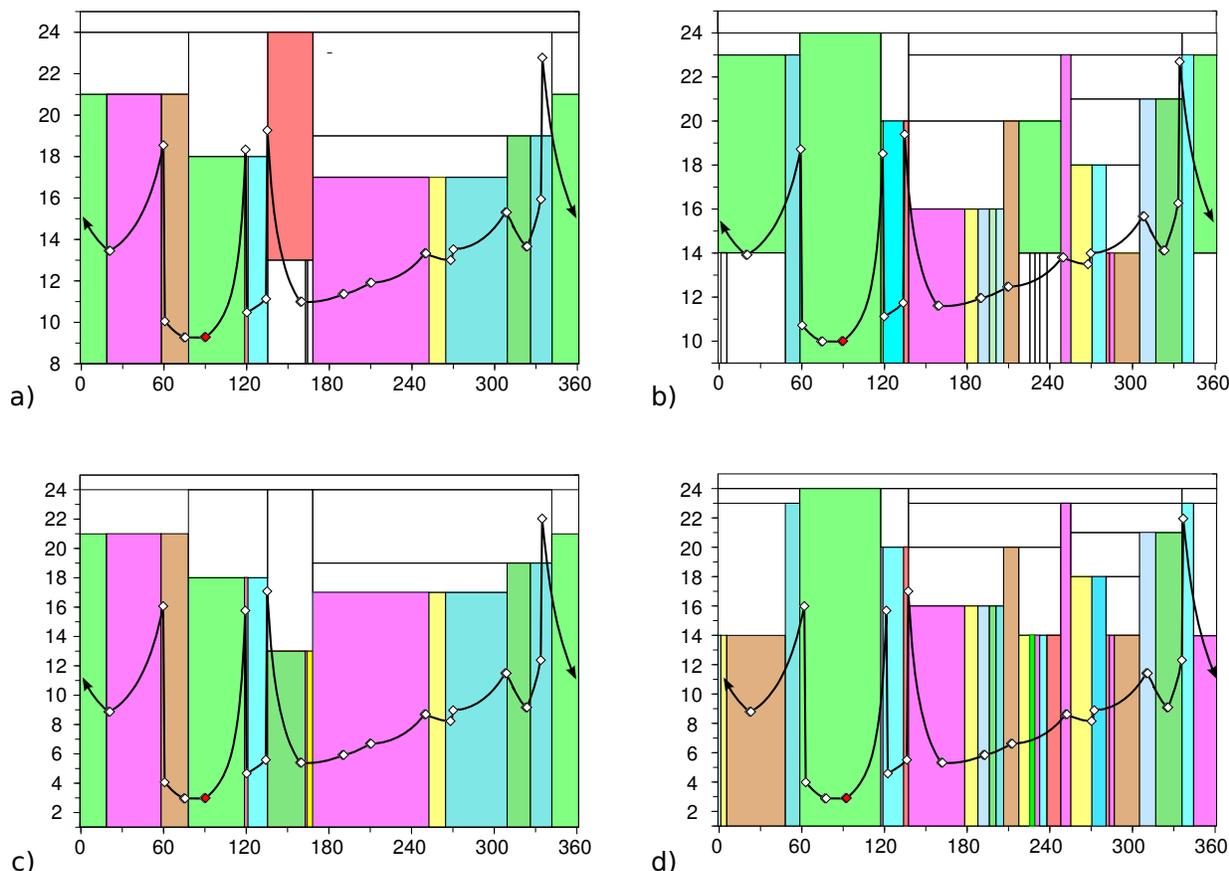


Figura 4.9: Rectángulos representando el árbol de intervalos obtenidos del conjunto de curvas de la figura 4.4 c) y d), $\sigma = 0,02$, tras filtrar el ruido por debajo de $\bar{h}_C + 2s_C$. En la parte superior se ha aplicado el mínimo adaptativo \bar{h}_C para los niveles de escala a la hora de calcular las alturas de los rectángulos, en la parte inferior el mínimo ha sido el nivel inferior, se usaron 24 niveles. En a) y c) se representan los rectángulos correspondientes a las curvas de la primera diferencia y en b) y d) los de la segunda. Los rectángulos coloreados son los estables. Se han sobrepuesto las medidas de rango, todos los extremos de los segmentos, salvo el marcado en rojo (90°), corresponden muy aproximadamente a la abscisa da un lado vertical de algún rectángulo estable. Aunque aparecen rectángulos estables que no corresponden a ningún extremo debidos a la gran cantidad de ruido presente, valores típicos de hasta ± 60 cm. a 10 m., el mínimo adaptativo disminuye su número y por lo tanto la fragmentación.

En la figura 4.9 se observa el resultado de aplicar la técnica de árbol de intervalos a las curvas de la figura 4.4 c) y d) con y sin mínimo adaptativo. Se ha mostrado el caso de desviación típica del error de la medición “grande” $\sigma = 0,02$, ya que en el caso del error pequeño el filtro logró eliminar todas las curvas debidas al ruido y no habrá diferencias entre usar y no el mínimo adaptativo. El valor del límite superior para el filtro de ruido que se usó fue $\bar{h}_C + 2s_C$. En la figura puede verse que en algunas zonas el uso del límite adaptativo evita la excesiva fragmentación. Cuando se empleó como límite superior $\bar{h}_C + 2,5s_C$ no hubo diferencias entre emplear o no el límite adaptativo y además se consiguió el mismo agrupamiento que en el caso anterior. Esto parece mostrar que el empleo mínimo adaptativo flexibiliza la elección de la constante de proporcionalidad k usado para definir el límite superior $\bar{h}_C + ks_C$, de manera que si se elige un poco más pequeño de lo adecuado, el límite adaptativo evitará en gran medida la fragmentación provocada por esa elección.

Cuando el conjunto de mediciones abarca 360° puede haber curvas de nivel que salgan por un extremo del rango y aparezcan por el otro, habrá que manejar los rectángulos como si realmente estuvieran dibujados sobre un cilindro. En la figura 4.9 a) y b) los rectángulos coloreados en verde, situados en los extremos izquierdo y derecho de los gráficos, tienen el mismo color indicando que son dos partes de un mismo rectángulo. Cuando el conjunto de mediciones abarca menos de 360° se introducen, si no aparecen debidos a las curvas, las posiciones correspondientes al primer y último sensor en todos los niveles de escala, a fin de limitar los rectángulos que lo necesiten.

Para hacernos una idea de la complejidad computacional de este proceso, pensemos que cada curva filtrada que alcanza el nivel superior produce dos ramas y las demás tres. Luego como mucho el número de rectángulos, nodos del árbol, será tres veces el número de curvas filtradas. El árbol sólo se recorre una vez, mientras se crea, en el mismo proceso se puede ir calculando el máximo de cada rama. Dado que la operación de filtrado elimina la gran mayoría de las curvas debidas al ruido, cuyo número está relacionado con el número de mediciones n , el número de nodos dependerá en gran medida de las características del entorno y del nivel de ruido, más que de n . Por esto la complejidad temporal de este proceso depende bastante más de la propia complejidad estructural del mundo, número de segmentos y nivel de ruido, que del número de mediciones.

4.2.2. Obtención de las semillas

Para terminar de exponer el procedimiento de clustering vamos a proceder a comentar cómo se fusiona la información procedente de los rectángulos estables, obtenidos a partir de los árboles de intervalos correspondientes a las dos diferencias, para obtener las semillas para la estimación mediante los filtros de Kalman. El procedimiento desarrollado es simple y se basa en el número de puntos mínimo y máximo para la semilla. La elección del mínimo depende de la precisión mínima que se desee obtener en la estimación de los valores iniciales para el filtro de Kalman, ver sección 5.2, la del máximo tiene que ver con la precisión máxima que se desee obtener para los valores iniciales y el tiempo de su obtención.

Llamaremos S al tamaño máximo de semilla que se usará y s al mínimo. A partir de ahora solo consideraremos los ángulos discretizados, φ_d , que definen los extremos de

los rectángulos estables, conviene seguir considerándolos como números reales, no enteros, para evitar los errores de redondeo, a no ser que se comparen con algún entero, y discretizarlos realmente al dar los valores de los sensores correspondientes a los puntos semilla. El procedimiento depende de si las medidas abarcan 360° o menos.

Se comienza la fusión ordenando los valores de los φ_d correspondientes a ambas diferencias en una sola lista eliminando duplicidades, φ_d^i , cuando las medidas abarcan 360° hay que tratar la lista de forma circular, teniendo cuidado de que al operar con ángulos estos tengan la misma determinación. Cuando se abarca menos de 360° , todos los $\varphi_d^i < s$, los muy próximos al primer sensor, se consideran iguales a 1, o sea estos intervalos desaparecen, y los $\varphi_d^i > n - s$, n indica el último sensor, se hacen iguales a n . Para los $s \leq \varphi_d^i \leq n - s$, independientemente de ángulo abarcado, si $\varphi_d^{i+1} - \varphi_d^i < s$ se hace $\varphi_d^i = \frac{\varphi_d^{i+1} + \varphi_d^i}{2}$ y φ_d^{i+1} se elimina de la lista. De esta manera la lista ya solo contiene valores que definen intervalos capaces de albergar una semilla de tamaño s . De estos intervalos los que puedan albergar como máximo S puntos semilla pasan a constituir una semilla cada uno, para los que pueden albergar más de S puntos, se eligen S puntos consecutivos centrados en cada intervalo para formar las semillas.

4.3. Experimentos

En esta sección se van a comentar algunos experimentos de clustering para la obtención de puntos semilla. Como se ha ido comentando, uno de los problemas que aqueja a diversos métodos para la obtención de escalas naturales es la posibilidad de fragmentación excesiva de las características, por lo que suelen necesitar de un procedimiento posterior de mezclado de los fragmentos Rosin (1994). Aunque los procedimientos propuestos en esta tesis muestran evitar en gran medida la aparición de la fragmentación excesiva se sigue necesitando un tal procedimiento de mezcla y, como se verá en los comentarios de los experimentos, de descubrimiento de características perdidas. Estos procedimientos se explicarán en el capítulo 5. Teniendo en cuenta, además, que el objetivo perseguido con la aplicación del procedimiento de clustering no es tanto la obtención de los límites exactos de las características, sino de semillas situadas convenientemente, no se van a realizar experimentos exhaustivos o comparativos sobre el comportamiento de los procedimientos propuestos en este capítulo. Este tipo de experimentos se realizarán en el mencionado capítulo 5. También se estudiará la cuestión del nivel máximo de la escala

4.3.1. Descripción

Sobre el mundo representados en la figura 4.1 se van a generar mundos simulados con errores de medición de desviación típica unitaria $\sigma = 0,002$, simulando un entorno doméstico o de oficina, $\sigma = 0,02$, simulando los setos de un jardín. La resolución angular de las mediciones es de 1° . El mundo ha sido diseñado de manera incluye problemas para su segmentación, tanto en general como para el procedimiento propuesto. Problemas de índole general son los siguientes:

- Dos segmentos forman un ángulo pequeño entre ellos, 15° , los que ocupan la zona frontal al robot, orientación 0° .

- Hay medidas que sobrepasan los $8m$. con ángulos de incidencia grandes, esto ocurre en los cuatro segmentos de mayor longitud. En esta situación la distancia entre dos puntos consecutivos puede superar el metro como ocurre en los segmentos que forman una especie de corredor en la figura. Este es un problema para algoritmos que fijan la máxima distancia posible entre dos puntos consecutivos para considerar que pueden pertenecer a un mismo segmento.

El problema que presenta el mundo para el método propuesto es el de segmentos con zonas de incidencia con ángulos muy cercanos o que contienen a 0° . Sospechamos que, visto el efecto de los errores en esa situación, ver figura 3.11, y aunque la desviación del error fuera constante, muchos métodos encontrarán dificultades en estas situaciones, aunque no se han encontrado comentarios sobre el particular en la literatura sobre el tema. Aparte de este problema la situación descrita plantea otro muy relacionado de método propuesto. Si se observa la figura 3.1 se verá que cuando el ángulo de incidencia es $\varphi - \theta = 0^\circ$, la curva presenta un mínimo local propio, que será detectado como extremo, inexistente, de un segmento por la primera diferencia. Esta es, pensamos, una de las principales fuentes de la posible aparición de fragmentación excesiva. Esta situación puede conllevar incluso el efecto opuesto, el pasar por alto algún extremo de segmento, como ocurre con el extremo marcado con rojo en la gráfica de las mediciones en polares que ha aparecido en varias figuras, la más próxima la 4.9. En la figura no se aprecian correctamente las posiciones y parece que el extremo mencionado y el anterior están a la misma altura, pero no es así, el anterior está ligeramente más abajo, por lo que el punto marcado no aparece ni como mínimo, ni máximo, ni punto de inflexión que puedan descubrir de forma apreciable las diferencias primera y segunda. Si ambos puntos estuvieran a la misma altura el mínimo se descubriría aproximadamente en la mitad de segmento, produciéndose la comentada fragmentación.

Por último se va a investigar brevemente la cuestión, aparcada anteriormente, del máximo de los niveles de escala, ver subsección 4.1.1.3, o lo que es lo mismo el tamaño máximo de núcleo necesario dependiendo del número de mediciones. Se repetirán los experimentos anteriores para varios números de mediciones y resoluciones angulares usando valores menores de M y del números de niveles L .

Para ello se repitió el anterior para $\sigma = 0,002$ empleando una resolución angular de $0,25^\circ$, o sea, 1440 mediciones, ahora la semi-amplitud máxima de núcleo, $M = 133$, es un 9,2% del número de mediciones, mientras que en el experimento anterior representa el 36,9%.

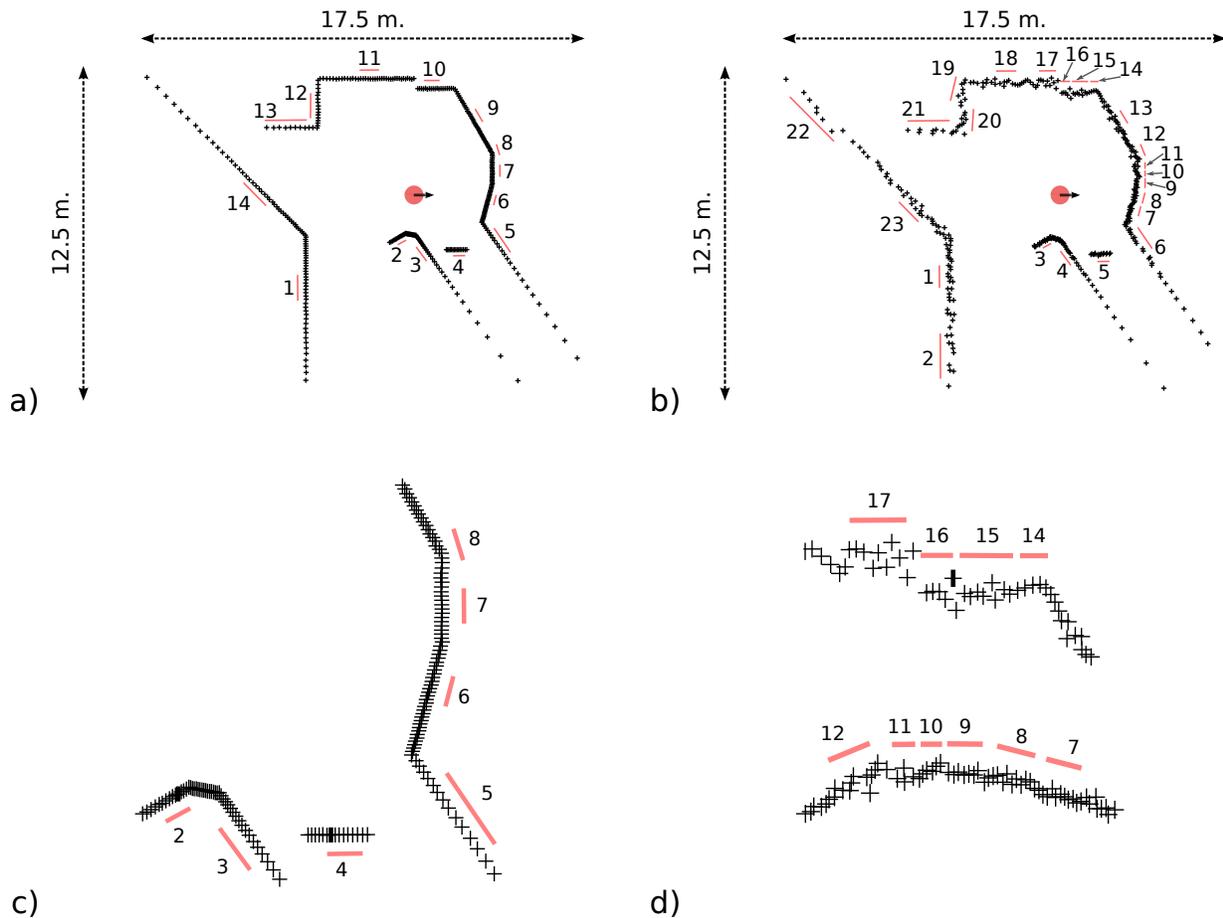


Figura 4.10: Semillas obtenidas cuando la desviación típica del ruido es a) $\sigma = 0,002$, un entorno doméstico o de oficina y cuando es b) $\sigma = 0,02$ setos de un jardín. c) y d) muestran detalles de las zonas problemáticas para el procedimiento propuesto.

4.3.2. Resultados

4.3.2.1. Usando el tamaño máximo para los núcleos

Se emplean 24 niveles de escala, logarítmica, desde 0.04 a 700, lo que implica una semi-amplitud para el mayor núcleo de 133. Los parámetros para el filtrado del ruido fueron $C = 50$ y $\bar{h}_C + 2,5s_C$.

Cuando se utilizaron errores de medición con desviación típica unitaria $\sigma = 0,002$, el procedimiento propuesto situó semillas aproximadamente en el centro, entendido como la mediana de la serie de puntos del segmento no como el centro geométrico, de todos los segmentos salvo uno, el situado entre las semillas 2 y 3 de la figura 4.10 a) y c). Uno de los extremos de este intervalo es el punto no detectado marcado de rojo en varias figuras anteriores, p. ej. 4.9, y sobre el que se ha comentado en la descripción de estos experimentos. Además una semilla, la 8, se encuentra a caballo entre dos segmentos este ha sido el único

caso de fragmentación excesiva para este nivel de ruido.

Con $\sigma = 0,02$ la fragmentación, como era de esperar, aumenta como puede verse en la figura 4.10 b) y d), hay pocos segmentos que contengan solo una semilla.

Es claramente necesario, especialmente cuando el ruido es grande, un procedimiento que mezcle los segmentos que se estimen a partir de semillas contenidas en el mismo segmento real y otro que intente descubrir segmentos que no contengan semillas, cuestión que se deja para el capítulo 5.

4.3.2.2. Reducción del tamaño máximo núcleo y del número de niveles

El parámetro de filtrado C es una cota para las áreas por lo que si el número de niveles L disminuye, sin cambiar el límite inferior, C deberá disminuir en la misma proporción. Un afortunado error de programación hizo que no se aplicara dicha disminución durante unas pruebas reduciendo el número de escalas y su máximo. De esta manera se encontró que se podía reducir L hasta 15 y, más importante aún, la semi-amplitud máxima M hasta $\frac{n}{9}$, siendo n el número de mediciones, sin que hubiera variaciones apreciables en el número y posiciones de las semillas. El efecto relativo sobre C fue como si aumentara en un factor de $\frac{24}{15} = 1,6$, o sea, como si pasase de valer 50 a 80. El valor $C = 50$ fue obtenido estudiando perturbaciones “artificiales” separadas, ver subsección 4.1.3.1, perturbaciones realmente aleatorias consecutivas pueden presentar configuraciones que generen curvas de nivel más extensas que las generadas por una perturbación “artificial” bipuntual, de hecho los histogramas de áreas correspondientes a perturbaciones realmente aleatorias mostraban colas superiores muy alargadas aunque de altura casi despreciables. De manera que incrementando el efecto relativo de C se incluyen más curvas debidas al ruido, y posiblemente algunas no originadas por él, aumentándose los valores \bar{h}_C y s_C del filtro.

El valor de M empleado en los experimentos anteriores fue $133 \simeq \frac{n}{3}$, por lo que con $M \simeq \frac{n}{9}$ habrá una reducción importante en el tamaño de los núcleos. Esta reducción implicará un importante aumento de la velocidad de proceso, sobre todo si implica usar la mitad de elementos en las convoluciones. Por ejemplo en el caso de $n = 181$ abarcando 180° , número de mediciones del láser SICK LMS200, y empleando $M \simeq \frac{n}{3} \simeq 60$, se tendría que usar un tamaño de transformada igual a la primera potencia de 2 que sea mayor o igual que $181 + 2 \cdot 60 = 301$, o sea 512, con $M < 38$ sería 256. Para el caso de de 360 mediciones abarcando 360° habrá que tener en cuenta el relleno para simular el carácter circular de los datos, el relleno está constituido por M elementos en cada extremo de la serie de mediciones, por lo que con $M = \frac{n}{9} = 40$ para calcular el tamaño de la transformada tendremos $360 + 2 \cdot 40 + 2 \cdot 40 = 520$, luego necesitaríamos 1024 elementos para las transformadas. Al intentar reducir M por debajo de 40 solía aparecer alguna semilla más.

No se encontraron diferencias apreciables al variar L entre 15 y 24 ni al variar M entre $\frac{n}{9}$ y $\frac{n}{3}$. Donde sí apareció alguna diferencia fue en el rango de valores de k en el límite del filtro $\bar{h}_C + ks_C$. Con $L = 24$, $M = \frac{n}{3}$ y valores de $k = 2, 2,5, 3$ los resultados no mostraban diferencias apreciables entre las posiciones y el número de semillas. Con $L = 15$, $M = \frac{n}{9}$ y $k = 3$ quedaba algún segmento sin semilla, que sí la tenía para los otros valores de k .

Dado que M , L y k son los parámetros del conjunto de procedimientos que pueden tener

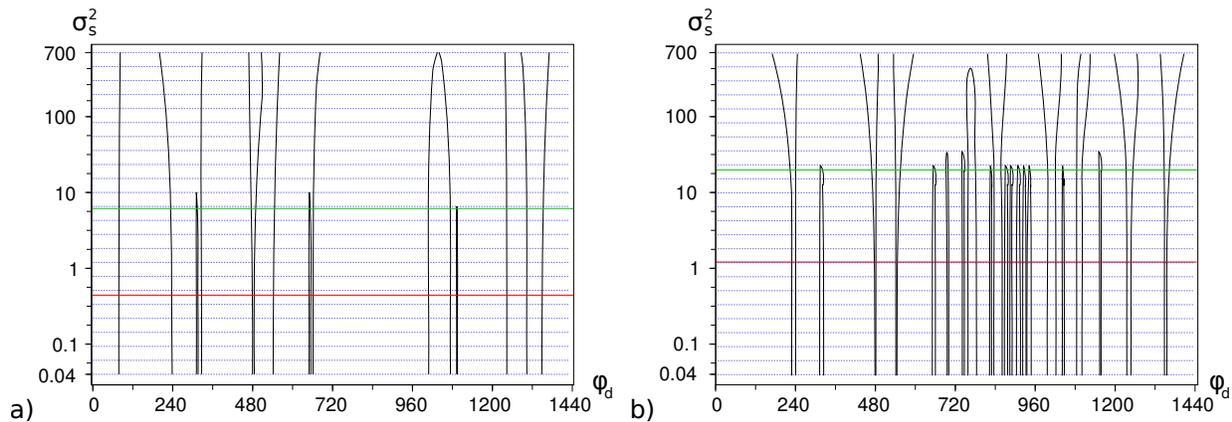


Figura 4.11: Curvas de nivel filtradas obtenidas empleando una resolución angular de $0,25^\circ$, 1440 mediciones, y $\sigma = 0,002$. En a) las correspondientes a la primera diferencia y en b) a la segunda.

un efecto más determinante sobre el resultado, aparte de C , la posibilidad de su variación conjunta e independiente, dentro de los rangos comentados, sin variación apreciable en los resultados, nos permite pensar que el procedimiento global está dotado de una notable robustez.

La figura 4.11 muestra las curvas de nivel filtradas con $\bar{h}_C + 2,5s_C$, si comparamos las más amplias con las de la figura 4.8 parece que esta última figura hubiera sido cortada por aproximadamente el nivel 17, 3 por encima de la marca 10. El número de semillas obtenidas en este caso fue de 29, cuando el filtro se empleó con $\bar{h}_C + 3s_C$ se obtuvieron 23 semillas. Resulta evidente que para conseguir resultados semejantes a los de resolución angular de 1° será necesario aumentar el nivel máximo de la escala, aunque no necesariamente el número de niveles. Si deseamos ampliar la semi-amplitud máxima M proporcionalmente debería ser $M \simeq 530$ lo que implicaría que el nivel máximo de escala debería llegar a 11236 y a 70225 si llegamos a la décima de grado.

Aquí se plantea el problema comentado del desbordamiento, ver subsección 4.1.1.3, al calcular los valores de los núcleos T , ecuación 2.2. El problema está en el cálculo de la función de Bessel y, supuesto que este se pudiera realizar, queda el problema de que la exponencial que aparece en el núcleo se anula para valores un poco superiores a 700, cuando se emplean cuatro bytes de precisión para guardar un número real.

Una posible solución sería encontrar una aproximación de los núcleos para tamaños grandes cuya expresión evitar los productos de valores extremos, esta aproximación debería respetar la propiedad de descartar variaciones, ver subsección 4.1.1.1, o algunas de las aproximaciones mencionadas en esa subsección.

Otra posibilidad sería emplear un programa de aritmética de precisión infinita para precalcular las expresiones de los núcleos, este precálculo no representa realmente ninguna diferencia ya que en una implementación eficiente, todos los núcleos y sus transformadas directas se deben calcular al inicializar el comportamiento del robot ya que son constantes.

Una tercera posibilidad es emplear la propiedad de suavización en cascada de los núcleos

T , ver subsección 2.2.3.1 Este enfoque presenta dos problemas: que solo permite avanzar en la escala de forma lineal, debido a la presencia de $t_2 - t_1$ en la fórmula de la propiedad, cuando ésta lo hace de forma exponencial, y que implicaría no poder precalcular las transformadas de $F(\varphi_d, t_1)$, lo que disminuiría su eficiencia.

Capítulo 5

Extracción de líneas

5.1. Introducción

En la sección 2.1 se definió como extracción de líneas al problema de encontrar los extremos de los segmentos a la vez que se estiman los parámetros de las correspondientes rectas

En este capítulo mostraremos cómo aplicar la estimación de líneas rectas descrita en el capítulo para la extracción de líneas de un conjunto de medidas de rango, mediante un procedimiento que podemos calificar de local seguido otro podemos calificar de global. En el procedimiento local se estiman segmentos a partir de puntos semilla consecutivos añadiendo puntos, sucesivamente a cada lado de la semilla, hasta que la condición de outlier se alcanza por cada lado. De esta manera se obtienen segmentos que pueden estar solapados. El procedimiento global se ocupa de deshacer los posibles solapes, otorgando cada punto común al segmento que menos variabilidad presentó al añadir dicho punto en el procedimiento local. Para terminar se ejecuta un procedimiento que prueba a lo largo de la lista de segmentos si dos adyacentes pueden considerarse partes de un mismo segmento, uniéndolos en caso afirmativo.

El tamaño de las semillas debiera ser pequeño para que le de tiempo al filtro a capturar aproximadamente el comportamiento del ruido, de manera que los estadísticos de detección de outliers puedan detectar adecuadamente los que se producen al salir de los límites del segmento. Aunque si el tamaño es demasiado pequeño puede comprometer seriamente la exactitud de la estimación inicial que proporcione la semilla.

Comenzaremos el capítulo con la exposición del método para obtener estimaciones de la semilla para terminar se expondrán los procedimientos local y global comentados al principio de esta sección.

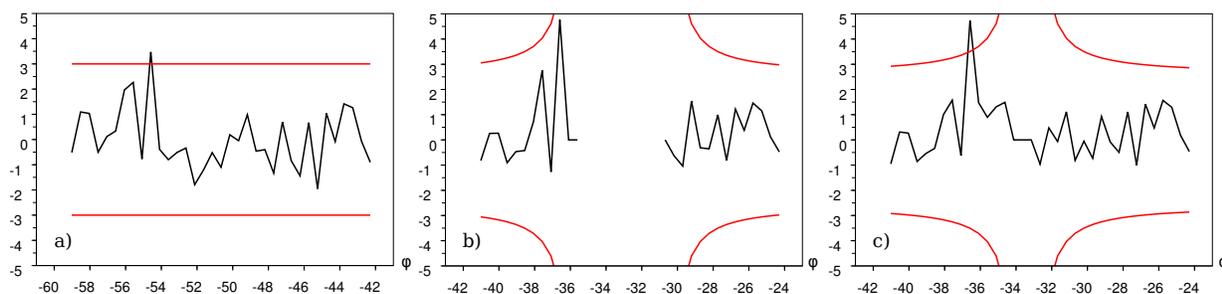


Figura 5.1: Mejora de la detección temprana de outliers al incluir la semilla en el filtro de Kalman. En a) el ruido base muestra un outlier en $\varphi = -55$. Al no incluir la semilla b) el estadístico de detección de outliers no alcanza el nivel crítico, debido al escaso número de grados de libertad. Cuando se incluye la semilla c) si se detecta el outlier.

5.2. Inicialización del filtro: el método de la mediana repetida

5.2.1. Razones para su empleo

El método de la mediana repetida Siegel (1982) es un método de estimación robusto de modelos de regresión lineal, no están sensible a la presencia de outliers como la regresión MCO. Se basa en la minimización de la suma de los residuos en valor absoluto en lugar de elevados al cuadrado.

La primera razón para la elección de este método para obtener una estimación inicial para el filtro de Kalman a partir de los puntos semilla ha sido su robustez, ya que se pretende emplear con semillas con pocos puntos. La segunda es que se pretende poder identificar los extremos de segmentos los más cortos posible y esto nos lleva a pensar si podríamos reutilizar los puntos semilla en la estimación del filtro de Kalman, a fin de que los grados de libertad de la distribución de los estadísticos de detección de outliers, t_{exp} , no sean demasiado pequeños cuando se sale de la semilla. De esta manera los valores críticos para la detección serían más pequeños y el filtro podría haberse estabilizado un poco más captando mejor el ruido y por lo tanto los outliers, ver figura 5.1. Claro que esto plantea un problema de independencia entre la estimación inicial y los puntos usados en el filtro, ya que los primeros ya han sido usados para la estimación inicial. Esta es la razón de emplear un procedimiento, como el de la mediana repetida, que no esté basado en la minimización de la *SCR*.

Para comprobar cómo afecta la inclusión de la semilla se volvieron a realizar los experimentos comentados en la subsección 3.5.3 incluyendo los puntos semilla, los resultados aparecen en la tabla 5.1. No parece haber una diferencia significativa entre los resultados para los errores de los parámetros entre las tablas 5.1 y 3.5, aunque sí en lo relativo a las d_t que parecen un poco mayores al incluir la semilla. Se tienen pues dos posibilidades incluir la semilla en incurrir en unas d_t algo mayores o no incluirla y no detectar algún outlier próximo a la semilla, o lo que es lo mismo un extremo de un segmento corto.

a)	media	min.	mediana	per. 95	per. 99	máx.	st. dev.
$E_{rel}(\hat{\rho})$	0,02539	0,00000	0,01363	0,08917	0,18929	0,67514	0,03733
$E_{abs}(\hat{\theta})$	0,05557	0,00000	0,01919	0,23665	0,38652	0,82385	0,08442
$E_{rel}(\hat{\sigma})$	0,24012	0,00000	0,12938	0,91137	1,65504	3,74228	0,33017
d_t	0,63573	0,14276	0,57154	1,16877	1,41657	1,99696	0,26459

b)	media	min.	mediana	per. 95	per. 99	máx.	st. dev.
$E_{rel}(\hat{\rho})$	0,00118	0,00000	0,00089	0,00328	0,00458	0,00992	0,00101
$E_{abs}(\hat{\theta})$	0,00190	0,00000	0,00111	0,00635	0,00977	0,02601	0,00219
$E_{rel}(\hat{\sigma})$	0,11614	0,00000	0,09392	0,29474	0,44961	1,46435	0,10055
d_t	0,519105	0,13696	0,49436	0,81934	0,98800	1,48792	0,15854

Tabla 5.1: Medidas de los errores cometidos al incluir la semilla. Se realizaron 32600 estimaciones de rectas de 35 puntos, 9 de semilla, rectas en a) con $\sigma = 0,02$ y en b) con $\sigma = 0,002$. Comparar estos resultados con los de la tabla 3.5.3

5.2.2. El tamaño de la semilla

El tamaño de la semilla está relacionado, como se ya se vió en 5.2.1, con el tamaño mínimo de segmento que es posible identificar mediante detección de outliers, con la precisión de la estimación inicial y con la complejidad del cálculo ya que los procedimientos deterministas practicable tiene una complejidad temporal de entre $O(n \log^3 n)$ y $O(n^2 \log^2 n)$. En esta sección se van a realizar experimentos para comprobar cómo afecta el tamaño de la semilla a la estimación inicial que proporciona. Contretamente calcularemos los errores que producen semillas de entre 3 y 15 puntos distribuidas a todo largo de una recta $\pm 88,5^\circ$, como en los experimentos de la subsección 3.5.3. Sólo se han estudiado los errores en los parámetros ρ y θ que son los que se usan en la estimación inicial, no los de σ ni las d_t . En la figura 5.2 a) y c) se muestran los errores relativos de la estimación de ρ , y en b) y d) los absolutos de la de θ , para $\sigma = 0,02, 0,002$ respectivamente, cuando la semilla tiene una longitud de nueve puntos. Al comparar estas gráficas con las correspondientes de las figuras 3.16 y 3.17, se aprecia que la distribución del error de la estimación inicial condiciona en gran medida la distribución del error de la estimación final. En las subfiguras e), f), g) y h) se observa la evolución de estos errores un función del tamaño de la semilla y de los valores habituales de σ .

5.3. Estimación de segmentos

5.3.1. Detección de extremos de segmentos

Antes de comenzar el procedimiento de estimación mediante filtro de Kalman se debe dar una condición para la detección de los extremos del segmento (break points). Existen dos tipos de outliers: outliers propios, puntos que pertenecen al segmento pero se halla un poco demasiado lejos de él, y outliers impropios, puntos que no pertenecen al segmento.

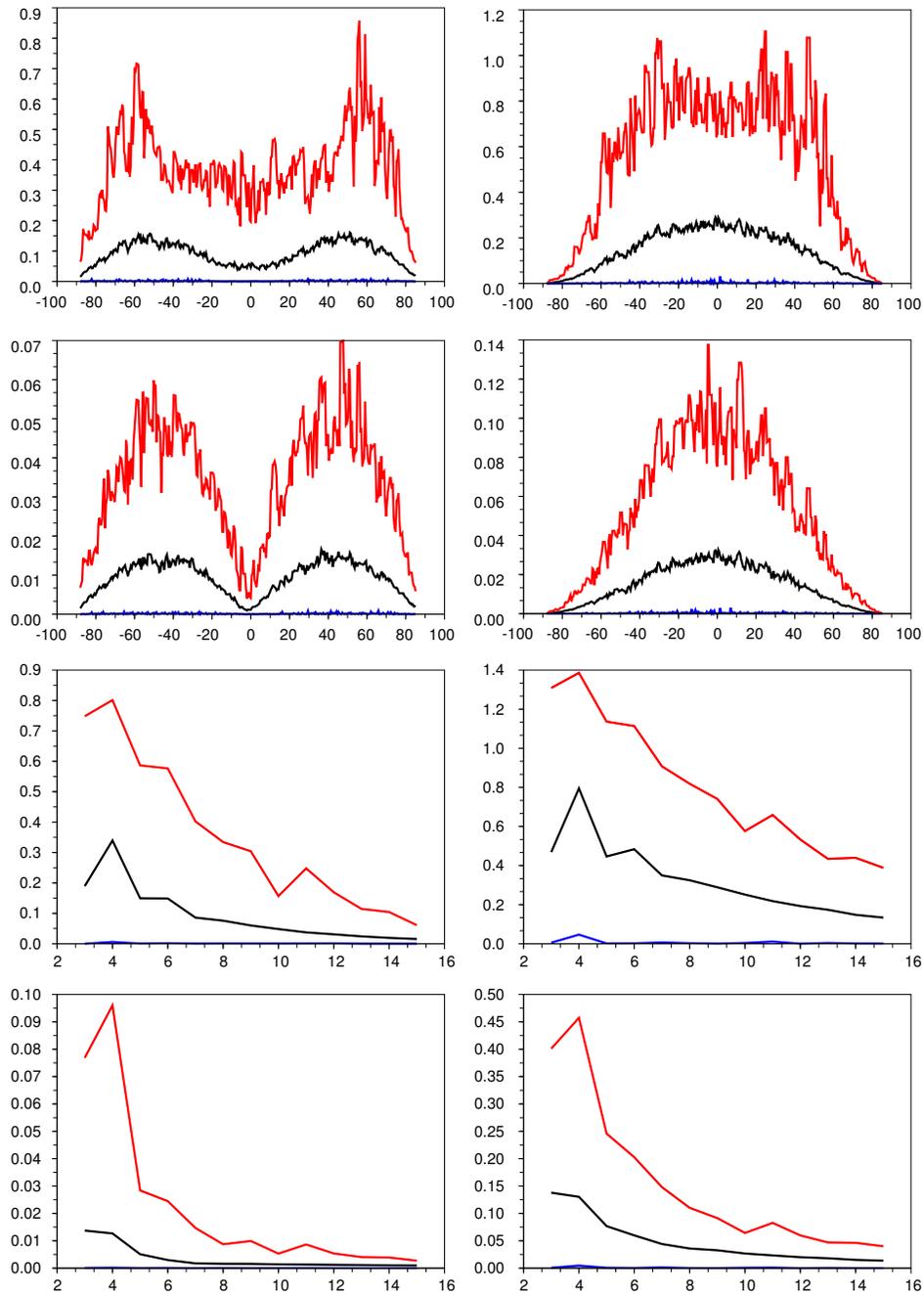


Figura 5.2: Errores en semillas en los parámetros en función del ángulo de incidencia, mitad superior, y en función del tamaño de la semilla, mitad inferior. Valores máximos (rojo), medios (negro) y mínimos (azul). Ver comentario en el texto.

Por esto la detección de un único outlier como condición de terminación, no parece adecuada. Se propuso como condición encontrar dos outliers consecutivos en un mismo lado del segmento, se aplica el procedimiento bilateral descrito en la subsección 3.5.3. Sin embargo esta condición resultó ser demasiado restrictiva. Como solución intermedia se propone como condición de detección de extremo de segmento la siguiente: encontrar dos outliers en dos o tres valores consecutivos. De esta manera si el nivel de significación para la detección de outliers es α , entonces $2\alpha^2$ es aproximadamente el nivel de significación para la detección de extremo de segmento.

Para determinar el fin del proceso de filtrado se deben encontrar extremos en ambos lados del segmento. Se han impuesto dos condiciones más para la finalización del proceso. La primera es que si se alcanza en principio de la semilla anterior o el final de la siguiente, caso de que existan, se detiene la estimación por el lado correspondiente. Esta condición se impone para evitar un solapamiento excesivo entre segmentos conectados cuando el ruido es grande. En el caso de que la apertura del haz no abarque 360° el primer y el último rayo actúan también como extremos de semillas. Segunda, si la estimación de la desviación típica del ruido excede un límite, el filtro es considerado como fuera de control, el límite se ha situado en 0.1, equivalente a 1 m. a 10 m.

Con respecto a la contribución de los outliers al proceso de estimación de los segmentos, se les atribuye un peso entre 0 y 1 para ponderar dicha contribución. En este trabajo se ha optado por usar el valor 0, o sea, no tenerlos en cuenta en los cálculos, aunque se ha llevado separadamente contabilidad de su contribución a la *SCR* para el procesos de resolución de solapes que se describe en la subsección 5.3.2.

5.3.2. Estimación y resolución de solapes

Tras la obtención de la estimación inicial a partir de las semillas se comienza el proceso del filtro de Kalman desde un punto central de cada semilla, inclusión de semilla, según el procedimiento bilateral descrito en la subsección 3.5.3, hasta encontrar la condición de finalización del segmento por cada lado del punto inicial.

Cuando todos los procesos de filtrado han concluido, se realiza una búsqueda de grupos de puntos que no pertenecen a ningún segmento, los denominaremos huecos, se determina una semilla en el interior de cada hueco y se realiza la estimación de los correspondientes segmentos mediante el filtro de Kalman.

Los segmentos pueden presentar solapes, por lo que es necesario un criterio para asignar un grupo consecutivo de puntos a cada segmento. Tras probar con distintas medidas estadísticas: coeficientes de determinación, valores de los estadísticos de detección, incremento de la varianzas estimadas del ruido, se observó que la comparación de las varianzas estimadas del ruido era la que ofrecía una mayor precisión en la resolución de solapamientos. Concretamente se calcula el signo de la diferencia entre las varianzas estimadas en un punto central del solapamiento y se busca el primer cambio de signo de dicha diferencia a ambos lados del punto central. Si se encuentra dicho punto se divide la región de solape, asignando a cada segmento la zona contigua a él. Si no existe dicho punto el segmento de menor varianza toma para sí la región de solape. Pudiera ser que un segmento fuera

absorbido completamente por el siguiente, en este caso se resolvería el posible solape entre el segmento anterior al absorbido y el que lo absorbió. Para haces de 360° hay que realizar el proceso circularmente, teniendo en cuenta que el primer segmento sigue al último. Antes y durante de la resolución de solapes mantiene ordenados los segmentos según su primer extremo, según el orden de los ángulos del haz, si dos segmentos comparten en primer extremo se sitúa primero el más corto.

Cuando el proceso de resolución de solapamientos concluye se realiza una nueva búsqueda de huecos, aunque en este caso la estimación del segmento se haya limitada al propio hueco. Tras este proceso todos los puntos quedan clasificados como pertenecientes a un segmento o como puntos aislados

5.3.3. Mezcla de segmentos

Especialmente cuando el ruido tiene un nivel elevado el procedimiento de clustering tiende a producir más de una semilla por cada segmento “verdadero” de manera que tras el procedimiento comentado en esta sección se hayan estimado más segmentos de los que en realidad existen. Por lo que es necesario un procedimiento de mezcla de segmentos semejantes. Como se ha comentado, al formular el filtro de Kalman de una manera equivalente a los modelos de regresión, podemos hacer uso de la teoría de dichos modelos y seleccionar un test de igualdad de modelos.

El test de Chow Johnston and DiNardo’s (1996), está diseñado para contrastar la igualdad de parámetros de un modelo cuando es estimado sobre conjuntos de datos distintos, el test permite contrastar la igualdad de un subconjunto cualquiera de parámetros. Para el caso que nos ocupa: dos conjuntos de datos (segmentos) y contrastar la igualdad de todos (dos) los parámetros del modelo de la recta la expresión de la hipótesis nula queda:

$$H_0 : \begin{aligned} \rho_1 &= \rho_2 \\ \theta_1 &= \theta_2 \end{aligned}$$

y la del estadístico de contraste:

$$F_{exp} = \frac{SCR - (SCR_1 + SCR_2)}{\frac{2}{SCR_1 + SCR_2}} \sim F(2, n - 4)$$

donde SCR se obtiene de la estimación usando los datos, n , de ambos segmentos y SCR_1 y SCR_2 de las estimaciones previas de cada segmento. La distribución del estadístico es una F de Snedecor con 2 y $n-2$ grados de libertad. Si para un determinado nivel de significación α se tiene que $F_{exp} < F_{\alpha, 2, n-4}$, siendo $F_{\alpha, 2, n-4}$ el valor crítico de la distribución para ese α , se consideran los segmentos iguales y se mezclan formando el segmento obtenido con los datos de ambos.

La validez de este contraste, como para todos los demás, disminuye al considerar modelos no lineales. El nivel de significación empleado en los experimentos ha sido el mismo que el resultante para la detección de extremos, ver subsección 5.3.1.

En la figura 5.3 se muestra un ejemplo, tomado de Cuadra Troncoso et al. (2011), del proceso descrito a lo largo de esta sección.

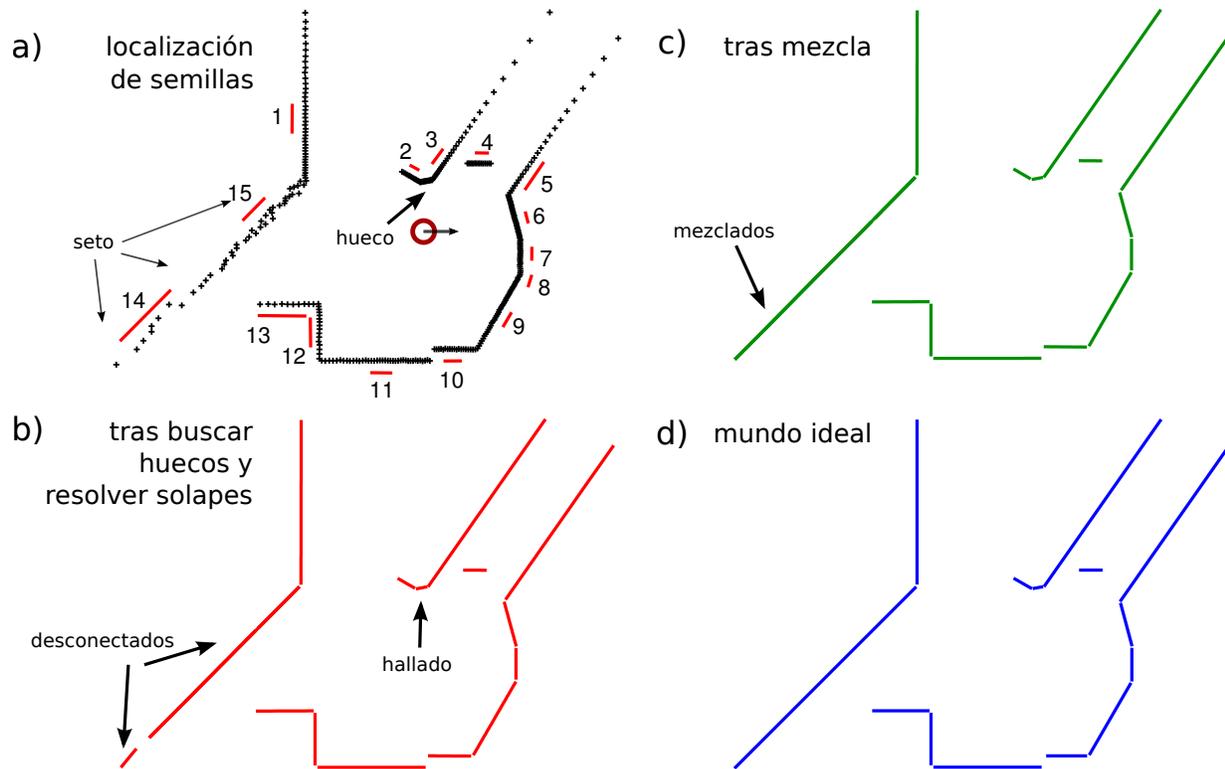


Figura 5.3: Secuencia de dibujos mostrando la aplicación del proceso de estimación, resolución de solapamientos y mezcla segmentos descrito en la sección 5.3. El mundo está compuesto de paredes planas y un seto donde el ruido es diez veces mayor. Las dimensiones del mundo son $17,5 \text{ m} \times 12,5 \text{ m}$, se han tomado medidas en 360° con resolución angular de 1° . El círculo rojo (negro) en a) representa al robot y la flecha en el centro del robot apunta a la orientación $\varphi = 0$. Se muestra el dibujo del mundo ideal, usado para generar las medidas, con propósito de comparación.

5.4. Experimentos sobre fiabilidad

En los experimentos que se describen a continuación se trata de comprobar la fiabilidad y eficiencia de método de segmentación propuesto. En Nguyen et al. (2007) se realiza una comparativa de distintos métodos de métodos de segmentación tanto desde el punto de vista de la fiabilidad y de la eficiencia y los procedimientos de comparación en él descritos han sido usados en trabajos posteriores Núñez et al. (2008). En los experimentos sobre fiabilidad se seguirán esos procedimientos adaptándolos al modelo de ruido propuesto en este trabajo, ver sección 3.3.2, y se investigará otro método de comparación.

En el modelo de ruido usado en estos experimentos la desviación típica del ruido se supone proporcional a la medición, la constante de proporcionalidad es la desviación típica del ruido base, y además es nuestro objetivo testear los procedimientos en presencia de diferentes niveles del ruido base para comprobar la robustez de los distintos métodos. Por lo que tendremos que generar los datos de forma sintética en lugar de trabajar con datos reales, para poder realizar pruebas sistemáticas.

Los procedimientos con los que se va a comparar el procedimiento propuesto son: Split and Merge (SM), el que obtuvo los mejores resultados en Nguyen et al. (2007) y los segundos mejores en Fernández et al. (2010) y es empleado en muchos procedimientos, y Random Sample Consensus (RANSAC). RANSAC es solo usado en el experimento inicial a modo de monitor, ver subsección 5.4.3, además cuando el ruido alcanzaba cierto nivel sus resultados carecían de significación debido al bajo número de emparejamientos y su observación en las gráficas podía inducir a confusiones.

La implementación de SM usada es la contenida en el simulador CARMEN y la de RANSAC la contenida en el simulador RMPT, se usaron los valores por defecto dados en estas implementaciones. Para SM CARMEN devuelve solo las coordenadas cartesianas de los extremos de los segmentos, para RANSAC MRPT devuelve el número de puntos de cada segmento y su ecuación en forma implícita.

5.4.1. Datos

Se han usado dos mundos ideales para generar las medidas. Uno con 180° de amplitud, M180, que se ha muestreado con una resolución angular de 1° , y otro con 360° de amplitud, que se ha muestreado con resoluciones de 1° y $0,5^\circ$, los denominaremos M360 y M720 respectivamente. Los segmentos que conforman M360 y M720 se pueden ver en la figura 5.3 d), una estimación de M180, indistinguible a simple vista del original, puede verse en la figura 5.6 a). Sobre los segmentos se generó ruido gaussiano con desviación típica $\sigma_i = \sigma E[r_i]$, para valores de $\sigma \in (0, 0,025]$, el primer valor fue 0.0001, el segundo 0.001 y de milésima en milésima hasta 0.025. Para este último valor tendríamos unos errores típicos dentro un rango aproximado de ± 75 cm. a 10 m. Para cada valor de σ se generaron 5000 series aleatorias de mediciones.

5.4.2. Estimación de segmentos por Mínimos Cuadrados Totales Generalizados Factibles

Aunque la implementación usada de RANSAC daba las estimaciones de los parámetros de la recta, pero no la SM, no daba las de las matrices de covarianzas por lo que se decidió descartar estas estimaciones. La estimación de los segmentos obtenidos mediante SM y RANSAC se ha llevado a cabo, como en Nguyen et al. (2007), por el método de los Mínimos Cuadrados Totales (MCT), ver subsección 2.4.2.

El procedimiento elegido es el descrito en Arras and Siegart (1997), que es el usado en Nguyen et al. (2007), ya que este método ha sido diseñado para matrices de covarianzas del ruido diagonales, heterocedasticidad, como es el caso del método propuesto. Es un tipo de MCT Ponderados, aunque en dicho trabajo los pesos no dependen del valor de las mediciones. El problema estriba en que dicha matriz debe ser conocida para poder calcular los pesos, en Arras and Siegart (1997) calcula experimentalmente dicha matriz. En nuestro caso, dado que los pesos obedecen a un modelos distinto y a la que se desean estudiar diversos niveles de ruido nos pareció oportuno realizar una modificación en el algoritmo que estamos comentando para adaptarlo a la situación descrita.

Siguiendo la técnica descrita en la subsección 2.4.1.2 para corregir la heterocedasticidad (MCG factibles), que fue usada en el filtro de Kalman propuesto, es necesario dar en primer lugar un estimador de la varianza del ruido. A continuación se va a sugerir un posible estimador para dicha varianza y su distribución, basándonos en los obtenidos en la teoría clásica, para el caso de estimación de líneas rectas obtenidas por mediciones radiales.

Supongamos por un momento en el caso de varianza del ruido constante, para estimar dicha varianza parece lógico usar los errores definidos así $e_i = r_i - \hat{r}_i^{MCT}$, siendo \hat{r}_i^{MCT} la estimación de r_i por MCT, análogamente a como se hace en el filtro de Kalman propuesto. Definimos entonces $SCR_{MCT} = \sum e_i^2$. Recordemos que la fórmula del estimador insesgado de la varianza del ruido y su distribución en la teoría clásica, para dos parámetros, son $\hat{\sigma}^2 = \frac{SCR}{n-2} \sim \chi^2(n-2)$, dos es el número de parámetros o también el número de derivadas igualadas a 0 al minimizar. En MCT en la versión elegida se realizan tres derivadas, con respecto a los dos parámetros y al cociente entre las varianzas de x e y , por lo que sugerimos como estimador de dicha varianza y distribución aproximada

$$\hat{\sigma}_{MCT}^2 = \frac{SCR_{MCT}}{n-3} \sim \chi^2(n-3) \quad (5.1)$$

Queda como posible trabajo futuro la investigación de las propiedades del estadístico propuesto, su uso en este trabajo queda justificado por los experimentos realizados, ver subsección 5.4.4. Dicho experimentos muestran que mientras SM realiza un clustering correcto, los resultados sobre número de varianzas correctamente estimadas no difiere apreciablemente entre los obtenidos aplicando el método propuesto y SM+MCT, y hay bastante similitud entre número de segmentos emparejados y el de varianzas correctas en el propio SM+MCT.

Una vez obtenido un estimador de σ^2 se puede aplicar la técnica de dos etapas, MCG factibles, usada en el filtro de Kalman, calculando los pesos a partir de las estimaciones

\hat{r}_i^{MCT} . Por esto denominaremos a esta técnica Mínimos Cuadrados Totales Generalizados Factibles (MCTGF). Así siguiendo el procedimiento descrito en Arras and Siegart (1997) se realizará un primera etapa en la que se estimarán los parámetros usando pesos $w_i \equiv 1$, seguida de una segunda en la que los pesos serán $w_i = \frac{1}{\hat{r}_i^{MCT}}$ siendo \hat{r}_i^{MCT} las estimaciones obtenidas en la primera etapa, la estimación de sigma se realiza a partir de los errores de la segunda etapa.

Las implementaciones usadas de SM y RANSAC no contemplaban la posibilidad de posible unificación del primer y el último segmento en caso de amplitud de 360° , por lo que se añadió el procedimiento para unirlos descritos en la subsección 5.3.3.

5.4.3. Criterios de comparación

Unos de los puntos clave al comparar métodos de segmentación es el número de segmentos correctamente estimados (emparejados), o sea, con una buena aproximación a los verdaderos parámetros y a los verdaderos extremos. En la metodología de Nguyen et al. (2007) no se menciona la cuestión de los extremos, así que se ha adoptado por realizar la comparación de parámetros cuando el segmento estimado y el real comparten al menos un 25% de sus puntos y además medir la cantidad de puntos de la intersección de ambos segmentos así como su longitud. Las magnitudes usadas en la comparación se enumeran en la tabla 5.2 y pasamos a comentarlas a continuación.

Magnitudes para evaluar la corrección del procedimiento.

- NSE , número de segmentos emparejados por el criterio estadístico, ver más abajo, a lo más uno por segmento real. Este valor mide la adecuación de las estimaciones de los parámetros de la línea, no la correcta situación de los extremos del segmento. A partir de él se obtiene $SE = NSE/NSR$ la proporción de segmentos emparejados, siendo NSR el número de segmentos reales.
- $NSNE$, número de segmentos no emparejados = número de segmentos descubiertos - NSE . Este valor mide tanto los fallos en las estimaciones como la fragmentación, o sea el descubrimiento de más segmentos que los reales. A partir de él se obtiene $SNE = NSNE/NSR$ la proporción de segmentos no emparejados. Se ha preferido usar en el denominador NSR en lugar de NSE , como es habitual, porque de esta manera se mide mejor la fragmentación lo que es de especial interés cuando el ruido es elevado. Valores de $SNE > 0,5$ pueden indicar una fragmentación elevada.
- NPE , número puntos en la intersección entre un segmento emparejados y su correspondiente segmento real. Nos da una medida de la buena situación de los extremos de los puntos. A partir de él se obtiene la proporción de puntos asociados a su verdadero segmento $PE = NPE/NTP$, siendo NTP el número total de puntos. Si $PE \geq SE$ es indicativo de que se están hallando con bastante corrección los extremos de los segmentos reales, para los segmentos emparejados.

NSE	Núm. segmentos emparejados
$NSNE$	Núm. segs. no emparejados = Núm. segs. descubiertos - NSE
NPE	Núm. puntos en intersección emparejados con reales
TLE	Longitud total de intersección de emparejados con reales
$Error \rho$	Desv. típica del error en la estimación de ρ en emparejados, cm.
$Error \theta$	Desv. típ. del error en la estimación de θ en emparejados, grados
NVC	Núm. segmentos con varianza de ruido base bien estimada

Tabla 5.2: Magnitudes usadas en la comparación entre procedimientos de segmentación.

- TLE , longitud total de intersección de emparejados con reales. A partir de él se obtiene la proporción $LE = TLE/LR$ de siendo LR la longitud total de los segmentos reales, o sea la proporción de mundo correctamente estimada. Además de lo comentado para NPE , TLE mide el rendimiento ante ángulos de incidencia elevados, cuando puntos consecutivos de un segmento pueden distanciarse bastante entre sí.
- NVC número segmentos con varianza de ruido base, σ^2 , bien estimada. A partir de él se obtiene la proporción de varianzas correctamente estimadas $VC = NVC/NSR$. Ver más abajo sobre el contraste de varianzas.

Magnitudes de para evaluar la exactitud del procedimiento.

- $Error \rho$, desviación típica del error en la estimación de ρ en segmentos emparejados medida en centímetros.
- $Error \sigma$, desviación típica del error en la estimación de σ en segmentos emparejados medida en grados.

5.4.3.1. Contrastes de parámetros

En la teoría de la regresión se pueden dar dos situaciones a la hora de contrastar la hipótesis de si los parámetros de la recta toman unos valores determinados:

$$H_0 : \begin{array}{l} \rho = \rho_0 \\ \theta = \theta_0 \end{array}$$

Si la matriz de varianzas y covarianzas de los estimadores, Σ es conocida entonces se tiene la siguiente distribución, asintótica en el caso no lineal:

$$\chi_{exp}^2 = (\rho - \rho_0, \theta - \theta_0) \Sigma^{-1} (\rho - \rho_0, \theta - \theta_0)' \sim \chi^2(2) \quad (5.2)$$

si $\chi_{exp}^2 < \chi_{\alpha,2}^2$ se acepta H_0 a un nivel de significación α .

Si la matriz de varianzas y covarianzas de los estimadores, Σ es desconocida entonces se tiene la siguiente distribución, asintótica en el caso no lineal:

$$F_{exp} = \frac{(\rho - \rho_0, \theta - \theta_0) \hat{\Sigma}^{-1} (\rho - \rho_0, \theta - \theta_0)'}{2} \sim F(2, n - 2) \quad (5.3)$$

si $F_{exp} < F_{\alpha,2,n-2}$ se acepta H_0 a un nivel de significación α .

La primera cantidad que puede ser desconocida en la expresión de Σ es σ^2 , ver ecuación 2.13 y sección 2.4.1.3, este estudio va dirigido hacia esa suposición, la segunda es Q , que depende del modelo de ruido.

La primera opción es la usada en Nguyen et al. (2007), la expresión que aparece en la ecuación 5.2 es conocida como distancia de Mahalanobis Mahalanobis (1936), usando como valor crítico $\chi_{0,25,2}^2 = 2,77$. La expresión de Σ en Nguyen et al. (2007) es hallada “a mano” a partir de 100 scans con datos reales obteniendo

$$\Sigma = \begin{pmatrix} 0,03^2 & 0 \\ 0 & 0,03^2 \end{pmatrix}$$

para las líneas reales. En nuestro caso, datos sintéticos, las líneas reales se conocen con total exactitud, así $\Sigma = 0$. por lo que se buscó un valor para la diagonal de Σ , que ofreciera resultados semejantes en cuanto a la proporción de emparejados en SM y RANSAC a los obtenidos en Nguyen et al. (2007) para los datos del “Intel-Laboratory dataset” cuando $\sigma = 0,002$, un valor similar al contemplado en Nguyen et al. (2007), el valor encontrado fue $0,01^2$. Para esta búsqueda es para lo único que se ha utilizado RANSAC. Cuando σ varía, la matriz se escala según el cuadrado de la proporción en la que varía σ con respecto a $0,002$, o sea:

$$\Sigma_{\sigma} = \left(\frac{\sigma}{0,002} \right)^2 \begin{pmatrix} 0,01^2 & 0 \\ 0 & 0,01^2 \end{pmatrix}$$

Dado que la matriz es diagonal y el valor crítico constante, el criterio es equivalente a fijar unos límites de tolerancia para ρ y θ , dependientes de σ de forma proporcional, sin tener en cuenta la posible relación entre los errores en ambos parámetros.

La segunda opción de contraste parece no ser necesaria en nuestra situación, conocemos la varianza, pero el suponer la matriz diagonal no parece un poco restrictivo. Además en este enfoque podemos decir que es el segmento quien acepta o no a los parámetros, esto depende de su matriz de varianzas lo que puede llevar a aceptar errores grandes, en especial si el ruido es elevado, haciendo que el procedimiento agrupe puntos de segmentos contiguos no concurrentes. Si los resultados según ambos criterios no difieren significativamente se tendrá una prueba de que el procedimiento, en cuanto a detección de extremos, tiene un comportamiento correcto y robusto frente al ruido. El nivel de significación se eligió bastante pequeño, $0,001$, ya con este valor se obtenían proporciones de emparejamientos para los valores pequeños del ruido similares a los del otro criterio.

5.4.3.2. Contraste de varianza

El criterio para contrastar la correcta estimación de la varianza del ruido base, σ^2 , usa la distribución del estimador correspondiente, ver subsección 5.4.2, para obtener el valor crítico $\chi_{\alpha,gd}^2$ donde gd son los grados de libertad, la estimación se acepta si $\chi_{exp}^2 < \chi_{\alpha,gd}^2$.

5.4.4. Resultados

Los resultados se van a ofrecer de forma visual: gráficas representando la magnitudes descritas en 5.4.3 y ejemplos de los resultados de reconstrucciones concretas sobre el conjunto de mediciones representado en coordenadas cartesianas, figuras de la 5.4 a la 5.13. Dado que es difícil comprender en detalle las relaciones entre los comportamientos de las diversas gráficas, los ejemplos dan una idea de lo que representan los valores de dichas gráficas, en especial muestran el grado de fragmentación. Las tablas de resultados incluidas en los ejemplos están confeccionadas aplicando el criterio de emparejamiento F , por ser el que muestra una mayor casuística, hemos preferido no mostrar ejemplos usando el criterio χ^2 ya que se incluirían figuras que visualmente no ofrecerían diferencias apreciables con las obtenidas usando el criterio F .

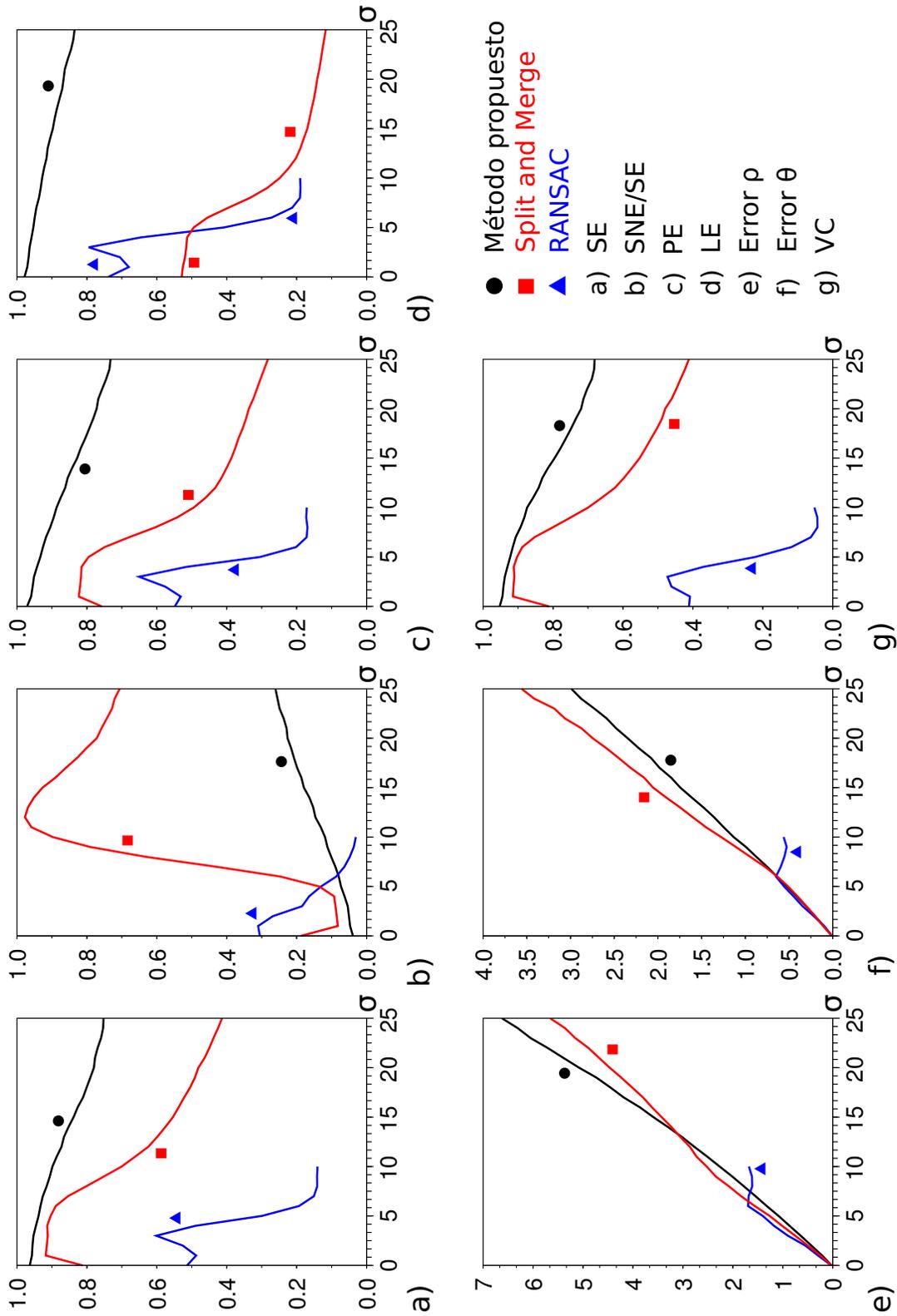
Los resultados en los mundos M180, M360 y M720 no difieren cualitativamente entre sí aunque hay algunas diferencias cuantitativas no muy acusadas, unas debidas a la forma de los mundos y otras a la resolución angular. Los resultados, en M180 en especial, para valores alrededor de $\sigma = 0,002$ concuerdan, en general, con los obtenidos en Nguyen et al. (2007) salvo por el número de segmentos no emparejados, en nuestros experimentos se obtienen valores mayores. Esto puede ser debido a que en nuestros mundos se incluyen segmentos que forma ángulos pequeños entre sí, lo que dificulta su separación, sin embargo los entornos de prueba de Nguyen et al. (2007) estaban formados principalmente por paredes en ángulo recto.

En todas las figuras en las que se representan las gráficas de la evolución de magnitudes se aprecia claramente diferencias en el comportamiento del método propuesto y los otros dos. Mientras que la gráficas correspondientes al método propuesto se observa un comportamiento casi lineal en los otros métodos se observan dos estados distintos: uno cuando el nivel de ruido es bajo y otro cuando el nivel es alto. Pasamos ahora a describir los resultados en detalle. En el eje de abscisas se representa σ con unidades de cm. a 10 m., habría que dividir por 1000 para pasar a la forma de representación usada en el resto del trabajo.

5.4.4.1. M180

La figura 5.4 muestra los resultados de las medidas de corrección y exactitud para el mundo M180 usando el criterio de emparejamiento basado en la χ^2 , ec. 5.2, ver características de M180 en subsección 5.4.1. En a) se observa la evolución de la proporción de emparejamientos, SE , para los niveles de ruido pequeño los valores para SM+MCT y RANSAC rondan los valores 0,9 y 0,6 similares a los obtenidos en Nguyen et al. (2007), lo que nos permite pensar en que existirá cierta correspondencia entre los resultados de ambos trabajos. También los errores en los parámetros son semejantes.

Método propuesto con contraste χ^2 Cuando el nivel de ruido es bajo el método propuesto ofrece para SE , figura 5.4 a), unos valores superiores 0,95 para los primeros niveles de ruido y este valor va decreciendo de forma casi lineal hasta situarse un poco



SE (prop. segs. emparejados) SNE (prop. segs. no emparejados) PE (prop. puntos emparejados)
 LE (prop. longitud total emparejada) Error ρ (desv. típica del error en ρ , cm.)
 Error θ (desv. típica del error en θ , gr.) VC (prop. estimaciones correctas de la varianza del ruido)

Figura 5.4: Resultados según el criterio χ^2 de comparación de parámetros, ec. 5.2, para M180. En el eje de abcisas aparece σ en unidades de cm. a 10 m.

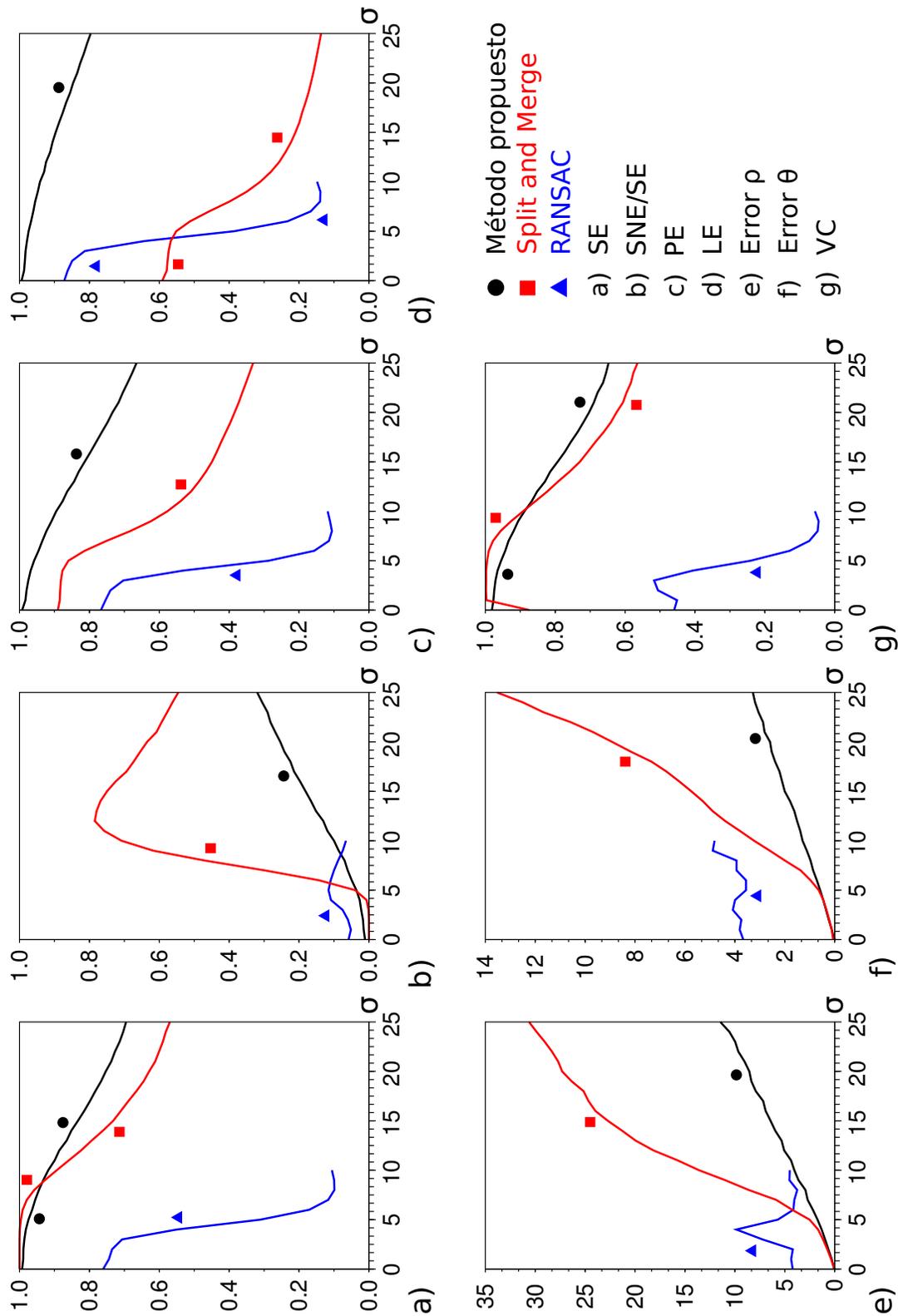
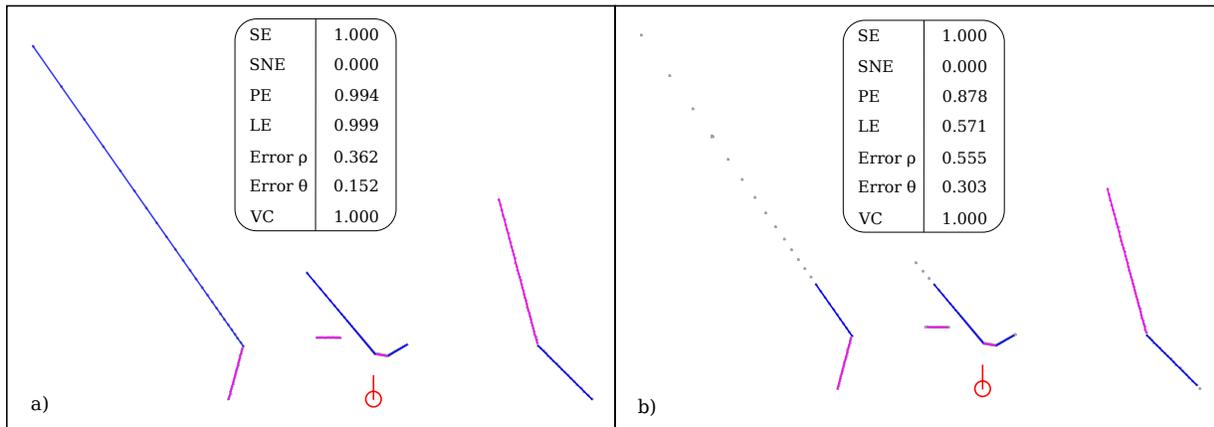


Figura 5.5: Resultados según el criterio F de comparación de parámetros, ec. 5.3, para M180. En el eje de abcisas aparece σ en unidades de cm. a 10 m.

SE (prop. segs. emparejados) SNE (prop. segs. no emparejados) PE (prop. puntos emparejados)
 LE (prop. longitud total emparejada) Error ρ (desv. típica del error en ρ , cm.)
 Error θ (desv. típica del error en θ , gr.) VC (prop. estimaciones correctas de la varianza del ruido)

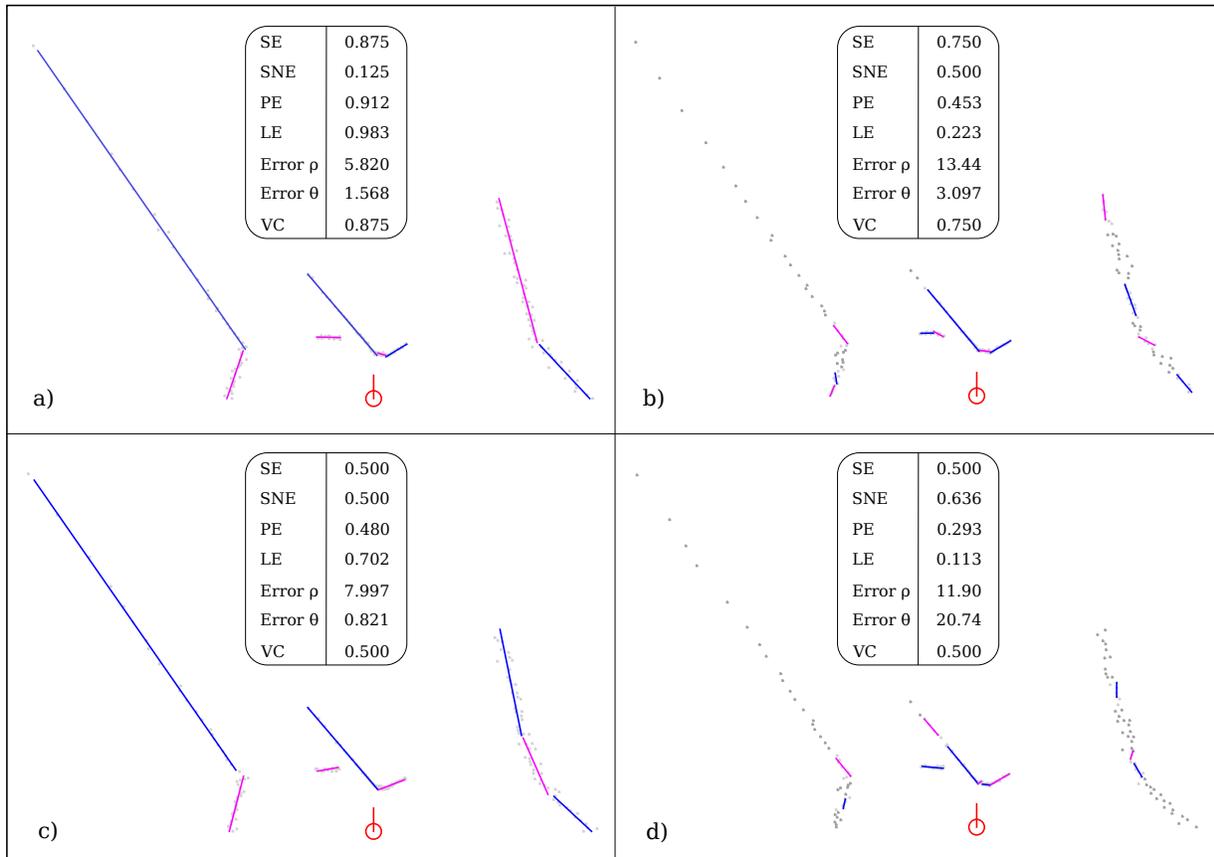


SE (prop. segmentos emparejados) SNE (prop. segmentos no emparejados) PE (prop. puntos emparejados)
 LE (prop. longitud total emparejada) Error ρ (desviación típica del error en ρ , cm.)
 Error θ (desviación típica del error en θ , gr.) VC (prop. estimaciones correctas de la varianza del ruido)

Figura 5.6: Reconstrucción de M180 cuando el nivel de ruido es pequeño, $\sigma = 0,002$. En a) el método propuesto y en b) SM+MTC. Los resultados de las magnitudes de corrección y exactitud aparecen en tablas, se muestran ejemplos de estimaciones con SE superior a la media. A la vista de los resultados el método propuesto detecta casi perfectamente los extremos de los segmentos, también es muy preciso en las estimaciones. SM+MCT se deja puntos sin asociar en las zonas que le ofrecen un mayor ángulo de incidencia, de manera que sólo empareja un 57,1% de la longitud total, los errores en los parámetros parecen un poco mayores en SM+MCT aunque tienen una buena precisión.

por debajo de 0,8. El valor de SNE , 5.4 b), cambia de forma lineal con el nivel de ruido desde un valor algo inferior a 0,05 hasta aproximadamente 0,25. El valor de PE , 5.4 c), para niveles bajos de ruido es muy similar al de SE , lo que indica que se están hallando los extremos de los segmentos con una precisión adecuada, con niveles de ruido alto PE tiende a ser ligeramente inferior a SE , lo que implica una ligera pérdida de precisión. Los valores de LE , 5.4 d), son superiores al los de PE lo que indica que el método responde adecuadamente ante ángulos de incidencia elevados, no dejando puntos sin clasificar. El error en ρ , 5.4 e), varía de forma casi lineal desde prácticamente 0 cm. hasta un poco por debajo de 7 cm. El error en θ , 5.4 f), varía de forma casi lineal desde prácticamente 0° hasta aproximadamente $3,5^\circ$. VC , 5.4 g), tiene un comportamiento muy similar al de SE , aunque disminuye un poco más acusadamente cuando el ruido es alto, esto indica que se está estimando adecuadamente la varianza del ruido para los segmentos emparejados. Estos resultados numéricos confirman el adecuado comportamiento del método ante el ruido así como su buena corrección y precisión.

En la figura 5.6 a) se observa una reconstrucción del mundo lograda con nivel de ruido bajo, la reconstrucción es prácticamente exacta. En la figura 5.7 a) y c) se observan dos reconstrucciones del mundo logradas con nivel de ruido alto, al exhibir el método una degradación casi lineal en el rendimiento y no muy acusada, sólo puede aparecer algo de fragmentación y/o no descubrir algún segmento como en c).



SE (prop. segmentos emparejados) SNE (prop. segmentos no emparejados) PE (prop. puntos emparejados)
 LE (prop. longitud total emparejada) Error ρ (desviación típica del error en ρ , cm.)
 Error θ (desviación típica del error en θ , gr.) VC (prop. estimaciones correctas de la varianza del ruido)

Figura 5.7: Reconstrucción de M180 cuando el nivel de ruido es grande, $\sigma = 0,02$. En a) y c) el método propuesto y en b) y d) SM+MTC. Los resultados de las magnitudes de corrección y exactitud aparecen en tablas, se muestran ejemplos de estimaciones con SE superior a la media a) y b) e inferior a la media c) y d). A la vista de los resultados la segmentación realizada por el método propuesto son mucho mejores que para SM+MTC, este algoritmo solo segmenta bien la zona situada frente al robot, donde los obstáculos están mas cerca y por lo tanto el efecto del ruido es menor. En los segmentos más alejados se produce una excesiva fragmentación y se pierden grupos numerosos de puntos. Aún en el mejor ejemplo la proporción de longitud emparejada no llega al 25 %, mientras que en el peor caso del método propuesto es del 70 %. El método propuesto, aún en el peor ejemplo con un valor de $SE = 0,5$, ofrece a simple vista una reconstrucción completa y creíble del mundo.

SM+MTC con contraste χ^2 Cuando el nivel de ruido es bajo el método propuesto SM+MCT ofrece resultados para *SE*, figura 5.4 a), muy similares a los obtenidos por el método propuesto, aunque ligeramente inferiores. Alrededor del valor de $\sigma = 7$ se observa en comienzo del cambio de estado en SM, que pasa de 0,9 a 0,6 rápidamente, a partir de ahí el descenso es menos acusado hasta llegar a un valor final inferior a 0,5. *SNE*, 5.4 a), presenta también el cambio de estado pasando de valores algo inferiores a 0,01 hasta aproximadamente 1 aunque luego desciende hasta situarse en torno a 0,75, parece que está apareciendo una fragmentación excesiva. *PE*, 5.4 a), muestra para niveles bajos de ruido valores algo inferiores a 0,8 pasa bruscamente a 0,5 para finalizar en torno a 0,3. Estos valores son bastante inferiores a los de *SE* por lo que el algoritmo no esa detectando adecuadamente los extremos de los segmentos. *LE* tiene valores iniciales en torno a 0,5 para terminar en torno a 0,2, el método falla ante ángulos de incidencia elevados, dejando zonas extensas sin clasificar. El error en ρ , 5.4 e), varía de forma casi lineal desde prácticamente 0 cm. hasta un poco por debajo de 6 cm. El error en θ , 5.4 f), varía de forma casi lineal desde prácticamente 0° hasta aproximadamente 3° . El comportamiento en cuanto a precisión es muy similar al del método propuesto.

En la figura 5.6 b) se observa una reconstrucción del mundo lograda con nivel de ruido bajo, las zonas con ángulo de incidencia alto quedan sin descubrir. En la figura 5.7 a) y c) se muestran dos reconstrucciones del mundo logradas con nivel de ruido alto, se observa una elevada fragmentación así como amplias zonas sin descubrir. Los segmentos emparejados con ruido base alto son los que están situados en la parte frontal del robot los más cercanos al él, y donde, por lo tanto, el ruido efectivo es pequeño. SM no parece resistir bien el ruido.

RANSAC con contraste χ^2 Recordemos que el papel de RANSAC en este estudio es simplemente para ajustar la matriz de varianzas Σ , así que solo realizaremos un breve comentario. RANSAC como SM muestra los dos mismos estados, aunque la transición ocurre antes y es más brusca llegando rápidamente a valores muy bajos de emparejamiento y de las demás medidas de corrección. RANSAC no resiste al ruido, como se comenta en Nguyen et al. (2007).

Método propuesto con contraste F El comportamiento de las medidas de corrección representadas en la figura 5.5 a), b), c), d) y g) es muy similar al que muestran en la figura 5.4, con quizás un ligero aumento en los valores finales y una ligera disminución en los finales, salvo *SNE* también se comporta de forma similar a la de la figura 5.4 b) aunque con una ligera disminución de los valores iniciales y un ligero aumento de los valores finales. El comportamiento de las medidas de precisión, e) y f), muestra un ligero aumento del error para valores grandes del ruido en comparación con los resultados mostrados en la figura 5.4 e) y f).

Al margen de la comparación con los otros métodos, esta similitud en los resultados indica una muy buena correspondencia entre ambos criterios de emparejamiento para el método propuesto. En otras palabras, lo que la especificación de los límites de tolerancia, a

la que equivale el criterio χ^2 , exige al método para aceptar una estimación de un segmento, corresponde muy aproximadamente a lo que el método “cree” que le exige la especificación, que es lo que da el criterio F , aunque no conoce dichos límites.

SM+MCT con contraste F En comportamiento de las medidas de corrección representadas en la figura 5.5 a), b), c), d) y g) es similar cualitativamente al que muestran en la figura 5.4, hay dos estados, pero cuantitativamente hay diferencias significativas. Los valores iniciales de las medidas aumentan aproximadamente en 0,1 aunque los finales son similares. *SNE*, 5.5 b), tiene un máximo algo menos elevado y valores iniciales menores que en 5.4 b).

Donde se aprecia una diferencia muy notable es en el comportamiento de las medidas de precisión, cuando comienza el cambio de estado se empiezan a disparar los errores en las estimaciones, la “creencia” del segmento no se corresponde con la especificación de límites de tolerancia. Mientras el ruido es pequeño los errores de estimación son muy similares a los del método propuesto.

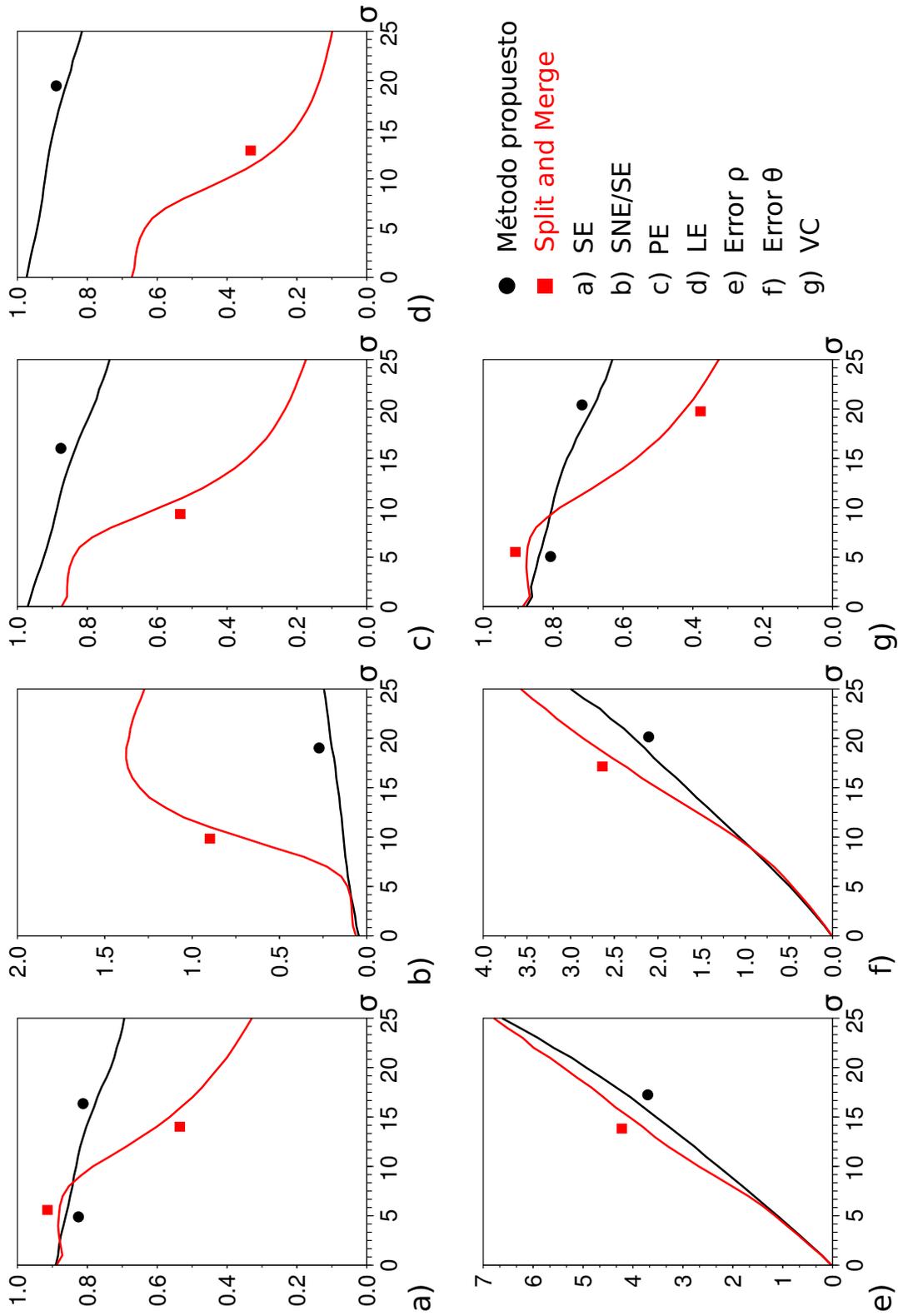
Para comentar ejemplos de reconstrucciones mostradas en las figuras 5.6 y 5.7, nos remitimos al epígrafe “Método propuesto con contraste χ^2 ”, unos párrafos más arriba en esta subsección.

RANSAC con contraste F RANSAC muestra diferencias similares a las comentadas para SM+MCT en el comportamiento de las medidas de corrección al que mostraba cuando se usaba el contraste χ^2 , sigue mostrando escasa resistencia al ruido. A partir de ahora dejará de ser utilizado en los experimentos, ya que sus resultados no son relevantes para el objetivo de evaluar nuestro método y además ralentizaba las pruebas ya que se llegaba a alcanzar numerosas veces el límite de intentos fijado en la implementación, 2000, cuando el ruido empezaba a ser algo elevado.

5.4.4.2. M360

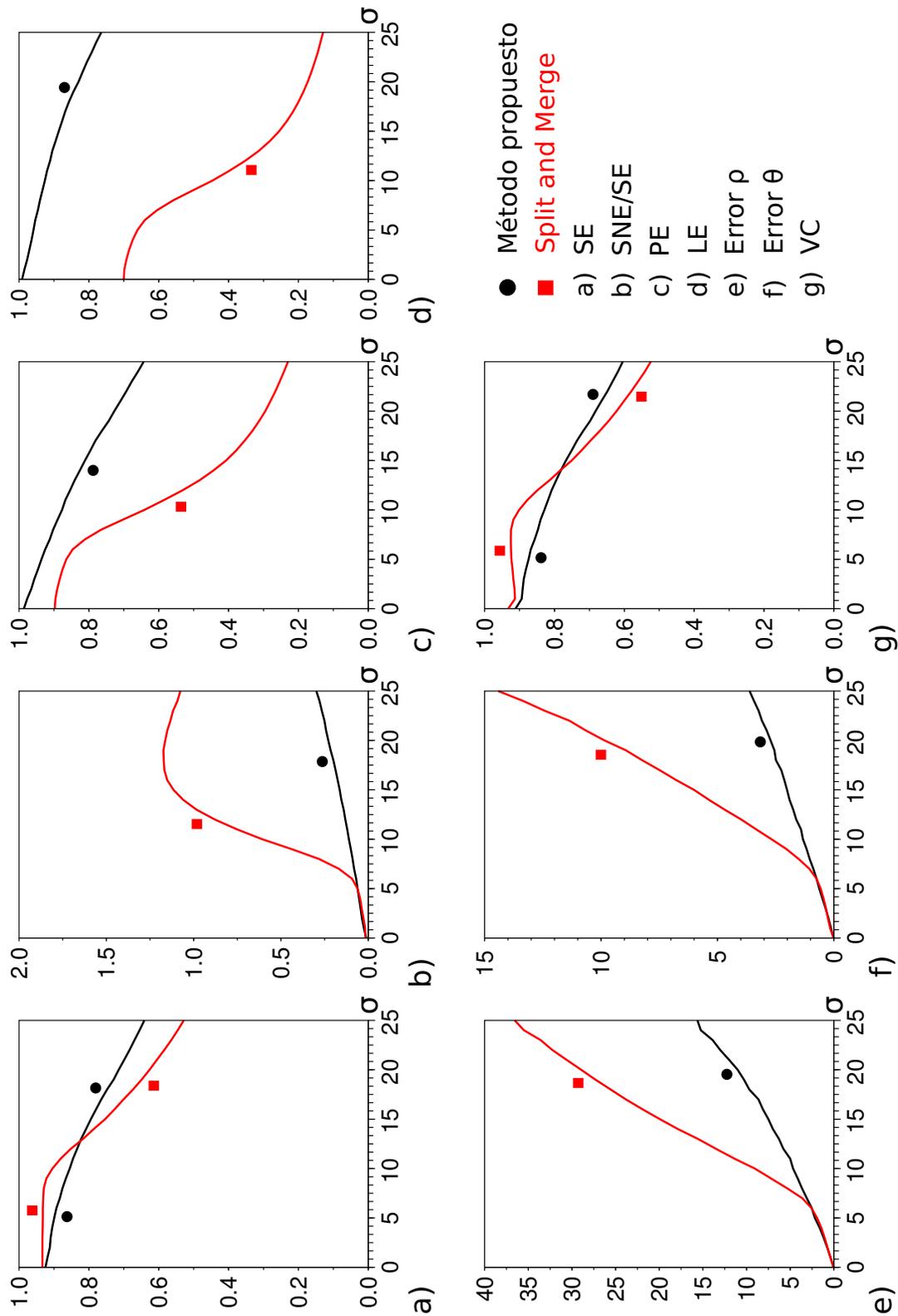
Los resultados para M360 obtenidos por el método propuesto y por SM+MCT no difieren significativamente de los obtenidos para M180, comentados en la subsección 5.4.4.1. Por lo que sólo comentaremos algunas diferencias que pudieran resultar algo más apreciables. Al observar la figura 5.9 a) podría pensarse que SM+MCT tiene un mejor rendimiento que el método propuesto para valores de σ en torno a 10, pero como puede apreciarse en 5.9 b), e) y f) la fragmentación elevada y los errores en las estimaciones empiezan a dispararse aproximadamente en $\sigma = 5$.

Para este mundo no se muestran ejemplos de reconstrucciones ya que, salvo por la forma del mundo, serían similares a las mostradas en las figuras 5.6 y 5.7, debido a que la resolución angular es la misma, ver características de M360 en subsección 5.4.1.



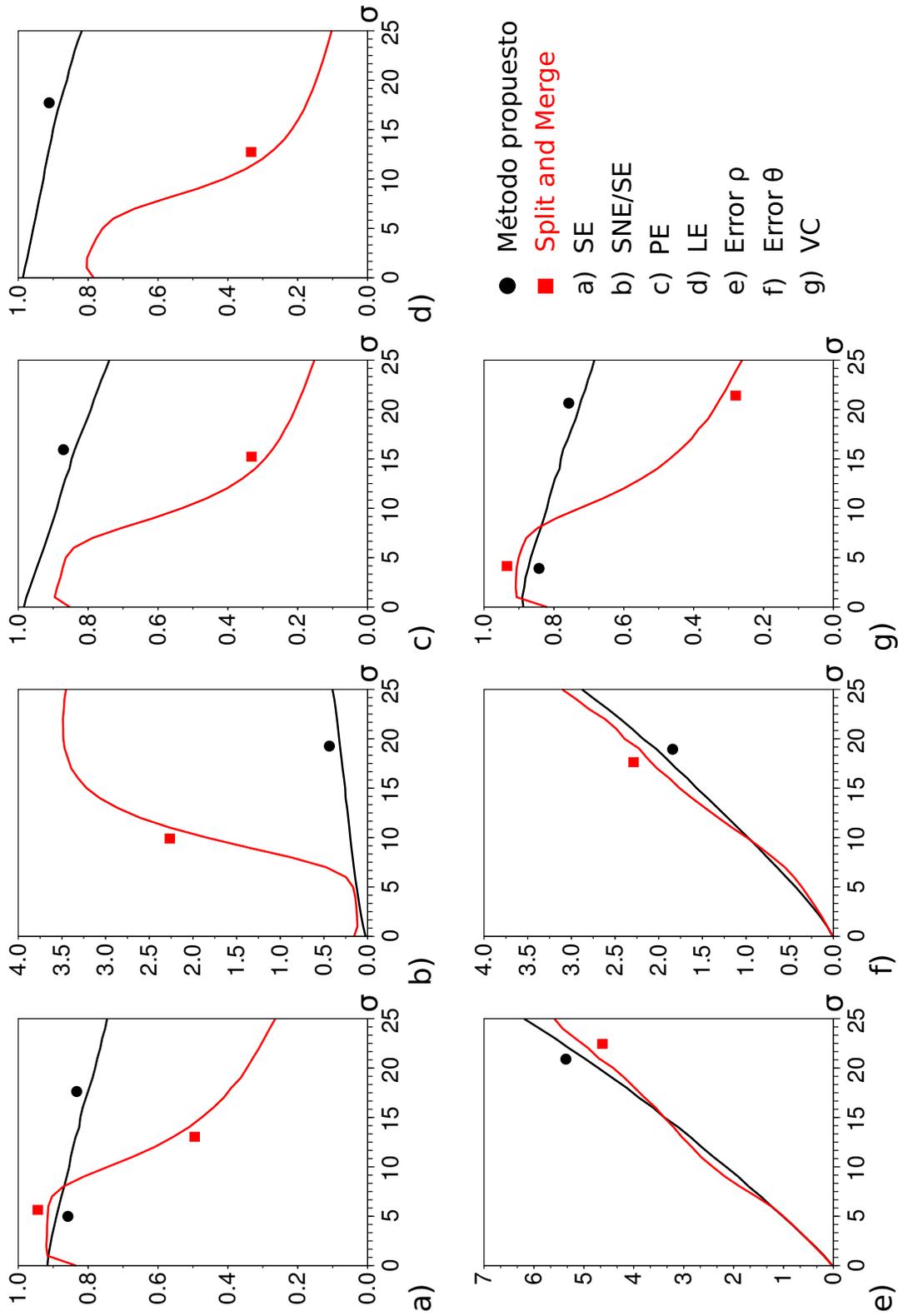
SE (prop. segs. emparejados) SNE (prop. segs. no emparejados) PE (prop. puntos emparejados)
 LE (prop. longitud total emparejada) Error ρ (desv. típica del error en ρ , cm.)
 Error θ (desv. típica del error en θ , gr.) VC (prop. estimaciones correctas de la varianza del ruido)

Figura 5.8: Resultados según el criterio χ^2 de comparación de parámetros, ec. 5.2, para M360. En el eje de abcisas aparece σ en unidades de cm. a 10 m.



SE (prop. segs. emparejados) SNE (prop. segs. no emparejados) PE (prop. puntos emparejados)
 LE (prop. longitud total emparejada) Error ρ (desv. típica del error en ρ , cm.)
 Error θ (desv. típica del error en θ , gr.) VC (prop. estimaciones correctas de la varianza del ruido)

Figura 5.9: Resultados según el criterio F de comparación de parámetros, ec. 5.3, para M360. En el eje de abcisas aparece σ en unidades de cm. a 10 m.



SE (prop. segs. emparejados) SNE (prop. segs. no emparejados) PE (prop. puntos emparejados)
 LE (prop. longitud total emparejada) Error ρ (desv. típica del error en ρ , cm.)
 Error θ (desv. típica del error en θ , gr.) VC (prop. estimaciones correctas de la varianza del ruido)

Figura 5.10: Resultados según el criterio χ^2 de comparación de parámetros, ec. 5.2, para M720. En el eje de abcisas aparece σ en unidades de cm. a 10 m.

5.4.4.3. M720

En M720 se explora la cuestión de una resolución angular menor, ver características de M720 en subsección 5.4.1.

Los resultados para M720 obtenidos por el método propuesto y por SM+MCT no difieren significativamente de los obtenidos para M360, comentados en la subsección 5.4.4.2. Salvo por un aumento muy significativo en el máximo de SNE , 5.10 b) y 5.11 b) para SM+MCT, señal de que SM aumenta la fragmentación al aumentar la resolución bajo ruido alto. El cambio de estado para SM+MCT ocurre antes que en M360, esto es especialmente apreciable en la figura 5.11 e) y f). El método propuesto no muestra alteraciones apreciables al cambiar de resolución.

En la figura 5.12 a) se observa una reconstrucción del mundo lograda con nivel de ruido bajo, la reconstrucción es prácticamente exacta para el método propuesto, SM deja puntos sin clasificar. En la figura 5.13 se observan reconstrucciones del mundo logradas con nivel de ruido alto. En a) y c) usando el método propuesto y en c) y d) SM+MCT, se observan situaciones semejantes a las comentadas para la figura 5.7 en la subsección 5.4.4.1, aunque la fragmentación provocada por SM puede ser mucho mayor, 5.13 d).

5.4.4.4. Resumen

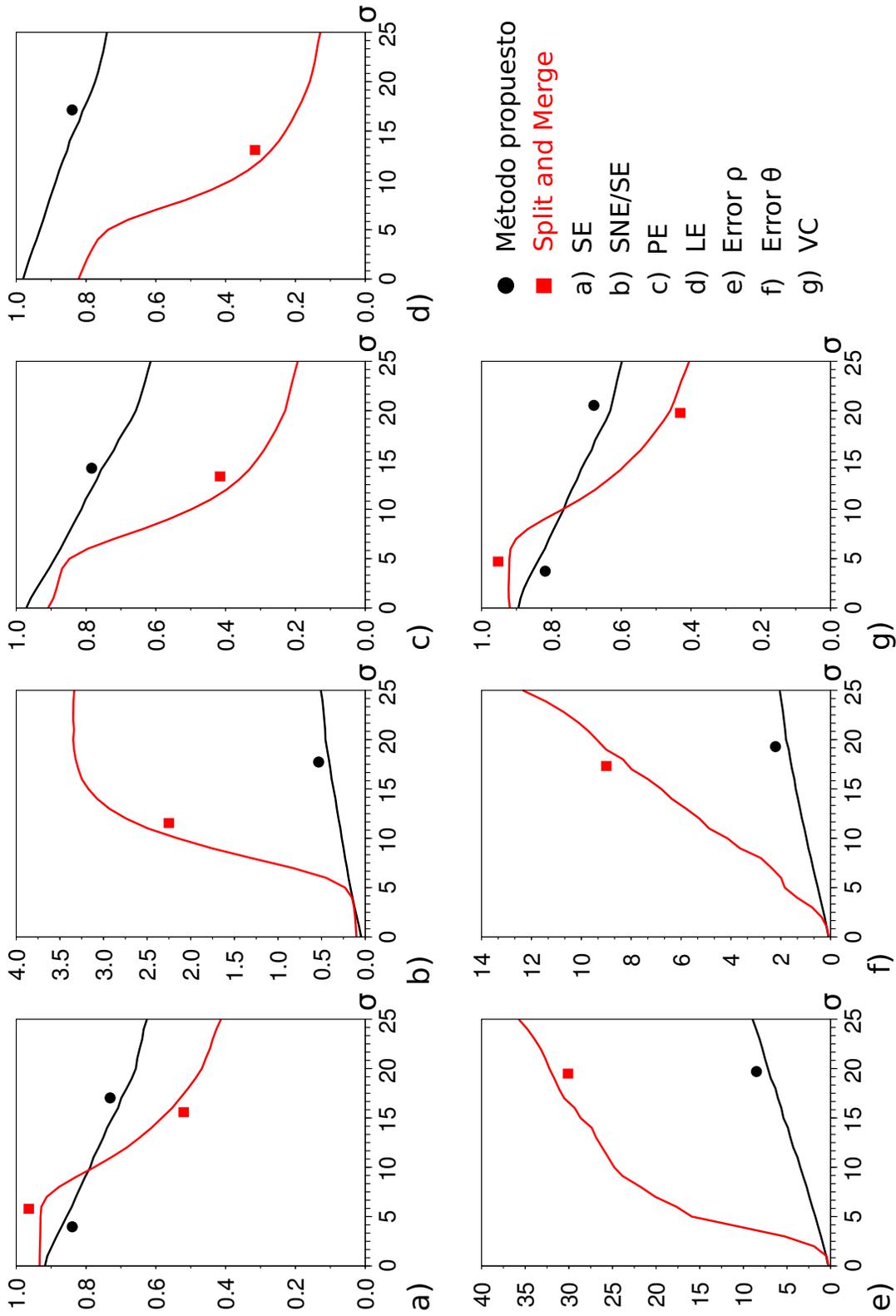
Desde en punto de vista de la especificación de límites de tolerancia, criterio χ^2 , el método propuesto tiene un rendimiento similar a SM+MCT en cuanto a proporción de segmentos emparejados, cuando el ruido es bajo, aunque logra clasificar una zona del mundo apreciablemente mayor y una mayor cantidad de puntos. En cuanto a los errores en los parámetros se tiene un comportamiento muy similar entre ambos métodos. El método propuesto experimenta una degradación casi lineal con pendiente poco acusada frente al aumento del nivel de ruido. SM+MCT se degrada notablemente ante el aumento del ruido produciéndose una elevada fragmentación. Esta fragmentación se ve incrementada al aumentar la resolución angular del haz, el método propuesto no experimenta variaciones apreciables ante el cambio de resolución realizado.

Desde en punto de vista de lo que el método “opina” sobre la especificación, criterio F , el método propuesto tiene una opinión bastante similar al del punto de vista de la especificación. SM+MCT parece que aumenta el rendimiento en emparejamientos a niveles de ruido bajo, pero lo hace a costa de un incremento muy considerable en la fragmentación y en los errores de los estimadores a poco que el nivel de ruido comienza a aumentar.

5.5. Experimentos sobre parámetros de scale-space y eficiencia

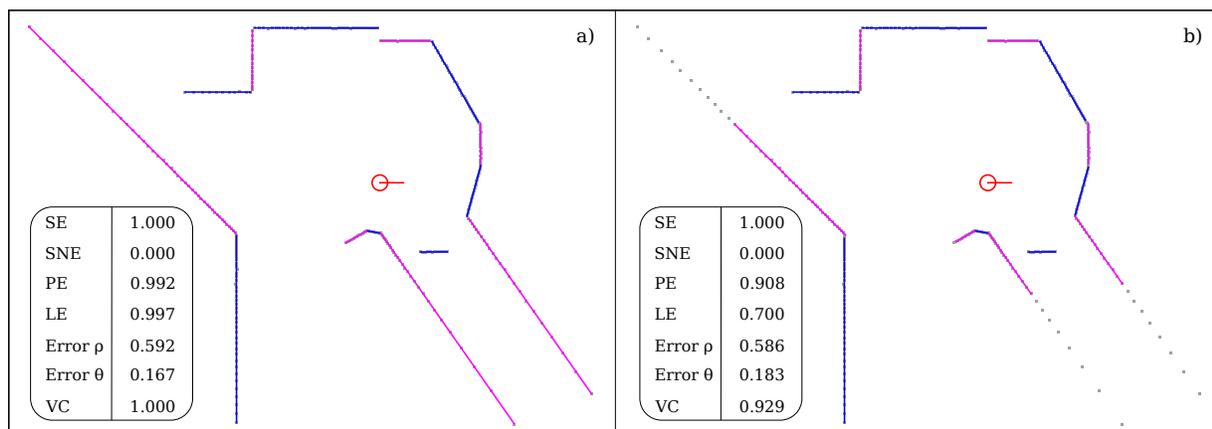
5.5.1. Experimentos sobre parámetros de scale-space

En la sección 4.3 se realizaron unos experimentos no exhaustivos sobre la variación de parámetros en el procedimiento de scale-space. En esta sección se describirá un experimento



SE (prop. segs. emparejados) SNE (prop. segs. no emparejados) PE (prop. puntos emparejados)
 LE (prop. longitud total emparejada) Error ρ (desv. típica del error en ρ , cm.)
 Error θ (desv. típica del error en θ , gr.) VC (prop. estimaciones correctas de la varianza del ruido)

Figura 5.11: Resultados según el criterio F de comparación de parámetros, ec. 5.3, para M720. En el eje de abcisas aparece σ en unidades de cm. a 10 m.



SE (prop. segmentos emparejados) SNE (prop. segmentos no emparejados) PE (prop. puntos emparejados) LE (prop. longitud total emparejada)
 Error ρ (desviación típica del error en ρ , cm.) Error θ (desviación típica del error en θ , gr.) VC (prop. estimaciones correctas de la varianza del ruido)

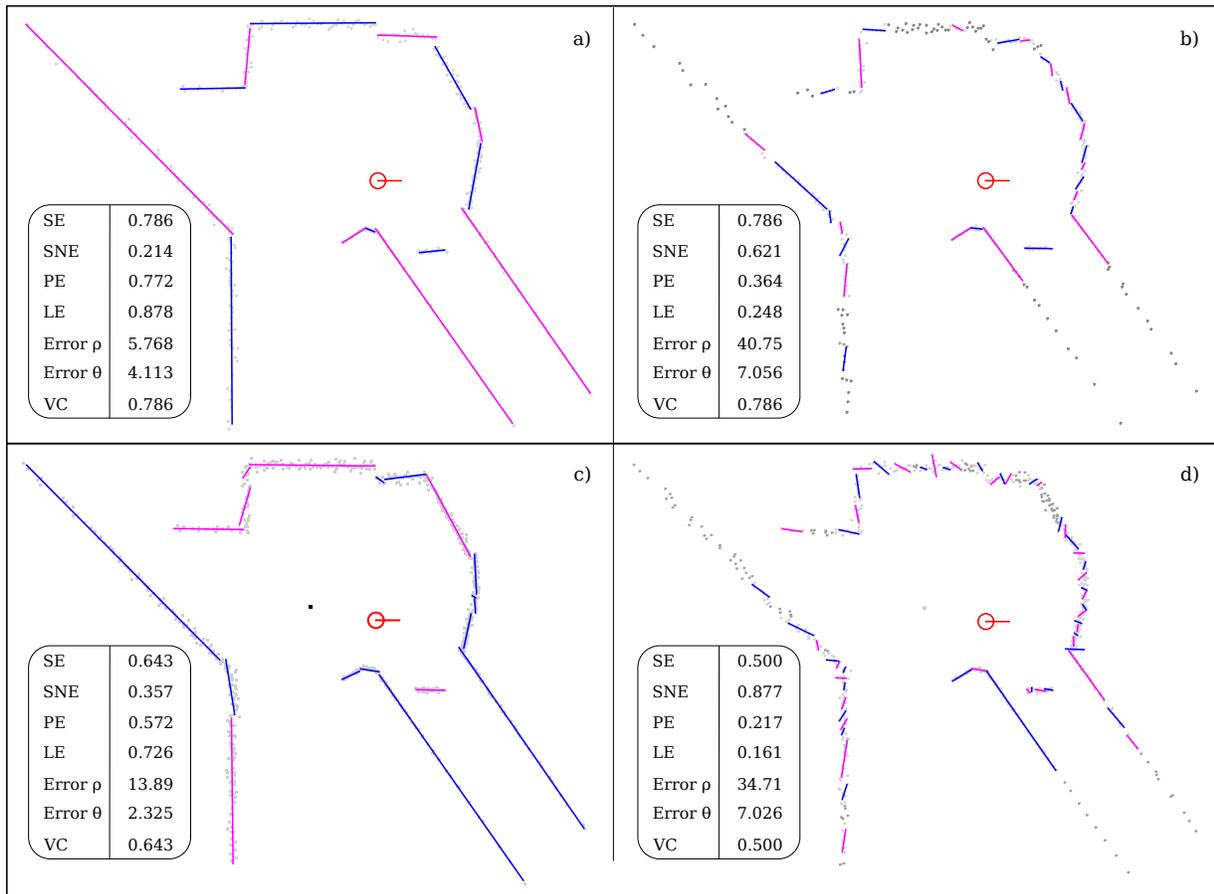
Figura 5.12: Reconstrucción de M720 cuando el nivel de ruido es pequeño, $\sigma = 0,002$. En a) el método propuesto y en b) SM+MTC. Vuelve a repetirse la situación comentada en el pie de la figura 5.6. Aunque ahora SM+MTC empareja ligeramente más puntos y más longitud debido a la menor resolución angular, $0,5^\circ$. La precisión de las estimaciones es ahora similar para ambos métodos.

sobre la variación del número de niveles del parámetros de escala, L , y de la constante k usada en el filtro de las curvas de nivel. Para ello se compararon los resultados de la segmentación usando $L = 24$ y $k = 2,5$, $L = 15$ y $k = 2,5$ y $L = 15$ y $k = 3$. Siguiendo el mismo esquema de generación de muestras, se obtuvieron las medidas de corrección y precisión usadas en los experimentos de la sección 5.4 para el mismo rango de valores de la desviación típica del ruido sin encontrar diferencias apreciables entre los resultados empleando los tres pares de parámetros. Por lo que se aporta una nueva prueba apuntando a que el efecto de la elección de los valores de los parámetros del procedimiento de scale.space no es crítico, dentro del rango de valores adecuados propuesto en el capítulo 4.

5.5.2. Experimentos sobre eficiencia

En las pruebas de eficiencia se testeará la velocidad de ejecución de la implementación en C++ desarrollada CuadraTroncoso (2011), la principal intención de estas pruebas de velocidad es probar que el tiempo de ejecución es pequeño comparado con el tiempo de muestreo del sensor o el del ciclo del robot, más que compararlo con otros métodos. Las pruebas se han realizado en un ordenador portátil con un procesador Core Duo 2 a 2.24 GHz, los robots de nueva generación están siendo equipados con este tipo de procesadores y varios conocidos modelos antiguos pueden ser actualizados.

Se van a realizar 1000 ejecuciones de cada método para cada uno de los valores de la desviación típica del ruido usados en los experimentos de la sección 5.4, en cada uno de los mundos M180, M360 y M720, cada uno tiene el doble de mediciones que el anterior. En el experimento con M180 se incluirá RANSAC, pero no en los experimentos en M360 y M720.



SE (prop. segmentos emparejados) SNE (prop. segmentos no emparejados) PE (prop. puntos emparejados) LE (prop. longitud total emparejada)
 Error ρ (desviación típica del error en ρ , cm.) Error θ (desviación típica del error en θ , gr.) VC (prop. estimaciones correctas de la varianza del ruido)

Figura 5.13: Reconstrucción de M720 cuando el nivel de ruido es grande, $\sigma = 0,02$. En a) y c) el método propuesto y en b) y d) SM+MTC. Los resultados de las magnitudes de corrección y exactitud aparecen en tablas, se muestran ejemplos de estimaciones con SE superior a la media a) y b) e inferior a la media c) y d). Vuelve a repetirse la situación comentada en el pie de la figura 5.7. Aunque ahora el nivel de fragmentación que puede producir SM es realmente excesivo y el error en las estimaciones aumenta especialmente para SM.

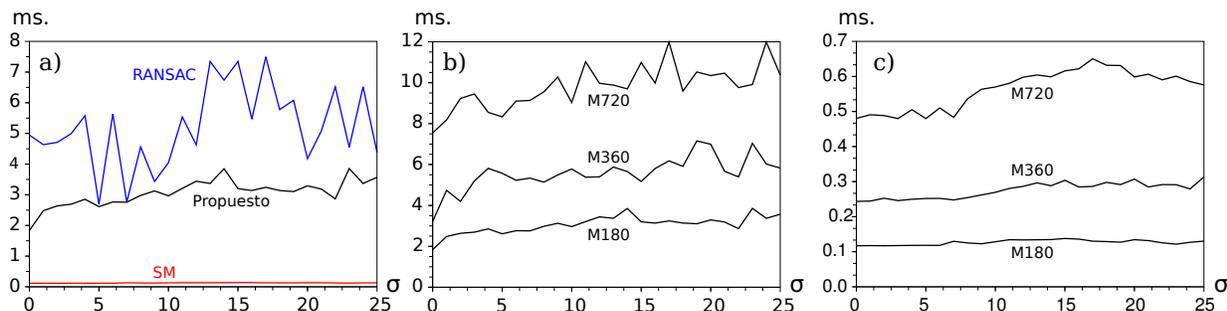


Figura 5.14: Resultados de las pruebas de velocidad para distintos niveles de ruido. En a) resultados en M181 para los tres métodos. En b) resultados del procedimiento propuesto en los tres mundos y en c) los de SM. En SM y RANSAC no se ha incluido el tiempo de estimación por MCT.

Para SM y RANSAC no se ha incluido los tiempos de las estimaciones por MCT añadidas. En el método propuesto no es posible separar el proceso de clustering del de estimación, ya que en este se refina el clustering, por lo que se dan los tiempos de ambos procesos. El procedimiento de scale-space se lleva entre un 55 % y un 65 % del tiempo total de proceso.

Los resultados para M180 se muestran en la figura 5.14 a), no hay duda de que SM es muy rápido, entre 15 y 20 veces más rápido que el procedimiento propuesto. RANSAC sin embargo aparece como el doble de lento aproximadamente que el procedimiento propuesto. En la figura 5.14 b) aparecen los resultados del método propuesto en los distintos mundos y en c) los resultados para SM. El nivel de ruido tiene una ligera influencia en los tiempos de ejecución en ambos métodos, parece que esta influencia se acentúa con el aumento del número de mediciones. El método propuesto muestra unos tiempos medios de ejecución del orden de 2,5 ms. a 3 ms. en M180, de 5 ms. a 6 ms. en M360 y de 9 ms. a 12 ms. en M720, estos tiempos parecen sugerir una complejidad temporal próxima a la lineal o como mucho a la de las transformadas de Fourier $O(\log n)$ del proceso de scale-space.

La reducción en la velocidad, con respecto a SM, es el precio a pagar por la corrección, precisión, robustez y flexibilidad del método propuesto, aunque estos tiempos son pequeños comparados con los tiempos de muestreo de los dispositivos láser habituales. La implementación desarrollada en C++ está todavía en desarrollo y se podrían conseguir tiempos de ejecución menores. Se podrían usar las GPUs de las tarjetas gráficas, con las que también vienen equipadas los nuevos robots, para realizar las transformadas de Fourier en un tiempo mucho menor y mejorar sustancialmente la velocidad del proceso de scale-space. El filtro de Kalman es muy sencillo y fácilmente realizable como dispositivo hardware, lo que redundaría en la velocidad del procedimiento.

Parte II

El método del centro de áreas

Capítulo 6

Navegación reactiva

6.1. Introducción

Una control reactivo se basa, según de la Paz López (2003), “en una colección de pares condición-acción, con un conjunto mínimo de estados internos, ejecutando simplemente la acción correspondiente para cada conjunto de lectura sensorial”. Se establece pues una correspondencia directa entre los sensores y los motores, según se siente así se actúa, no hay planificación y además solo se tiene una información parcial del mundo que rodea al robot.

La colección de pares, que Arkin (1998a) denomina la codificación de la conducta, puede ser discreta o continua. En el caso discreto se pueden usar dos estrategias para establecer la correspondencia condición-acción: la primera es que la información sensorial proporcione la manera de encontrar la acción apropiada, similar a como actúan los patrones de acción fija en animales, la segunda se basa en sistemas de reglas. Estas reglas son del tipo if-then, puede darse el caso de que las reglas entren en conflicto, por lo cual es necesario arbitrar un mecanismo de resolución de conflictos. En el caso continuo se tiene una función matemática que relaciona cada lectura sensorial con una reacción. La correspondencia y reglas de resolución de conflictos puede estar definida a priori, cableada, como en Walter (1950); Braitenberg (1986); N. Franceschini (1992) o puede aprenderse, como veremos más adelante.

La navegación reactiva experimentó un gran empuje a raíz del trabajo de Brooks (1986), en él se propone una arquitectura de subsunción en la que se describen varios niveles de abstracción creciente, que resuelven las subtarear en las que se descompone la navegación, p. ej.: detección y evitación de obstáculos, deambular (wandering), explorar, etc, esta idea está muy relacionada con las arquitecturas basadas en conductas de Arkin, más tarde Mataric (1997) propuso una modificación a esta arquitectura. Un control básico de navegación debe, por lo tanto, mover al robot detectando y evitando obstáculos, esta es la primera cuestión a resolver. La segunda es dirigir al robot a un objetivo determinado. Si bien los controles reactivos están bien preparados para resolver la primera cuestión, incluso en entornos cambiantes Singhal (1997), debido a su velocidad de respuesta, no lo están para abordar la

segunda, debido a su conocimiento eminentemente local del entorno, y en estos casos puede hacer falta una planificación para llevar a cabo la tarea Arkin (1998b), sistemas híbridos.

El método de navegación que se propone en esta tesis se enmarca dentro de los que usan codificación continua, por eso se expondrán primero este tipo de métodos comentando sus virtudes y defectos. Para completar esta visión de la robótica reactiva se comentará la cuestión del aprendizaje.

En trabajos anteriores del equipo de investigación donde se desarrolló esta tesis, se comenzó el estudio del uso del centro de áreas del área libre percibida por el robot como método de navegación, ya que esta tesis continúa esa línea de investigación se expondrán las bases de dicho método en las secciones finales del capítulo. Para no repetir figuras remitiremos al lector a las de otros capítulos posteriores.

6.2. Controles continuos

Estos controles también son conocidos como planeadores de rutas locales, para distinguirlos de los planeadores de rutas globales Arkin (1998a) que no entran dentro del espectro de la navegación reactiva y por lo tanto del ámbito de este estudio.

6.2.1. Métodos de potencial

En los métodos de potencial de campo (Potential Field o Virtual Forces Field VFF) métodos de potencial la dirección del movimiento viene determinada por un campo de fuerzas, los obstáculos generan fuerzas repulsivas y los objetivos atractivas, por lo que sus posiciones deben ser conocidas. Los potenciales se calculan usando una malla rectangular que cubre el área percibida. Fueron introducidos por Krogh (1984); Khatib (1986), más tarde Koren and Borenstein (1991) pusieron de manifiesto varios problemas del método: como la presencia de mínimos locales en el campo de fuerzas, donde el robot queda atrapado, y conductas cíclicas oscilatorias. Estas situaciones se producen, por ejemplo, cuando el robot intenta pasar por un pasillo estrecho, cuando un obstáculo, especialmente cóncavo, se interpone en su camino hacia el objetivo, pueden verse imágenes relacionadas con estos problemas en las figuras 7.3, 7.8 y 7.9. Debido a estas limitaciones se han propuesto muchas modificaciones al método. Los campos de potencial armónico, donde el uso de funciones armónicas evita los mínimos locales, se han usado por ejemplo en Kim and Khosla (1992); Jacob et al. (1997); Masoud and Masoud (2000); Fahimi et al. (2009). Campos de potencial variantes en el tiempo Tianmiao and Bo (1992); Ferrara and Rubagotti (2007); Gonçalves et al. (2010), en los que como su nombre indica el campo además de ser función de las coordenadas espacial lo es también del tiempo. Fuerzas virtuales ficticias Serradilla and Maravall (1996); Maravall et al. (2000) que hacen que los mínimos locales repelan al robot. Métodos adaptativos Clark et al. (1992); Ram et al. (1997).

6.2.2. Histograma de campo vectorial

El método del histograma de campo vectorial (Vectorial Field Histogram, VFH) fue introducido en Borenstein and Koren (1991), usa un histograma sobre una malla 2D en coordenadas cartesianas, que es reducido a un histograma 1D en coordenadas polares, en el que se estudian sectores con baja densidad de obstáculos, los sectores se seleccionan en función de la dirección a objetivo. Suele estar aquejado de los mismo problemas que los métodos de potencial. El método fue mejorado con dos variantes: VHF+ Ulrich and Borenstein (1998), usa un umbral de histéresis para suavizar la trayectoria, una función de costo para analizar el rendimiento y tiene en cuenta en tamaño del robot, entre otras. Más adelante se presentó otra versión mejorada del método VHF* Ulrich and Borenstein (2000) donde se busca un comando de movimiento que minimize una función de costo usando el algoritmo A*. Otra variante es el VHF Doble Yang et al. (2000).

6.2.3. Ventana dinámica

El enfoque de las ventanas dinámicas (Dynamic Window Approach DWA) fue introducido en Fox et al. (1997) tiene en cuenta las restricciones cinemáticas y dinámicas del robot, la ventana contiene los valores de las velocidades lineales y radiales admisibles en cada momento. Aplicando una función objetivo que depende de la velocidad del robot, la distancia al obstáculo más próximo y el progreso hacia la posición objetivo, se elige el comando de velocidad.

En Khatib et al. (2002) se desarrolla la idea de espacio ego-dinámico en el que las restricciones dinámicas están implícitas en la estructura de dicho espacio, en Ogren and Leonard (2005) es visto en el contexto de de los modelos de control predictivo. Las características de este método le hacen apropiado para robots que se desplazan a velocidades elevadas Smolic-Rocak et al. (2010).

6.3. Aprendizaje

El aprendizaje o adaptación puede mejorar el rendimiento de un control de navegación Arkin (1998a). El aprendizaje puede producir habituación, disminución de la intensidad de la respuesta conforme se desarrolla el aprendizaje, lo que puede verse como la inhibición de una conducta. También puede producir sensibilización por el contrario implica el aumento de la respuesta, la conducta es facilitada. El aprendizaje intenta generalmente ajustar los valores de ciertos parámetros de los métodos. Los métodos de aprendizaje más comúnmente usado en navegación son los siguientes: aprendizaje por refuerzo, redes neuronales, lógica borrosa y algoritmos genéticos.

6.3.1. Aprendizaje por refuerzo

En el aprendizaje por refuerzo (Reinforcement Learning Sutton and Barto (1981); Sutton (1988)) las conductas son recompensadas o castigadas según una función de costo que

el algoritmo trata de maximizar. Es quizás el modo más empleado en navegación, primeros ejemplos de su uso se pueden encontrar en Maes and Brooks (1990) para controlar el movimiento de las patas de un robot, en Kröse and van Dam (1992) para la evitación de obstáculos y en Gat (1991) en un comportamiento dirigido a objetivos. Más adelante sido usado en tareas reactivas como las siguientes. “Búsqueda de comida” (foraging) Mataric (1994). Evitación de obstáculos y alcance de objetivos en distintos entornos más o menos dinámicos del R. Millán and Torras (1995); Kirchner (1998); Davesne and Barret (1999); Maravall and de Lope (2003); Yen and Hickey (2004). Movimiento de robots con patas articuladas Benbrahim and Franklin (1997); Barfoot et al. (2006); Anshar and Williams (2007); Erden and Leblebicioglu (2008).

6.3.2. Redes neuronales

Mucho de los métodos del paradigma de las redes neuronales han sido empleados en la navegación reactiva en robótica, estas redes pueden estar inspiradas más o menos en la biología. Comentaremos algunos de sus usos en el tema que nos ocupa. El aprendizaje por condicionamiento clásico o pavloviano está fuertemente bioinspirado, ha sido usado en Verschure et al. (1995); Porr et al. (2003); Porr and Wörgötter (2006); Cuadra Troncoso et al. (2007, 2009). Para la navegación reactiva también se han usado: perceptrones en Ortega and Camacho (1996), funciones de base radial en Larbi et al. (2005); Rossomando et al. (2011), clasificadores auto-organizados en Dubrawski and Crowley (1994) y memorias asociativas en Nehmzow and McGonigle (1994), entre otros. Incluso se han realizado experimentos preliminares para el estudio de controles de navegación reactivos usando cultivos de neuronas reales de Santos et al. (2010); Ferrández et al. (2010).

6.3.3. Lógica borrosa

La lógica borrosa es muy útil en los sistemas basados en reglas para resolver conflictos entre conductas. Los controles de navegación basados en lógica borrosa fueron usados inicialmente en navegación reactiva por Reignier (1994), por Saffiotti et al. (1993) en el robot Flakey y por Goodridge and Luo (1994) en el robot MARGE. Desde entonces han sido estos controles han sido muy utilizados en el tema tratado, por ejemplo, Castellano et al. (1997); Xu et al. (1998); Maaref and Barret (2000); Selekwia et al. (2008); Huq et al. (2008). Redes neuronales y lógica borrosa se ha unido en Song and Sheen (2000); Hui et al. (2006); Demirli and Khoshnejad (2009).

6.3.4. Algoritmos genéticos

Los algoritmos genéticos han sido usados también en la navegación reactiva dentro de la denominada Robótica Evolutiva Nolfi and Floreano (2004), fueron introducidos en Pearce et al. (1992), aunque pueden tener el problema de la velocidad de ejecución, que puede no hacerlos muy adecuados para la operación en tiempo real. Uno de los primeros casos de utilización en se puede encontrar en Ram et al. (1994). Normalmente son usados en

conjunción con un método basado en redes neuronales Floreano and Mondada (1998); Han and Oh (2008); Fernandez-Leon et al. (2009); Ruini and Cangelosi (2009) o lógica borrosa Braunstingl et al. (1995); Matellán et al. (1998); Pratihari et al. (1999); Abdessemed et al. (2004).

6.4. El centro de áreas total

En esta sección consideramos que el robot está equipado con sensores de rango que cubren un sector de 360 grados alrededor de él, de ahí el calificativo de total que se aplica al centro de áreas en esta sección. El centro de áreas es el equivalente en dos dimensiones del centro de masas o de gravedad. La fórmula en coordenadas cartesianas para el caso de un polígono de n vértices $\{(x_j, y_j)\}$ es:

$$x_{CA} = \frac{1}{6A} \sum_{j=1}^{n-1} (x_j y_{j+1} - x_{j+1} y_j) (x_j + x_{j+1}) f(x_j, y_j, x_{j+1}, y_{j+1})$$

$$y_{CA} = \frac{1}{6A} \sum_{j=1}^{n-1} (x_j y_{j+1} - x_{j+1} y_j) (y_j + y_{j+1}) f(x_j, y_j, x_{j+1}, y_{j+1})$$

donde $A = \frac{1}{2} \sum_{j=1}^{n-1} (x_j y_{j+1} - x_{j+1} y_j) f(x_j, y_j, x_{j+1}, y_{j+1})$ es el área y f es la función de densidad de área. La densidad de área puede ser útil para modelar la importancia relativa de las zonas de espacio libre cercanas al robot con respecto a las lejanas, aunque suele usarse $f \equiv 1$.

El trabajo seminal en el campo lo constituye Álvarez et al. (1999) donde se usa el centro de áreas para conseguir un sistema de referencia adecuado para la integración de datos sensoriales. Durante el desarrollo de la tesis doctoral de la Paz et al. (2003) se comenzó a utilizar en centro de áreas para definir el movimiento del robot. Primeramente el centro de áreas total calculado ejercía una fuerza de atracción sobre el robot y cuando el robot llegaba a las inmediaciones del centro de áreas, la fuerza se tornaba repulsiva haciendo que el robot buscara otro centro de áreas al que dirigirse. De esta manera la trayectoria del robot iba pasando por sucesivos centros de áreas lo que, cuando las áreas son aproximadamente convexas, implica una trayectoria segura, ver la subsección 7.2.2 para una discusión más detallada sobre la convexidad y la seguridad. Las propiedades de invariancia de las centro de áreas total lo hacen muy interesante para la navegación en mapas, pero requieren el uso control de alto nivel que impliquen la construcción de mapas topológicos durante el movimiento del robot de la Paz López and Álvarez Sánchez (2001), de manera que es difícil usarlo como la única guía para el movimiento en entornos complejos.

Al considerar que el centro de áreas genera campos de fuerzas se podría pensar que este método no es sino una variante de los métodos de potencial, pero en de la Paz López et al. (2009) se aporta la prueba matemática de que los campos de fuerzas generados por el centro de áreas pueden no ser representables mediante un potencial, por lo que constituye una

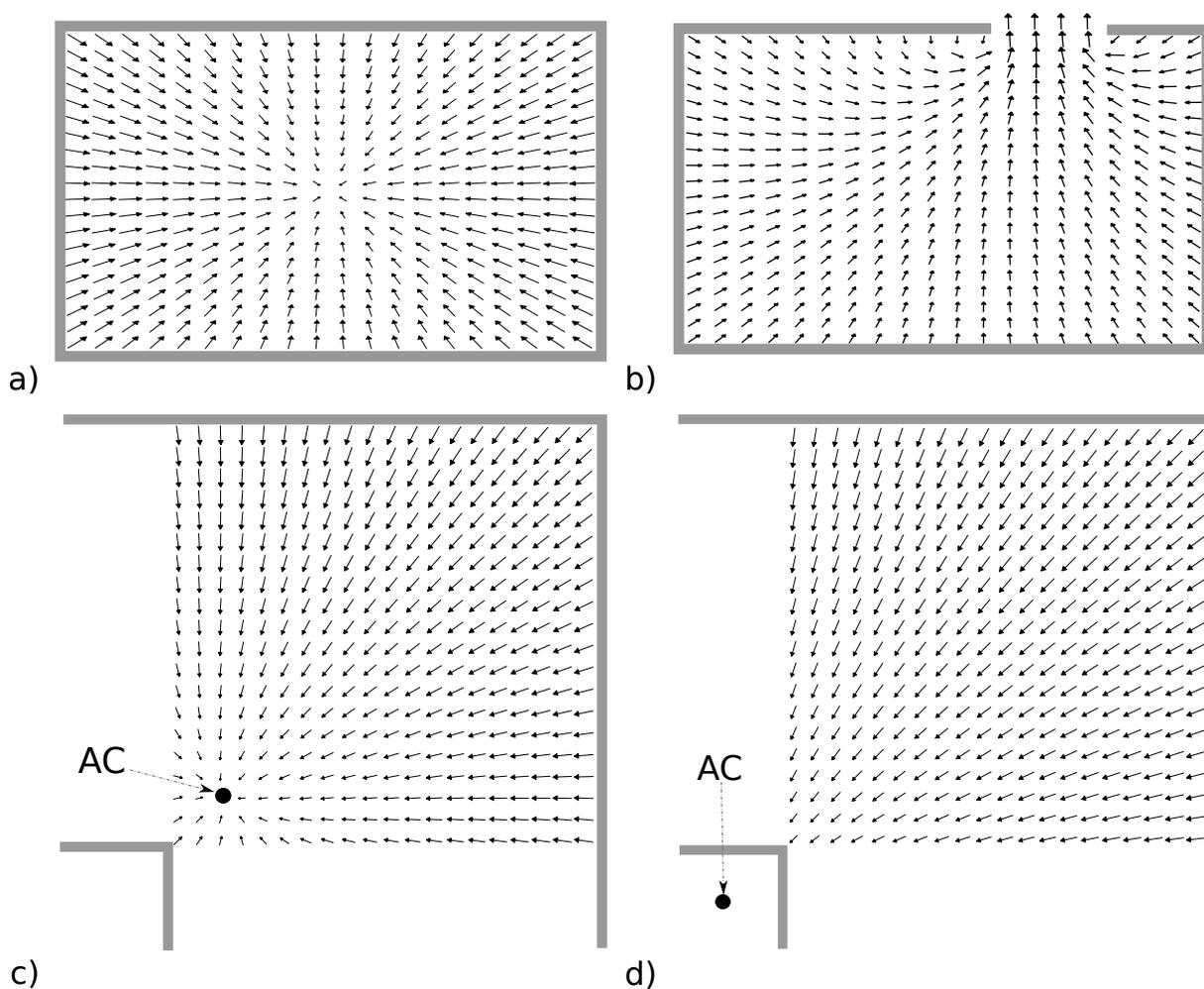


Figura 6.1: Campo de fuerzas generados por distintas configuraciones del área libre. En a) una habitación cerrada, en b) la misma habitación con un puerta abierta. En c) y d) se muestran codos en corredores, en c) el centro de áreas es accesible pero no en d).

extensión de dichos métodos. En la figura 6.1 se pueden ver campos de fuerzas generados por el centro de área en diversas situaciones.

Sobre la complejidad temporal del procedimiento es claro que es la misma que la de calcular el centro de áreas, o sea, $O(n)$ siendo n el número de mediciones. Los procedimientos anteriormente comentados en la sección 6.2 trabajan con una mallas bidimensionales, por lo que el uso del método del centro de áreas, que ha probado trabajar bien incluso con 16 mediciones, supone una muy importante reducción en los tiempos de proceso.

Capítulo 7

Centro de áreas básico

7.1. Introducción

Las ideas desarrolladas en esta parte de la tesis tienen como fuente de inspiración la siguiente hipótesis de trabajo: los humanos modelamos el área libre que nos rodea, no el área ocupada (obstáculos), cuando navegamos reactivamente, y usamos el centro de áreas de dicha área libre como una referencia para el movimiento. Esta hipótesis y su relación con la implementación del método de navegación serán desarrolladas en la sección 7.2, en ella se describirá la versión más básica del método. En la sección 7.3 se describe la modificación del método para hacerlo capaz de alcanzar objetivos (goal reaching). La descripción detallada del control de movimientos se expone en la sección 7.4. Para terminar del sin número de pruebas y experimentos realizados se comentarán los que se describen en Álvarez Sánchez et al. (2009), realizados en un simulador, sección 7.5, y con robots reales Álvarez Sánchez et al. (2010), sección 7.6.

7.2. Descripción del método

7.2.1. La hipótesis de trabajo

En la búsqueda de fuentes de inspiración en la forma que tenemos los humanos de resolver el problema de encontrar la forma correcta de evitar obstáculos, podemos primeramente imaginar la siguiente situación. Supongamos que conducimos un coche y nos encontramos con un camión grande que viene en dirección opuesta, ver figura 7.1 a). Un buen consejo para un conductor novato en esa situación es: ¡No mires al camión, mira al hueco! También se puede recordar como procedemos cuando caminamos con prisa por una calle concurrida, pero no atestada, ver figura 7.1 b). Para cada sección de nuestro camino, buscamos y seguimos reactivamente el centro de áreas del espacio libre circundante, modelando este área libre como una función de la distribución del entorno. Áreas más concurridas nos harán bajar la mirada y, por consiguiente, nuestra velocidad se reducirá. Áreas más despejadas nos permitirán alzar la mirada e incrementar nuestra velocidad.

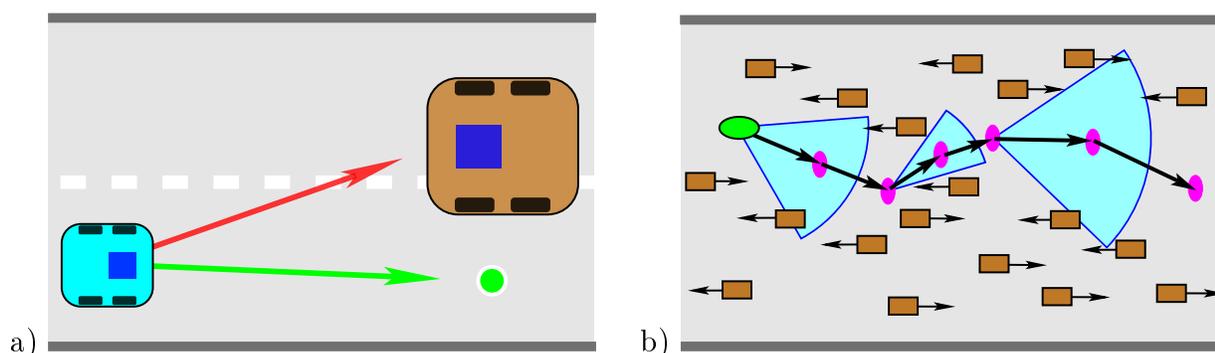


Figura 7.1: Algunas fuentes de inspiración para la hipótesis de trabajo. a) Si en la carretera nos encontramos un camión grande que viene en dirección opuesta, es mejor mirar al hueco. b) Navegación reactiva humana cuando andamos con prisa por una calle concurrida, pero no atestada. Los rectángulos representan a los otros transeúntes. En cada sección de nuestro camino tendemos a buscar el centro del espacio libre cercano a nuestro camino, representado por las elipses pequeñas.

La hipótesis de trabajo, tal como fue planteada en 7.1, es: los humanos modelamos el área libre que nos rodea, no el área ocupada (obstáculos), cuando navegamos reactivamente, y usamos el centro de áreas de dicha área libre como una referencia para el movimiento. Más precisamente:

- Los humanos calculamos y seguimos, aproximadamente el centro de áreas de un sector de la parte frontal del área libre circundante percibida, que denominaremos *sector de avance*.
- El sector de avance es aproximadamente un sector circular con radio y ángulos de comienzo y fin variables. El radio y los ángulos depende de la configuración espacial del sector de avance. Cuanto más concurrida sea el entorno más pequeños serán el radio y la apertura del sector.
- Tras seleccionar un sector, nos orientamos de manera que dicho sector ocupe una posición central de nuestro movimiento. Entonces seguimos su centro de áreas con una velocidad relacionada con su distancia, por ejemplo, centros de áreas cercanos limitan la velocidad.

En las siguientes secciones usaremos estas ideas como puntos de partida para desarrollar un procedimiento y una implementación para una navegación reactiva segura.

En la sección 6.4 se expusieron los trabajos anteriores de los directores de esta tesis sobre el centro de áreas de toda el área libre circundante al robot, centro de áreas total. Dado que el centro de áreas que se utiliza en este trabajo solo usa un área frontal, se denomina método del centro de áreas parcial o frontal, omitiremos en los que sigue el adjetivo por brevedad.

7.2.2. El proceso de división para centros de área inaccesibles

La navegación a través de áreas convexas es segura, la línea que une cualquier par de puntos de dicha área está contenida en ella, por esto las áreas convexas no contienen obstáculos dentro, aunque estarán limitadas por obstáculos. Por consiguiente volverse hacia la posición del centro de áreas y avanzar en línea recta es la opción de navegación más segura. Este razonamiento constituye un punto de apoyo para el primer y tercer ítem de la hipótesis de trabajo, ver subsección 7.2.1. Este razonamiento funciona tanto en el caso teórico en el que el entorno es percibido completamente como en el real donde solo una pequeña porción del entorno es percibida, siempre que el área percibida cubra la dirección del movimiento.

El problema aparece cuando el área percibida no es convexa. En este caso, el centro de áreas puede estar fuera de ella y por lo tanto ser inaccesible. Básicamente, la solución propuesta consiste en dividir el área percibida en dos, con la esperanza de que alguna de las dos partes tenga un centro de áreas accesible hacia el que encaminarse. La figura 7.2 d) muestra un ejemplo en el que un área no convexa se divide en dos que sí lo son.

Esta idea viene de experiencias personales en situaciones como las representadas en la 7.1 b). En esas situaciones, también hemos observado que las personas tendemos a bajar la vista mientras cambiamos la dirección de nuestro movimiento para evitar un obstáculo. Esta idea se añadió a la de proceso de división, comprimiendo las sub-sectores resultantes antes de calcular sus centros de áreas, mediante una limitación del rango máximo considerado para los sensores. Además el proceso de compresión puede producir áreas con concavidades menos perceptibles y que por lo tanto tiene una mayor probabilidad de tener un centro de áreas accesible. Por el momento, y por sencillez de la exposición, el robot es considerado como un punto matemático. En este caso punto accesible significa lo mismo que punto contenido en el interior del área libre, más adelante tendremos en cuenta las dimensiones del robot, ver subsección 7.2.3.

7.2.2.1. Descripción

Con las ideas anteriores en mente podemos pasar a describir el proceso para encontrar centros de áreas accesibles para una navegación segura ilustrado en la figura 7.2. La zona completa de área percibida frontal será denominada el sector de avance completo y el subsector seleccionado, tras una operación de división y/o compresión será denominado sector de avance restringido. Cuando se usan sensores de rango el área libre percibida puede ser vista como un polígono (estrellado). Las mediciones faltantes debidas a malos rebotes, etc. se sustituyen por el alcance máximo que se considere para el sensor u otro valor similar. Los vértices del polígono son subconjunto de mediciones y el origen del sistema de referencia local del robot, a partir de ahora para evitar repeticiones, se supondrá que dicho origen es vértice de cada polígono o sector considerado, sin mención explícita. El sistema de referencia local es el único sistema de referencia usado por el método del centro de áreas. A continuación se da la descripción del método, para realizar “wandering”, que puede seguirse en la figura 7.2.

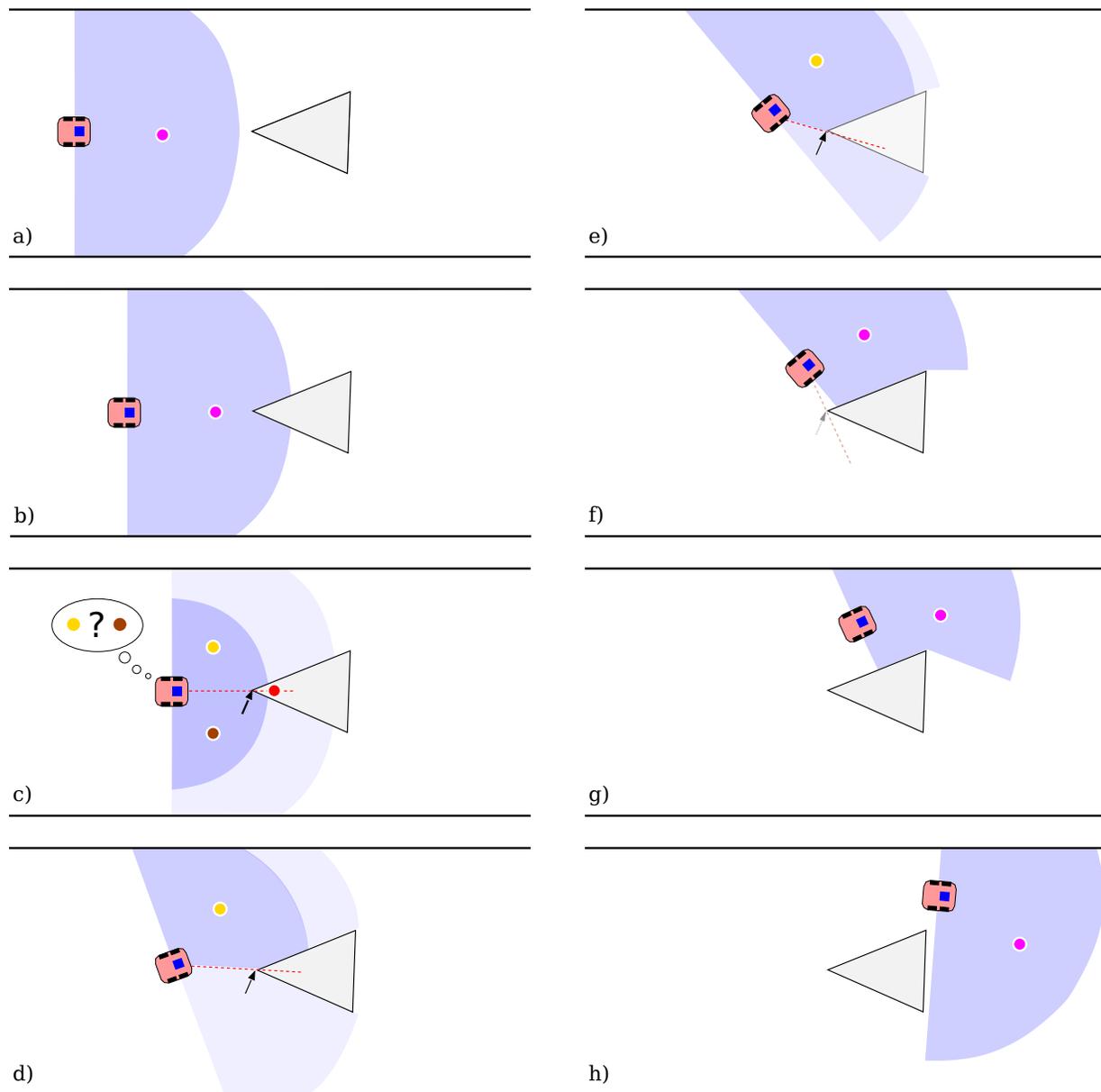


Figura 7.2: Evitación de obstáculos usando en centro de áreas. Estas imágenes forman un secuencia mostrando cómo el robot supera un obstáculo usando el método del centro de áreas. La secuencia se explica detalladamente en la subsección 7.2.2. Las áreas sombreadas, oscuras o claras, representan el sector de avance completo. El área oscura es el sector de avance actual, las claras son sectores descartados. Los círculos pequeños son los centros de áreas, el color claro (dorado) indica la selección del sub-sector izquierdo y el oscuro (marrón) la del derecho. El punto de división se indica con una flecha.

1. El robot comienza en una posición en la que el centro de áreas del sector de avance completo es accesible¹, en nuestros experimentos el sector frontal tiene una amplitud de 180°. El robot comienza a seguir el centro de áreas, ver figura 7.2 a).
2. Mientras que el centro de áreas sea accesible el robot lo sigue, ver figura 7.2 b).
3. Cuando el centro de áreas se vuelve inaccesible, se determinan dos los dos puntos del polígono que definen el sector angular que contiene a la recta que une el origen con el centro de área. De entre los dos se elige el punto más cercano al robot como punto de división y se divide el polígono en dos. Los polígonos resultantes son comprimidos, por ejemplo usando la mitad del alcance máximo considerado, se calculan sus centros de áreas y se determina su accesibilidad, ver figura 7.2 c).
 - Si los centros de los dos sectores son accesibles se elige uno de ellos, aleatoriamente o mediante una preferencia externa.
 - Si solo uno de los centros es accesible se elige su sector.
 - Si ambos centros son inaccesibles, no se puede continuar el camino, entonces se determina el centro de áreas de la parte del área libre percibida posterior y se gira hacia él, es una maniobra de escape.

Las coordenadas locales del punto de división han de ser guardadas en una memoria a corto plazo y actualizadas usando la odometría del robot. El robot casi se para y comienza a girar hacia el centro de áreas seleccionado. Mientras gira las coordenadas del punto de división se actualizan. Durante el giro, vértices del polígono descartado pasan al lado del punto de división donde se encuentra el sector de avance restringido, estos vértices se añaden a dicho sector y son eliminados del descartado, 7.2 d).

Este proceso incrementa la amplitud de sector de avance restringido por lo que se incrementa el alcance máximo considerado proporcionalmente, así cuando la amplitud del sector restringido alcance a la del completo el rango máximo considerado habrá alcanzado su valor inicial, ver figura 7.2 e) y sección 7.4.

4. Cuando el movimiento del robot hace que todos los vértices hayan pasado el punto de división el polígono descartado es eliminado. En este momento el punto de división puede ser borrado de la memoria a corto plazo y el robot está en la situación original, ver figura 7.2 f). Ahora, siguiendo el centro de áreas el robot puede sobrepasar el obstáculo, figuras 7.2 g) y h).

Cuando se usa el sector de avance restringido y si su centro de áreas se vuelve inaccesible, se puede realizar un nuevo proceso de división. Se podrá usar el nuevo sector resultante siempre que su amplitud no sea demasiado pequeña, ver figura 7.3. Si el proceso de división se realiza con éxito se obtiene un nuevo punto de división, este punto y el anterior punto

¹En esta exposición, por simplicidad, el centro de áreas es accesible al inicio, pero no es una condición obligatoria.

de división delimitan el nuevo sector de avance. Mientras el robot se mueve vértices son añadidos al sector elegido y eliminados de los descartados. En entornos complejos, especialmente cuando el robot gira, es también posible que vértices abandonen el sector elegido, en ese caso son añadidos al correspondiente sector descartado. Cuando la amplitud del sector de avance se vuelve demasiado pequeña, como resultado de sucesivos procedimientos de división el robot tiene que efectuar la maniobra de escape.

7.2.3. Teniendo en cuenta el tamaño del robot

En la discusión previa se ha usado una definición *naive* para accesibilidad, ahora se tomarán en consideración las dimensiones del robot. Primeramente, para que un punto sea accesible es necesario que el robot pueda reposar sobre él, permaneciendo en el área libre, esta condición se puede establecer como: es necesario que para que un punto pueda ser considerado accesible que un círculo, de radio suficiente para contener al robot, centrado en el punto esté totalmente contenido en el área libre percibida.

Si el robot tiene que alcanzar el centro de área entonces, un corredor recto, suficientemente ancho, del robot al centro debe estar contenido en el área libre. Durante los experimentos realizados esta condición resultó ser muy restrictiva porque no tiene en cuenta el carácter dinámico del centro de área, el centro se mueve con el robot. Por lo que la condición se relajó reduciendo la longitud del corredor y haciéndola dependiente de las velocidades lineal y radial del robot. Más o menos, el corredor solo tiene que ser lo suficientemente grande como para permitir al robot parar e iniciar la maniobra de escape sin chocar. Un punto para ser considerado accesible ha de cumplir ambas condiciones. Cuando el sector de avance tiene su máxima amplitud, o cercana, y la accesibilidad no es satisfecha por ninguno de los dos centros de área resultantes del proceso de división, se da una oportunidad antes de iniciar la maniobra de escape: el sector de avance es comprimido a la mitad y entonces se vuelve a intentar el proceso de división una sola vez más.

Al dividir el sector de avance es posible que ambos centros de áreas sean accesibles, por lo que habrá que decidirse por uno de ellos. Se puede realizar la selección de una forma totalmente aleatoria, la misma probabilidad para ambos sectores o se puede elaborar un criterio de conveniencia en el que basar la elección. Para una navegación segura este criterio podría estar basado en la amplitud del sector o su área. Se pensó que ambas magnitudes eran importantes para la seguridad y se usó su producto para construir una medida relativa de la conveniencia de un sector, los productos son usados habitualmente para expresar la interacción entre variables Cox (1984). Más precisamente, como medida de la conveniencia relativa de un sector se usará: la contribución del sector a la suma del producto mencionado para cada sector. En la implementación realizada, cuando el robot está vagabundeando sin rumbo, la elección de sectores es realizada aleatoriamente usando las medidas de conveniencia como probabilidades, se usa un umbral para no tener en cuenta sectores de poca conveniencia.

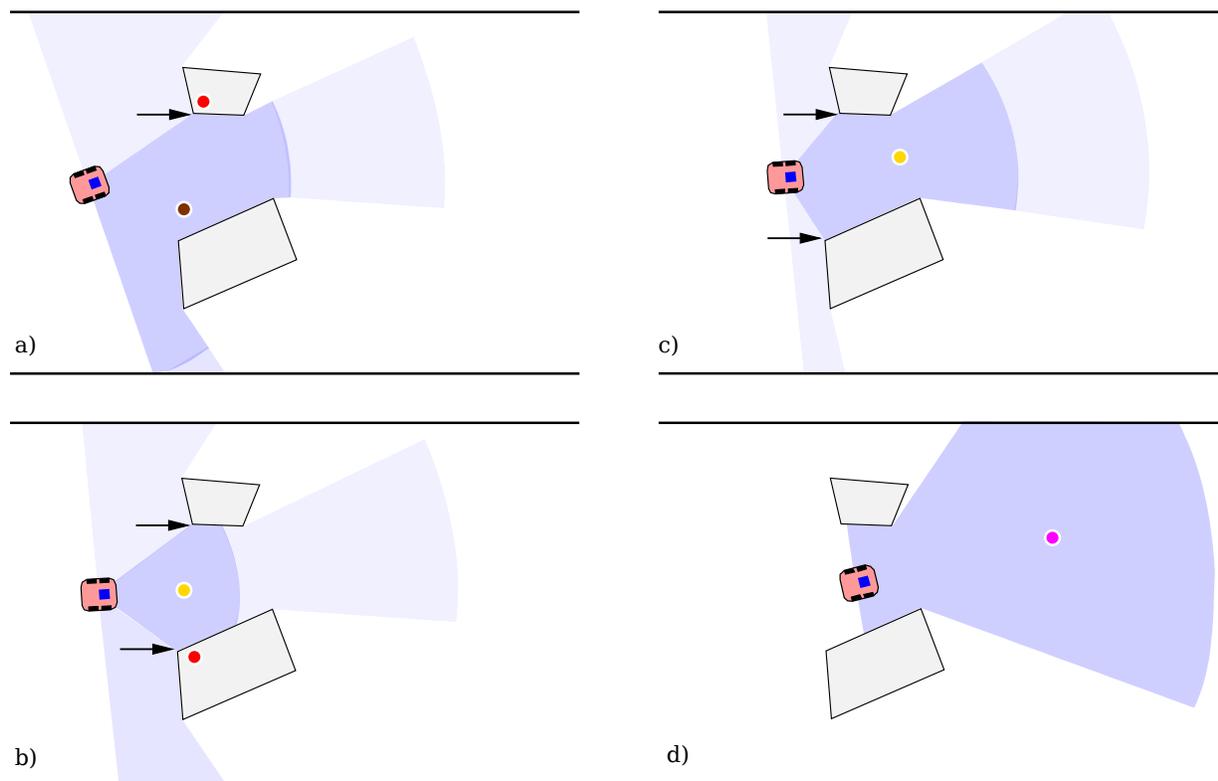


Figura 7.3: Divisiones sucesivas para pasar entre dos obstáculos próximos. Los símbolos usados son los mismos que en la figura 7.2 . En a) el centro de áreas del sector de avance completo acaba de volverse inaccesible dentro del obstáculo superior. El proceso de división produce un punto de división y un nuevo centro de área cerca del obstáculo inferior. En b) debido al movimiento del robot para encara el nuevo centro, el nuevo centro se vuelve inaccesible. Un nuevo proceso de división se lleva a cabo resultando un nuevo punto de división y un nuevo centro de área. El sector de avance ha sido comprimido de nuevo debido a la reducción de amplitud, de manera que el robot se moverá más lento, ver sección 7.4. Ahora el centro de área está aproximadamente situado en la mitad de el camino entre los dos obstáculos. En c) el robot ha estado siguiendo el centro y expandiendo su sector de avance mientras ha ido dejando atrás los puntos de división. En d) el sector completo ha sido recuperado y ahora el robot tiene el camino despejado.

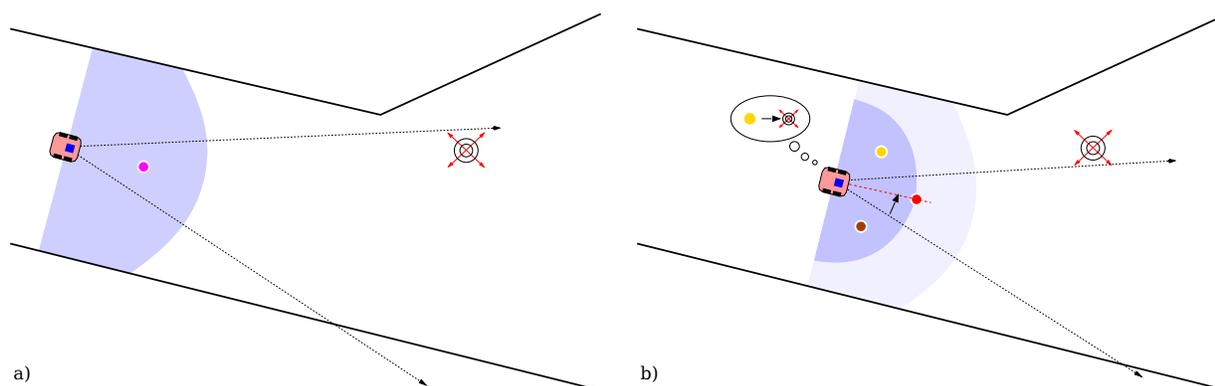


Figura 7.4: Alcance de objetivos usando el método del centro áreas. Los símbolos y la nomenclatura son los mismo que en la figura 7.2. El objetivo es un emisor de estímulos. El robot tiene sensores que aproximadamente detectan la dirección del estímulo. a) Mientras la dirección del estímulo esté dentro del sector más frontal del robot, marcado con flechas, el robot seguirá al centro de áreas. b) Cuando la dirección del estímulo sale del sector frontal se considera que existe un obstáculo virtual sobre el centro de áreas y se aplica el proceso de división eligiendo el sector del lado orientado al emisor, el punto de división se sitúa en la línea que une al robot con el centro de áreas original, algo cerca del robot para se recupere pronto la amplitud completa del sector. Aplicando este proceso iterativamente el robot alcanza la velocidad del objetivo, entonces se ejecuta un procedimiento de parada basado en la intensidad del estímulo.

7.3. Alcance de objetivos

El alcance de objetivos (goal reaching) de forma reactiva se puede conseguir con una ligera modificación del método expuesto hasta el momento. El objetivo tiene que ser una fuente de estímulo y el robot debe percibir su intensidad y dirección aproximadas. La intensidad es principalmente usada para elicitar la ejecución un procedimiento de parada cuando es lo suficientemente alta, y por lo tanto el robot estará cerca de la fuente. La dirección es usada para sugerir la preferencia por un determinado lado en el proceso de división.

Se usará un procedimiento de parada muy simple. Cuando la intensidad del estímulo alcance un determinado valor el procedimiento de parada tomará un control relativo sobre la velocidad del robot forzando una disminución, el robot se parará cuando la intensidad alcance otro umbral superior.

Se mencionó en la subsección 7.2.2.1 la selección entre dos sectores con centros de área accesibles puede realizarse aleatoriamente o mediante una preferencia externa. Del conocimiento de la dirección aproximada del estímulo se deduce el lado del robot del que viene el estímulo y se dispone de una preferencia de lado para el proceso de división. Aunque de esta manera el objetivo solo es alcanzado por suerte, es un método pasivo.

Para conseguir una búsqueda activa se van usar obstáculos virtuales. Si la dirección del estímulo se desvía de la del robot, el método sitúa un obstáculo virtual en la posición

del centro de áreas causando un proceso de división con una preferencia de lado definida, ver figura 7.4. Se usa un sector angular frontal para determinar cuando la dirección del estímulo se ha desviado lo suficiente de la dirección del robot. La amplitud de este sector frontal depende inversamente de la intensidad del estímulo, forzando al robot a encarar de forma más precisa al objetivo cuando esté cerca. Para no interferir con la evitación de obstáculos en esta clase de proceso de división, el sector sugerido debe tener un alta medida de conveniencia para ser seleccionado, para compensar el intento de dividir un sector resultante de una división previa debida a un obstáculo real. El punto de división se sitúa en la línea que une al robot con el centro de áreas original, algo cerca del robot para se recupere pronto la amplitud completa del sector de avance.

7.4. El control del movimiento

Como se ha comentado anteriormente el robot ha seguir al centro de áreas, pero ¿qué significa exactamente seguir al centro de áreas? Podría significar que el robot tien que seguir exactamente la misma trayectoria que el centro, la más segura, pero esta implica diversas cuestiones. Los sensores tiene un periodo de muestreo no nulo, por consiguiente solo se obtienen algunos puntos de la trayectoria del centro de áreas y solo en algunos instantes. Para sensores de rango, espacialmente se obtienen los puntos de un polígono, no una curva continua, y fallos en los sensores son usuales, esto produce una trayectoria del centro de áreas que no es suave, cuanto menor sea el número de puntos menor la suavidad. Por lo que se debería estimar secuencialmente la trayectoria para dar los comandos de movimiento oportunos. Pero, incluso aunque se pudiera estimar dicha trayectoria, todavía queda un problema, el proceso de división hace que la trayectoria del centro de áreas sea esencialmente discontinua.

Como este trabajo versa sobre conductas reactivas se va a tratar de diseñar un procedimiento, en la línea de la hipótesis de trabajo, basado simplemente en la distancia del centro de áreas al robot, d , y el ángulo entre la dirección de éste y la del centro de áreas en coordenadas locales ψ .

La velocidad lineal del robot está directamente relacionada con d , un valor alto de d mayor generalmente indica un área libre grande. El valor de d necesita ser normalizado. En el caso de un sector circular de amplitud $\alpha > 0$ y radio r , es $d = \frac{2\sin(\alpha)}{3\alpha}r$, por lo que d es una función decreciente de α . Así áreas estrechas, y por lo tanto menos seguras podrían tener valores de d elevados, haciendo al robot moverse peligrosamente rápido. Por consiguiente, se usará un parámetro de normalización de distancia $\eta = \frac{2\sin(\alpha)}{3\alpha}r_{max}$, siendo r_{max} el rango máximo considerado del sensor. Usualmente $d \leq \eta$ para los polígonos que se forman habitualmente, así que usualmente $0 \leq \frac{d}{\eta} \leq 1$.

El ángulo ψ está relacionado con la velocidades lineal y radial, ver 7.5.

Si ψ es pequeño, $|\psi| \leq \psi_{min}$ $\psi_{min} > 0$, se avanza en línea recta, así se evita el cabeceo del robot. Si ψ es grande, $|\psi| \geq \psi_{max} > 0$, entonces la velocidad lineal del robot debe ser nula y debe girar en dirección al centro de áreas. Cuando $\psi_{min} < |\psi| < \psi_{max}$ la velocidad lineal, v_{lin} , debe estar inversamente relacionada con $|\psi|$ y la radial, v_{rad} , directamente. Se

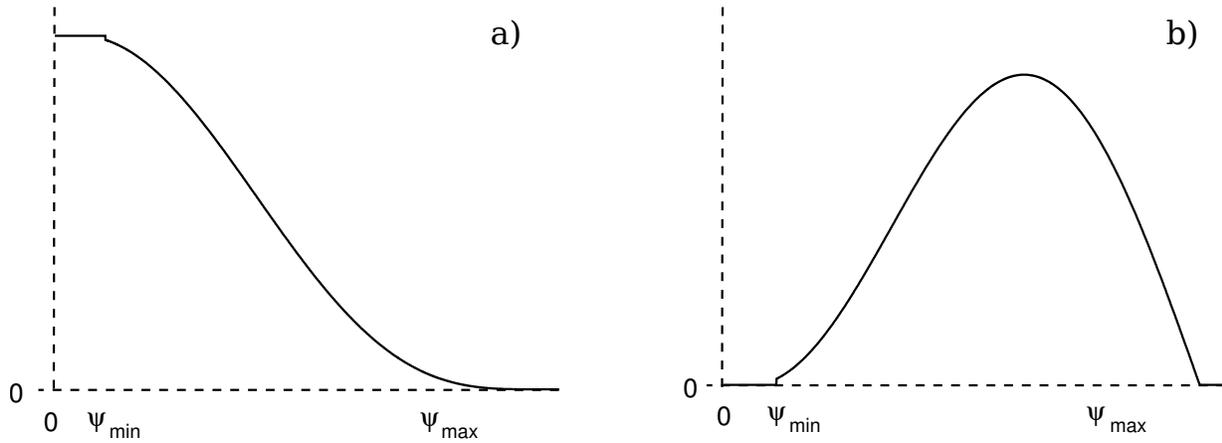


Figura 7.5: Control de movimiento. Las velocidades lineal y radial son funciones de la distancia del robot al centro de áreas, d , y el ángulo formado entre la dirección de avance del robot y la línea que une al robot con el centro de áreas ψ . a) Velocidad lineal en función de ψ para d constante. b) Velocidad radial en función de ψ para d constante.

usa el ángulo ψ' que guarda la misma proporción con ψ que ψ_{\max} con $\frac{\pi}{2}$, así $\psi' = \frac{\pi}{2} \frac{\psi}{\psi_{\max}}$.

Las fórmulas, originalmente derivadas para las velocidades de cada rueda por separado en un sistema de dirección diferencial, se obtuvieron experimentalmente y son:

$$v_{lin} = \frac{v d}{2 \eta} \cos(\psi)(1 + \cos(2\psi'))$$

$$v_{rad} = \text{sgn}(\psi) \frac{v d}{l \eta} \cos(\psi)(1 - \cos(2\psi'))$$

siendo $\text{sgn}(\cdot)$ la función signo, v la velocidad deseada cuando no hay ningún obstáculo, y l la longitud del eje del robot. Como es usual que $0 \leq \frac{d}{\eta} \leq 1$, pero no siempre, se tendrá que el máximo de v_{lin} podrá ser un poco superior a v . Los valores de ψ_{\max} que se han encontrado adecuados en los experimentos han sido los comprendidos entre 90 y 100 grados, nunca se han observado valores de $|\psi|$ por encima de 80.

7.5. Experimentos simulados

El método descrito en las secciones anteriores se implementó en el simulador Cybersim, en desarrollo por nuestro equipo Cuadra Troncoso (2011), para comprobar su validez bajo diversas circunstancias.

7.5.1. Descripción del robot

El robot simulado es similar a un Pioneer-3AT con dos ruedas cuyas velocidades se controlan independientemente. Los sensores de rango fueron modelados como se describe

en Cuadra Troncoso (2008); Thrun et al. (2005) y se simuló un sonar de haz estrecho, en un ambiente estático ($w_{short} = 0$), sin ruido ni cross-talking entre sensores ($w_{rand} = 0$). El modelo de robot usado en estas simulaciones incluía 36 sensores de rango alrededor del robot distribuidos cada 10° , para compensar el estrecho haz de detección. Los elementos del mundo son simulados según dos modelos diferentes:

- Sensores ideales: todos los objetos en el rango son detectados independientemente del ángulo de incidencia, ($\alpha_{hit} = 90^\circ$ y $\alpha_{max} = 90^\circ$) en la notación de Cuadra Troncoso (2008); Thrun et al. (2005).
- Sensores realistas: los obstáculos son detectables dependiendo del ángulos de incidencia, si el ángulo de incidencia del haz es menor que 30° el obstáculo es detectado, si está entre 30° y 60° la probabilidad de detección va de 1 a 0 linealmente, para ángulos mayores que 60° no se detecta el obstáculo, ($\alpha_{hit} = 30^\circ$ y $\alpha_{max} = 60^\circ$).

En ambos casos se usó un alcance máximo de 5 m. El error en la distancia medida tenía una distribución gaussiana una desviación típica $\sigma = 5$ mm, este dato se tomó de las especificaciones de dispositivos comerciales como el s3onar SensComp, Inc. (2004) y el l3aser SICK AG (2006).

7.5.2. Comparaci3n con el procedimiento de wandering de la librer3a Aria

El procedimiento de “wandering” propuesto en la secci3n 7.2 se va a comparar con el m3todo que proporciona la conocida librer3a rob3tica Aria (Mobile Robots) usada por el robot Pioneer 3AT y otros, con sensores de sonar y l3aser simulados mediante el software MobilSim que acompa3a a dicha librer3a .

En la simulaci3n en MobilSim, se us3 la clase ArActionGroupWander de Aria. El grupo de acciones se compone de dos, ArActionAvoidFront con los siguiente par3metros: avoidFrontDist = 450 mm, avoidVel = 200 mm/s, avoidTurnAmt = 15° , priority = 79; y ArActionConstantVelocity con los siguiente par3metros: forwardVel = 500 mm/s, priority = 50. Estos son los valores por defecto salvo la velocidad que fue cambiada de 400 a 500.

Se usaron dos clases de mundos para la prueba. El primer mundo tiene la distribuci3n de obst3culos t3pica de un entorno dom3stico o de oficina con paredes en 3ngulo recto, puertas y algunos muebles. Cada experimento dur3 30 minutos. La figura 7.6 muestra cuatro ejemplos de wandering, con dos diferentes puntos de partida simulados en CyberSim. El arco punteado representa el m3ximo rango considerado para el sensor, 5 m. En a) y b) usando el centro de 3reas, se observa que el robot ha deambulado por todas las regiones alcanzables del mundo (habitaciones) siguiendo trayectorias suaves. En c) y d) usando el m3todo de la librer3a Aria con MobilSim, el robot solo deambula por algunas partes del mundo, dependiendo del punto inicial, y las trayectorias no son suaves.

El otro tipo de mapa usado es un circuito cerrado con unos pocos obst3culos, con una longitud total de 81 m. La figura 7.7 a) muestra el recorrido realizado usando el m3todo del centro de 3reas, la trayectoria es suave y cubre todas las v3as alternativas. Por otro

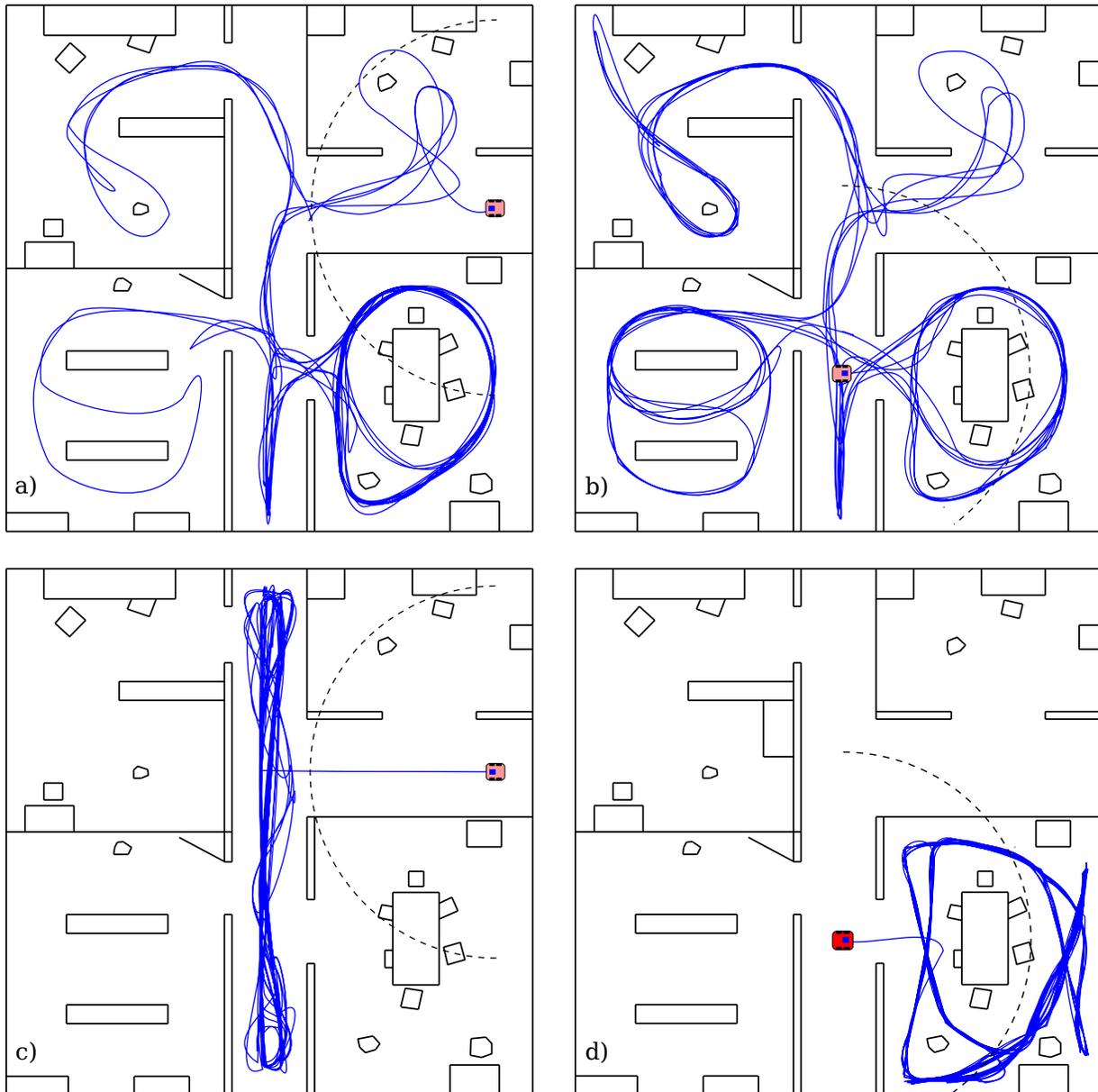


Figura 7.6: Trayectorias seguidas por el robot al deambular por un entorno doméstico o de oficina. En a) y b) usando el método del centro de áreas, en c) y d) el procedimiento de wandering de la librería Aria. El arco punteado indica el máximo rango considerado, 5 m.

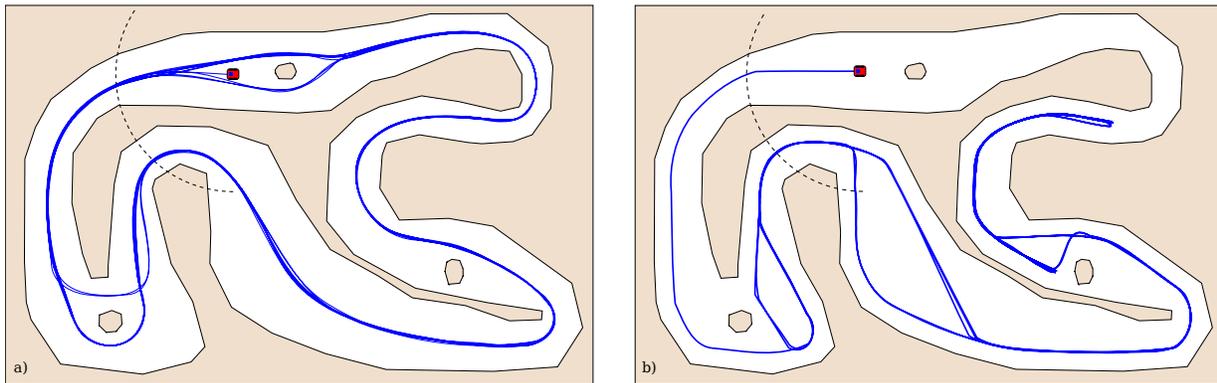


Figura 7.7: Comparación de recorridos en un circuito cerrado, en a) usando el método del centro de áreas, en b) el procedimiento de wandering de la librería Aria. El arco punteado indica el máximo rango considerado, 5 m.

lado en b), usando la librería Aria con MobilSim no se obtienen trayectorias suaves, sino que suelen seguirse las paredes y el robot no llega a pasar por la parte más estrecha del circuito. El arco punteado representa el máximo rango considerado para el sensor, 5 m. Cada experimento duró 30 minutos.

7.5.3. Evitación simple de obstáculos para alcanzar objetivos

Estos experimentos se llevaron a cabo a fin de comprobar el comportamiento del método del centro de áreas ante algunos problemas comunes a métodos, reactivos y no, de navegación que se comentaron en 6.2.1 como paso entre obstáculos cercanos y trampas en forma de C y U mientras se intenta alcanzar un objetivo.

El modelo de robot usado en estas simulaciones tiene 36 sensores de rango equidistribuidos alrededor del robot cada 10° . El robot recibe un estímulo, por ejemplo un sonido, de manera que puede identificar la dirección de procedencia. Se supone que el estímulo no está afectado por la presencia de obstáculos pero la detección de su dirección es solo aproximada. La dirección del estímulo se usa, como se comentó en la subsección 7.3, para establecer la preferencia de lado en el proceso de división del sector de avance mediante obstáculos virtuales.

El primer conjunto de experimentos se colocó un obstáculo aproximadamente circular entre el robot y el objetivo, se usaron círculos de diversos tamaños y se colocó el objetivo a varias distancias del objeto. En todos los la trayectoria del robot estuvo bastante cerca de la trayectoria cuasi-óptima.

En el segundo conjunto de experimentos las formas de los obstáculos eran cóncavas, C o U, con diferentes tamaños y colocando el objetivo a diversas distancias del obstáculo en el lado opuesto al robot. La figura 7.9 muestra cuatro casos de alcance del objetivo sorteando un objeto cóncavo. En los casos mostrados el robot conseguía escapar de la trampa en forma de C o U, obstáculos más anchos podrían impedir al robot escapar.

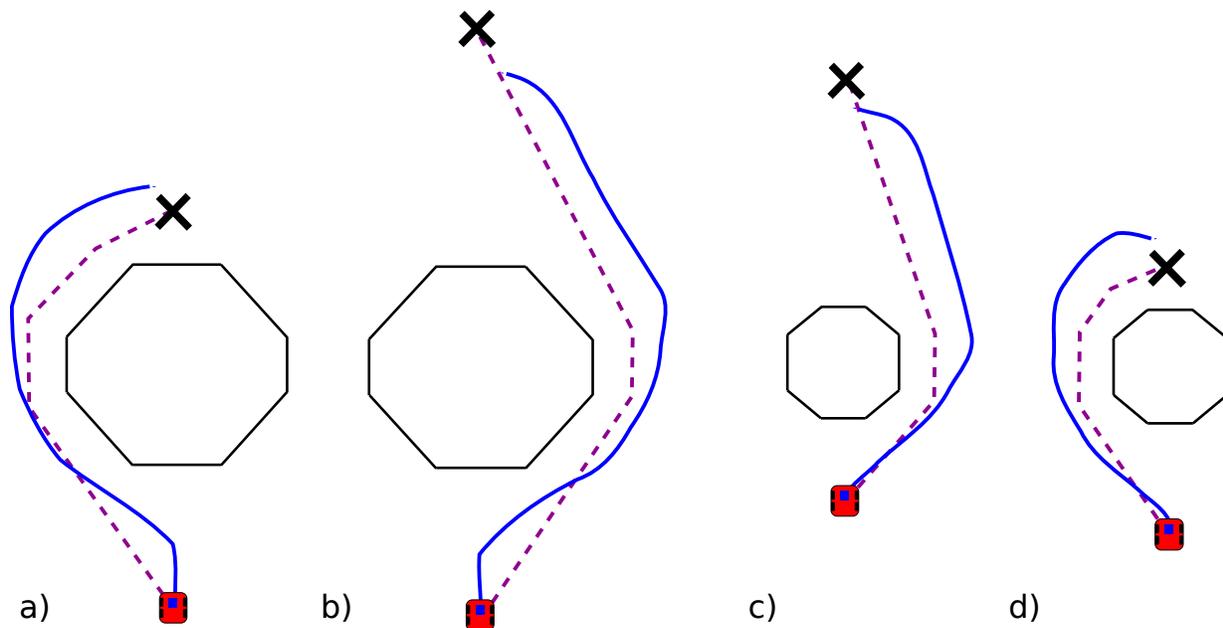


Figura 7.8: Evitación de obstáculos con el objetivo a diferentes distancias del objeto aproximadamente circular. En a) y b) el diámetro del círculo es de 4 m. y en c) y d) de 2. En todos los casos la línea continua muestra la trayectoria del robot al objetivo (cruz) y la punteada es una trayectoria cuasi-óptima alrededor del obstáculo.

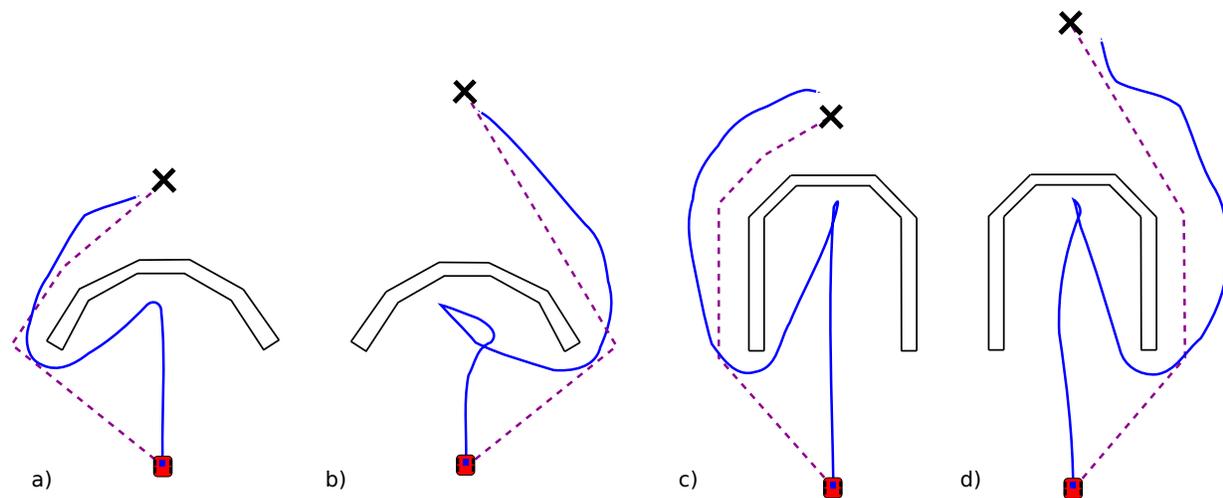


Figura 7.9: Evitación de obstáculos con el objetivo a diferentes distancias de objeto cóncavos. En a) y b) un objeto en forma de C de 5,375 m. de anchura y 2,058 m. de profundidad. En c) y d) un objeto en forma de U de 3,7 m. de anchura y 4,5 m. de profundidad. En todos los casos la línea continua muestra la trayectoria del robot al objetivo (cruz) y la punteada es una trayectoria cuasi-óptima alrededor del obstáculo.

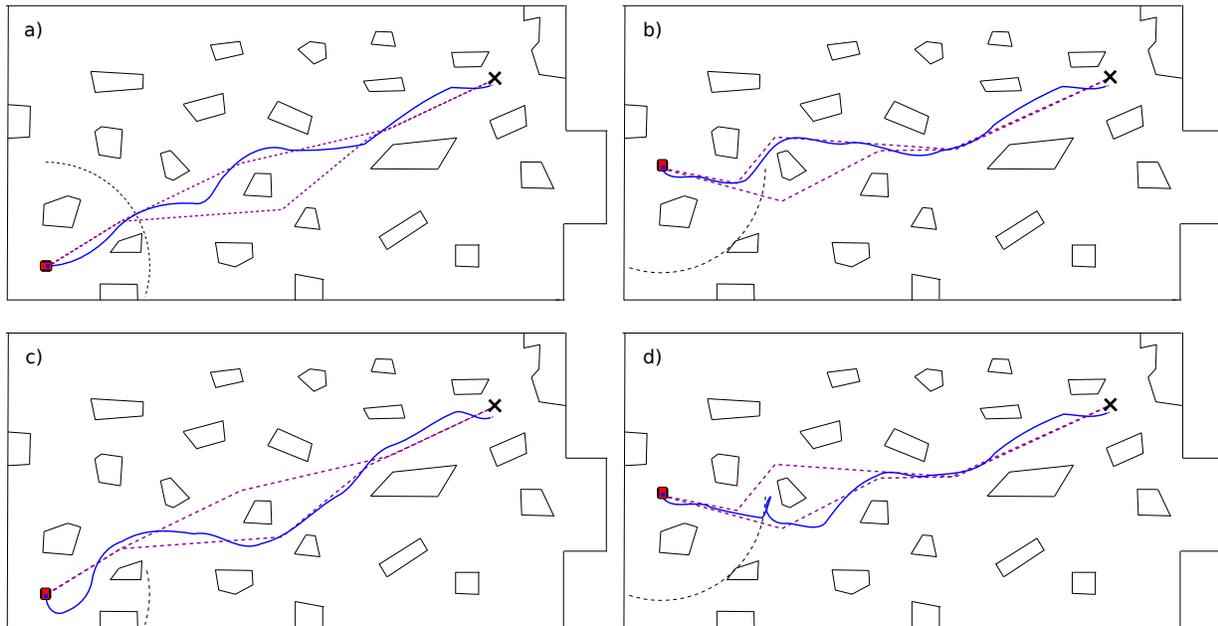


Figura 7.10: Alcanzando objetivos en mundos complejos simulando sensores ideales. En a) y b) recorridos con un buen rendimiento, muy poco exceso de distancia recorrida. En c) y d) recorridos con un mal rendimiento, poco exceso de distancia recorrida. En todos los casos la línea continua muestra la trayectoria del robot al objetivo (cruz) y la punteada es una trayectoria cuasi-óptima al objetivo. El arco punteado indica el máximo rango considerado, 5 m.

7.5.4. Evitación de obstáculos complejos para alcanzar objetivos

En estos experimentos se pretende probar la capacidad del método del centro de áreas para alcanzar objetivos en mundos complejos con poca estructuración, los experimentos se realizarán simulando sensores ideales y realistas comentados en la subsección 7.5.1 y se compararán los resultados. Se eligieron cuatro posiciones iniciales y cuatro orientaciones: 0, 90, 180, 270 grados. Para cada combinación de posición y orientación se realizaron 3 ejecuciones, lo que da un total de 96 experimentos para cada tipo de sensor. El diámetro del robot es de 55 cm. y se ajustó su velocidad máxima a 63 cm/s . Las longitudes cuasi-óptimas (medidas a mano) para cada posición inicial eran: 23,527 m., 23,166 m., 21,979 m. and 11,897 m. La proporción entre la longitud recorrida por el robot con respecto a la longitud óptima se usó como medida de rendimiento, que es siempre ≥ 1 . Ejemplos de ejecuciones con buenos y malos rendimientos, se muestran en la figura 7.10 para los sensores ideales y en la figura 7.10 para los realistas.

En los experimentos con los sensores ideales los resultados estadísticos para el rendimiento fueron mínimo 1,0299, máximo 1,5619, media 1,2105, mediana 1,1721 y desviación estándar 0,1396. En la figura 7.12 a) se puede ver que en el 83,33% de los casos las longitudes recorridas se han incrementado menos del 30%, usando sensores ideales. Usando sensores realistas 7.12 b) en el 83,33% de los casos las longitudes recorridas se han incre-

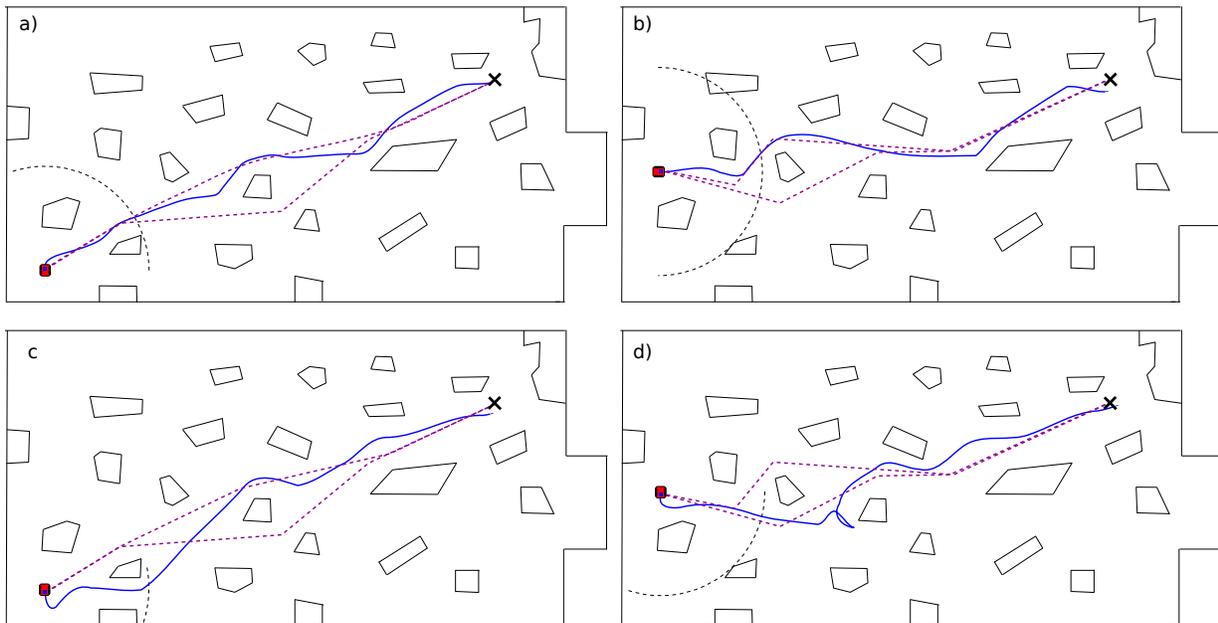


Figura 7.11: Alcanzando objetivos en mundos complejos, simulando sensores realistas. En a) y b) recorridos con un buen rendimiento, muy poco exceso de distancia recorrida. En c) y d) recorridos con un mal rendimiento, más exceso de distancia recorrida. En todos los casos la línea continua muestra la trayectoria del robot al objetivo (cruz) y la punteada es una trayectoria cuasi-óptima al objetivo. El arco punteado indica el máximo rango considerado, 5 m.

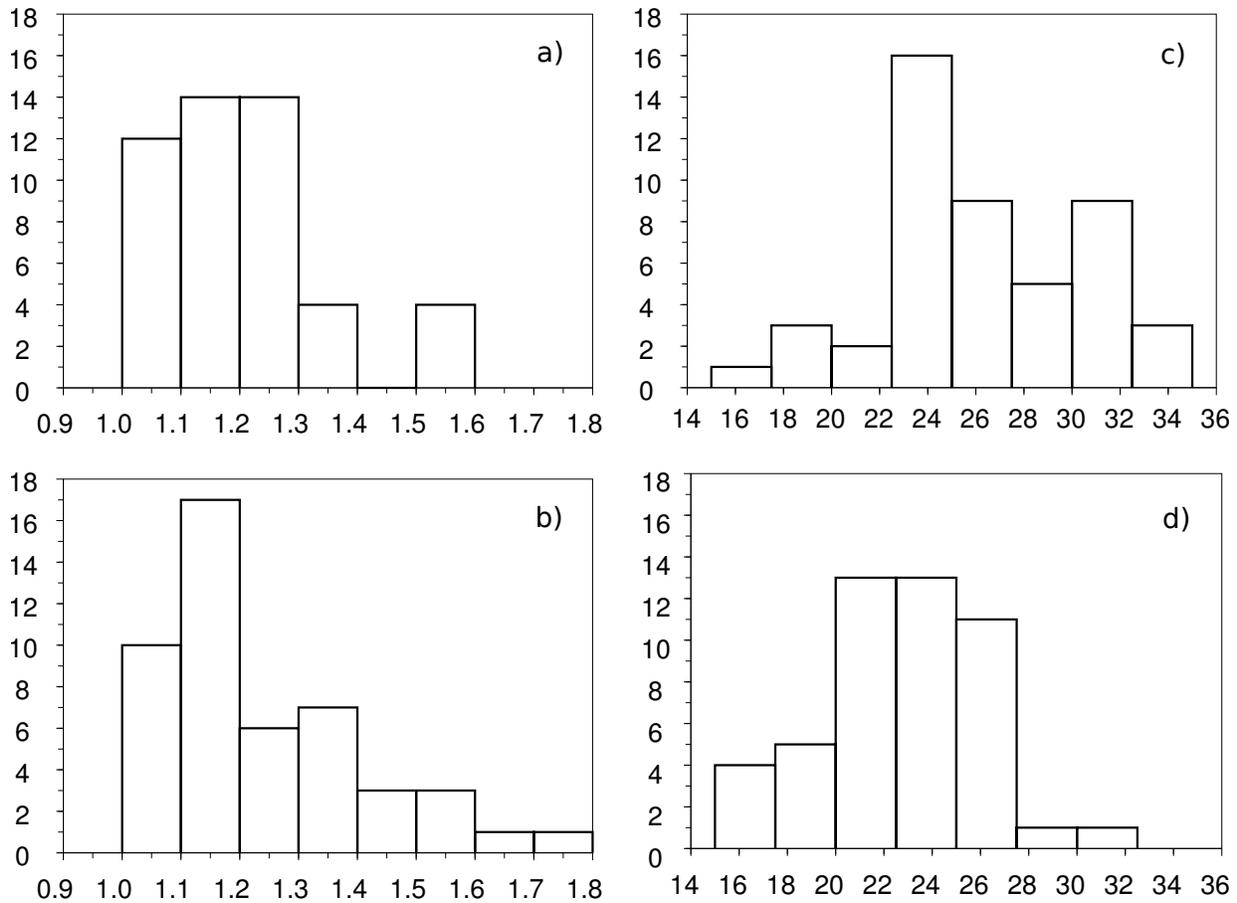


Figura 7.12: Alcanzando objetivos en mundos complejos, simulando sensores realistas. En a) y b) recorridos con un buen rendimiento, muy poco exceso de distancia recorrida. En c) y d) recorridos con un mal rendimiento, más exceso de distancia recorrida. En todos los casos la línea continua muestra la trayectoria del robot al objetivo (cruz) y la punteada es una trayectoria cuasi-óptima al objetivo. El arco punteado indica el máximo rango considerado, 5 m.

mentado menos del 40 %. Para las velocidades, usando sensores ideales c), se consigue una velocidad media de $26,3016 \text{ cm/s}$ y una desviación estándar de $4,2338 \text{ cm/s}$, usando sensores realistas d), se consigue una velocidad media de $22,8285 \text{ cm/s}$ y una desviación estándar de $3,3839 \text{ cm/s}$.

La comparación entre sensores ideales y realistas muestra que aún perdiendo un 50 % de las medidas debido a malos rebotes el rendimiento del método del centro de áreas solo se ve reducido ligeramente. Estos resultados muestran además en buen rendimiento del método para alcanzar objetivos sin ninguna planificación.

7.6. Experimentos con robots reales

7.6.1. Descripción general

En estos experimentos el robot tenía que encontrar y alcanzar un objetivo, un panel verde de $2 \times 2 \text{ m.}$, localizado en la esquina opuesta del punto de partida del robot. El mundo era un restaurante de facultad fuera de uso, sus dimensiones eran $24,4 \times 8,8 \text{ m.}$, ver figura 7.13. Se usaron 50 cajas cúbicas de cartón de 50 cm. de lado, para crear obstáculos, algunos para cubrir desperfectos del suelo. La altura de las cajas era lo suficientemente pequeña como para no ocultar el objetivo. Había otros obstáculos fijos como columnas, la barra y una máquina. La barra tenía un rodapié por debajo del nivel de los sensores de rango, por lo que no era detectable. La pared enfrente de la barra estaba cubierta completamente de ventanas con persianas venecianas. La iluminación artificial consistía en tubos fluorescentes blancos y de luz de día mezclados sin orden.

Para este experimento se usó un robot Pioneer 3AT equipado con un láser SICK LMS200 cubriendo todo el área frontal, 180 grados, 8 sonar Polaroid cubriendo la parte trasera del robot y una cámara pan-tilt-zoom Canon VC-C50i con ángulo de visión de 48 grados. El láser se usó para los cálculos del centro de áreas del sector de avance, el sonar para el cálculo del centro de áreas trasero para la maniobra de escape y la cámara para buscar el objetivo. El máximo de los sensores de rango fue fijado en 5 m., como en los experimentos anteriores. Se usó la librería de localización Arnl de Mobile Robots solo con el propósito de monitorizar, grabar trayectorias, en ningún momento para la navegación.

7.6.2. Localización del objetivo

En este experimento no se tenía el objetivo de desarrollar un sistema de seguimiento visual sino simplemente testear el método del centro de áreas en situaciones reales, para conocer mejor sus virtudes y defectos.

Para encontrar el objetivo se usó ACTS un programa para procesamiento de vídeo. El programa puede dar información acerca de manchas de color (blobs) en una imagen: áreas, coordenadas del centro de áreas, etc. Debido a las dispares condiciones de iluminación: luz del día entrando por las ventanas, diferentes tipos de fluorescentes, reflexiones, etc., el programa se entrenó desde varias posiciones con el color observado del objetivo. Se usó la

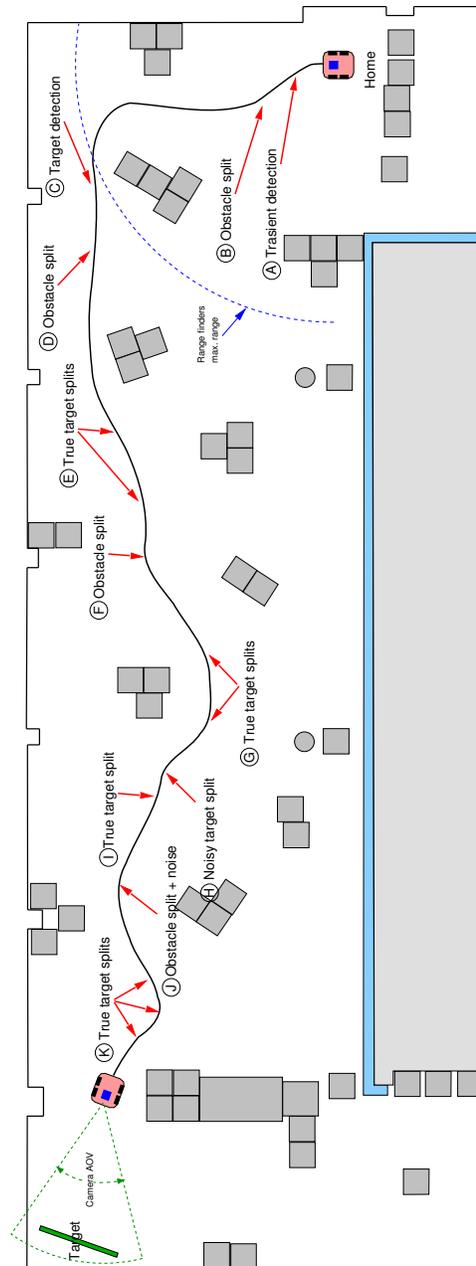


Figura 7.13: Experimento con un robot real. El robot tenía que buscar y alcanzar un objetivo visual. Las dimensiones de la habitación son $24,4 \times 8,8$ m., la distancia en línea recta del robot al objetivo es de 23,5 m. Los puntos de la trayectoria donde el sector de avance fue restringido (dividido y/o comprometido) están marcados con flechas y letras, entre ellos el robot navega siguiendo el centro de áreas del sector de avance. Ver sección 7.6.3 para una explicación completa.

función de balance de blancos automático de la cámara para compensar los cambios de luz, pero este procedimiento toma un tiempo durante el cual es fácil que la cámara pierda el blob teniendo que comenzar el proceso de búsqueda, ver más abajo.

El movimiento alrededor del eje vertical al suelo, panning, era el único realizado por la cámara. El intervalo mínimo para enviar comandos a la cámara con la seguridad de que serían ejecutados era de 300 ms., que es tres veces el tiempo del ciclo del robot. Cuando el objetivo se perdía, se enviaban a la cámara comandos para ejecutar un giro incremental; el sector cubierto con este movimiento era de ± 70 grados. Cuando un blob del color deseado era encontrado, y si su área estaba por encima de un umbral para filtrar el ruido tanto como fuera posible, comandos apropiados para el giro eran enviados a la cámara para mantener centrado el blob en la imagen. El umbral fue establecido como el número de píxeles que presentaba el objetivo desde la posición de inicio, 800 píxeles en una imagen de 320×240 . Cuando el ángulo de giro de la cámara, estando el objetivo localizado, quedaba fuera de cierto sector central, ± 20 grados, el proceso de división del sector de avance era disparado.

7.6.3. Resultados

El principal problema en estos experimentos fue la localización y seguimiento del objetivo debido, principalmente, a las cambiantes condiciones de iluminación a lo largo del camino del robot. Los experimentos en los que el robot alcanzó más rápidamente el objetivo se realizaron de noche. Como ejemplo, se va a comentar la trayectoria seguida por el robot en uno de esos experimentos, que se muestra en la figura 7.13. Los comentarios se refieren a los puntos en los que se produjo el proceso de división del sector de avance, y que se marcan con una letra en la figura, en todos los otros puntos el robot simplemente seguía al centro de áreas del sector de avance.

El robot comienza su camino desde el punto de inicio (Home) y la cámara empieza a girar hacia izquierda del robot. En el punto **A** tiene lugar una detección pasajera del objetivo, el área del blob está un poco por encima del umbral. El sector de avance es dividido (split) y su sub-sector izquierdo se convierte en el nuevo sector de avance, dirigiendo al robot hacia el objetivo, pero la cámara ha perdido el objetivo y continúa girando en su búsqueda. En **B** en centro de áreas se vuelve inaccesible dentro del obstáculo con forma de T invertida y un nuevo proceso de división comienza. Como el objetivo no está localizado no hay preferencia de lado, el sub-sector derecho es elegido aleatoriamente. El centro de áreas conduce al robot entre los obstáculos de una forma segura. En **C** el área del blob del objetivo tiene unos 1000 píxeles y cuando la cámara lo localiza comienza un proceso de división, ahora la preferencia de lado viene determinada por la dirección de la cámara, la cámara, a su vez, comienza en proceso de centrado en el blob.

El obstáculo frontal en **D** produce un nuevo proceso de división, con preferencia de lado establecida por el ángulo de giro de la cámara. Una vez que el obstáculo es sobrepasado, sucesivos procesos de división **E**, debidos al ángulo de la cámara, orientan al robot en la dirección correcta. Un nuevo obstáculo **F** se encuentra en la mitad del camino, nuevo proceso de división con preferencia a la izquierda. Dos procesos de división sucesivos **G**, debidos al ángulo de la cámara, hacen que el robot supere el obstáculo.

Un ajuste de la cámara \textcircled{H} provoca la aparición de ruido expandiendo el blob del objetivo hacia la izquierda, el ángulo de giro de la cámara varía lo suficiente como para provocar un nuevo proceso de división, esta situación es apreciable en el vídeo del experimento. Pronto \textcircled{I} el verdadero blob es descubierto y se inicia un nuevo proceso de división debido al ángulo de la cámara. En \textcircled{J} un obstáculo y el ángulo de giro de la cámara hacen al robot girar hacia la izquierda, pero el objetivo casi se pierde debido al ruido. Una vez que el blob verdadero es recuperado, tres procesos de división sucesivos \textcircled{K} conducen al robot cerca del objetivo. Entonces se ejecuta el proceso de parada.

La trayectoria del robot muestra las mismas características que las que se observaban en los experimentos simulados: es suave y el robot evita los obstáculos de forma segura. Una vez que el objetivo es localizado con un área suficiente es alcanzado siguiendo una ruta corta, siempre que los ajustes de la cámara lo permitan.

Capítulo 8

Centro de áreas con segmentación

El estudio de los experimentos descritos en el capítulo 7.6, nos llevó a darnos cuenta de que en ciertas circunstancias la posición del centro de áreas llevaba al robot a situaciones de riesgo o de choque. Simulando dicho experimento se pudo comprobar que en esas ocasiones la posición del centro de áreas ocupaba lugares inesperados que hacían que el proceso de división produjera sectores poco convexos y por los tanto peligrosos, ver subsección 7.2.2. Se llegó a la conclusión de que para avanzar en la navegación por centros de áreas era necesario mejorar el modelo del entorno. El modelo de área libre percibida con el que ha estado trabajando es el de un polígono, pero este polígono se halla formado por numerosos vértices afectados de ruido. La idea que se nos viene a la mente al pensar en un polígono representando al entorno, poco tiene que ver con la de un conjunto numerosos de medidas ruidosas, más bien con la de un conjunto de pocos segmentos bien definidos. Si las medidas de rango se filtraran previamente mediante un procedimiento de segmentación sí se tendría un polígono que representara un nivel de abstracción superior.

Varios de los problemas observados en el método básico derivan todos de la misma causa, sectores poco convexos, o sea, con esquinas salientes y otras protuberancias. Pero se han observado dos situaciones distintas en las que se producen y que llevan aparejadas dos modificaciones del método del centro de áreas básico. En este capítulo se estudiarán los problemas mencionados, las soluciones encontradas mediante el uso de la segmentación y algunas pruebas realizadas.

8.1. Localización adecuada de los puntos de división

En los experimentos descritos en la sección 7.6 se observó una situación que llevaba en bastantes ocasiones al robot a iniciar la maniobra de escape tras un proceso de división, lo que le hacía perder el objetivo. El problema viene de la selección del punto de división que se realiza: se determina de entre los dos rayos del haz entre los que está el centro de áreas, cuál de los dos vértices correspondientes del sector de avance está más cercano al robot. La cuestión está en que la posición del centro de áreas no tiene porqué ser la del punto más intrusivo de su borde, como puede verse en la figura 8.1. En esta figura se

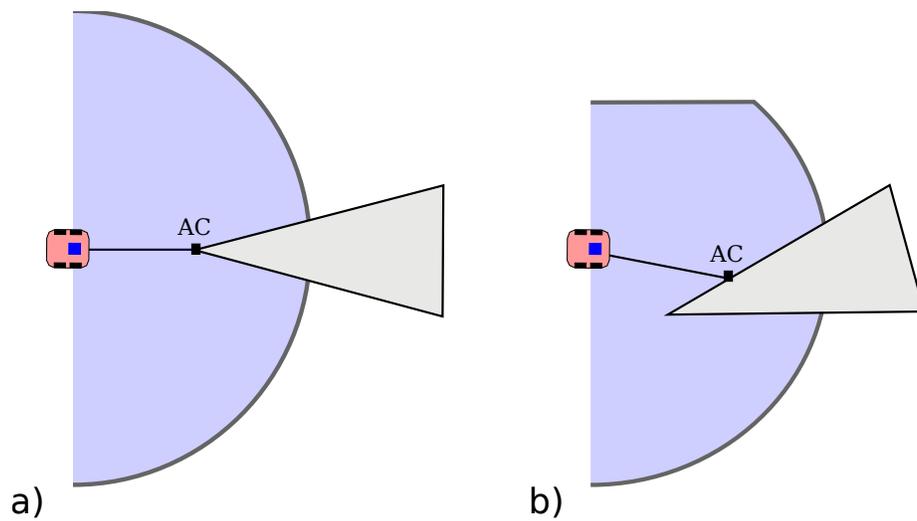


Figura 8.1: Mala situación del centro de áreas. En a) tenemos el caso ideal, el centro de áreas se vuelve inaccesible en el punto más crítico para la marcha del robot, justo en la pico acusado que constituye el punto más peligroso. En esta situación los dos sub-sectores en los que se dividirá el sector de avance, serán convexos y no habrá problemas para la navegación. Pero en b), con una configuración similar de obstáculos puede producir una situación en la que cuando el centro de áreas se vuelve inaccesible, uno de los subsectores en los que se dividirá el sector presenta un pico acusado, si este sector es elegido puede producirse una situación peligrosa.

plantea la cuestión de forma muy simplificada, en la realidad se necesitan configuraciones más complejas para que se observe la situación.

En la figura 8.2 se aprecia una situación real ocurrida en los experimentos descritos en la sección 7.6 en la que se produce la situación de mala posición del centro de áreas.

El centro de áreas cuando se vuelve inaccesible en el punto de división marcado con la flecha en a). Debido a la dirección del objetivo el subsector derecho es seleccionado, estos problemas suelen darse más a menudo cuando se persigue un objetivo que cuando simplemente se deambula. Como se aprecia en la figura 8.2 a) el sector elegido no es convexo y presenta un pico acusado en la que será la dirección de movimiento del robot cuando siga al nuevo centro de áreas. La porción de sector de avance comprendido entre el punto de división y el pico saliente hace que la trayectoria del centro de áreas no se aparte lo suficiente del obstáculo b) y c) conduciendo al robot a un choque o a una maniobra de escape.

La solución propuesta para este problema consiste en una forma de determinar la posición del pico saliente comentado, 8.2 d). Está claro que, en cierto sentido, el pico saliente es el punto más próximo al robot en su camino. La cuestión está en como determinar dicho punto. La solución propuesta usa la segmentación de las medidas de rango para disponer de un conjunto poco numeroso de segmentos bien definidos. Supongamos por simplicidad de la exposición que en centro de áreas se vuelve inaccesible a la izquierda del robot, como en la figura 8.2 a). Se define un desplazamiento angular fijo $\psi \geq 0$, se determina el segmento donde el centro de áreas se ha vuelto inaccesible y la bisectriz del sector de avance, que formará un ángulo ν en el sistema de coordenadas locales del robot. Todos los extremos izquierdos de segmentos, desde el que contiene al centro de áreas hacia la derecha, que correspondan a un ángulo menor que $\nu + \psi$ son seleccionados y todos los extremos derechos correspondan a un ángulo menor que $\nu + 2\psi$ son seleccionados. De entre los vértices seleccionados se elige el más próximo al robot como punto de división. Conviene separar el punto de división del segmento al menos una distancia igual al radio del robot, perpendicularmente a la dirección de movimiento, para hacer más segura la trayectoria.

La solución propuesta es muy simple y se podría mejorar analizando qué segmentos son concurrentes y cuáles están separados y otras características que se puedan extraer de la segmentación.

8.2. Puntos de división dinámicos

En la formulación del método del centro de áreas dada hasta ahora la posición de un punto de división permanece fija durante el tiempo de vida de éste, podemos decir que los puntos de división son estáticos. Sin embargo aunque el punto de división pueda colocarse en un lugar adecuado inicialmente, el movimiento del robot puede situar en una mala posición a dicho punto de división.

El problema es muy similar al descrito en la sección 8.1, aunque en este caso la posición inicial del punto de división sí esté en el pico más peligroso, pero el posterior movimiento del robot hace que sea otro pico el que entrañe más peligro. Dado que en ese movimiento

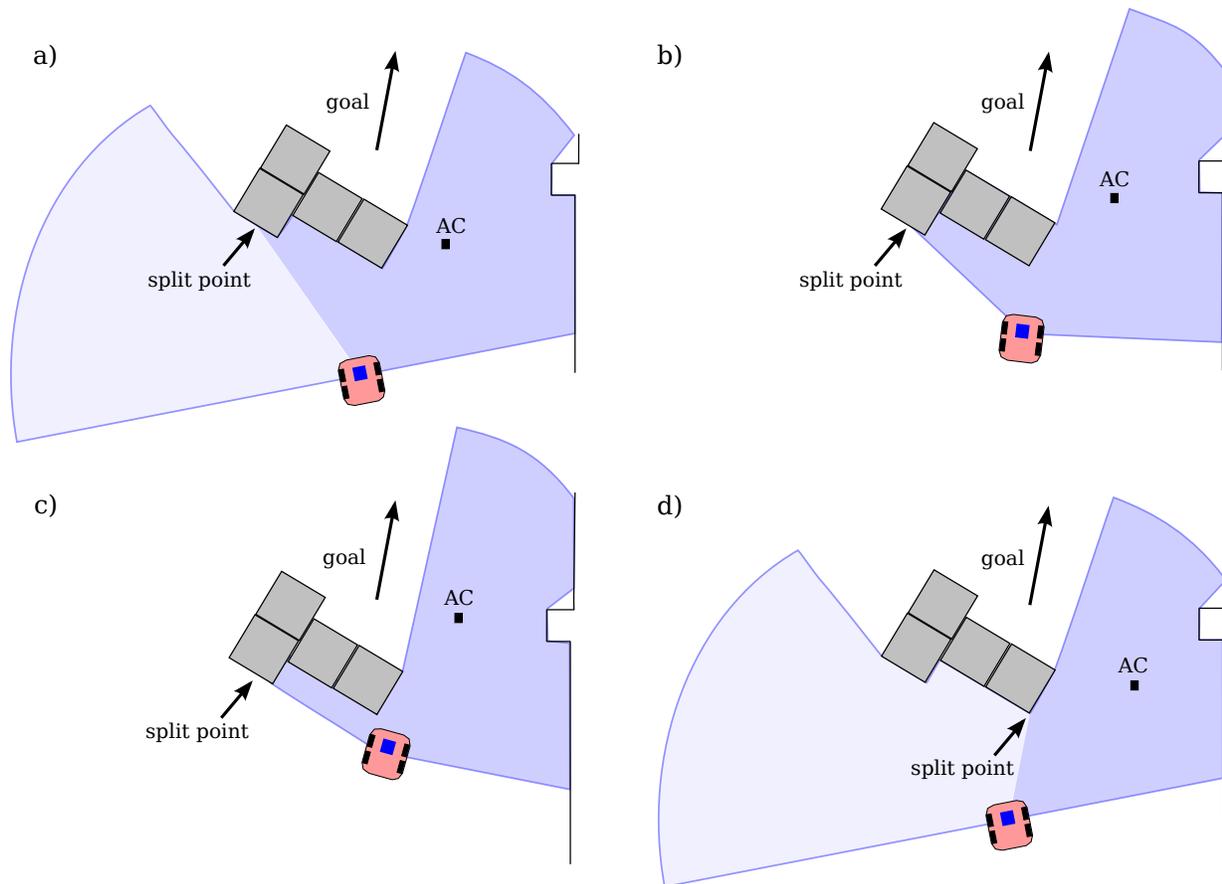


Figura 8.2: Mala situación del centro de áreas. La desafortunada situación del centro de áreas cuando se vuelve inaccesible en el punto de división marcado con la flecha, a). Debido a la dirección del objetivo el subsector derecho es seleccionado y en b) el robot sigue su centro de áreas (AC). Pero la trayectoria del centro de áreas no se aparta lo suficiente del obstáculo, conduciendo al robot a un choque en c). Si el punto de división se sitúa en el pico más saliente en la trayectoria del robot, el centro de áreas se apartaría lo suficiente del obstáculo d), conduciendo al robot por una vía segura. El sector seleccionado tiene un color azul más oscuro que el descartado.

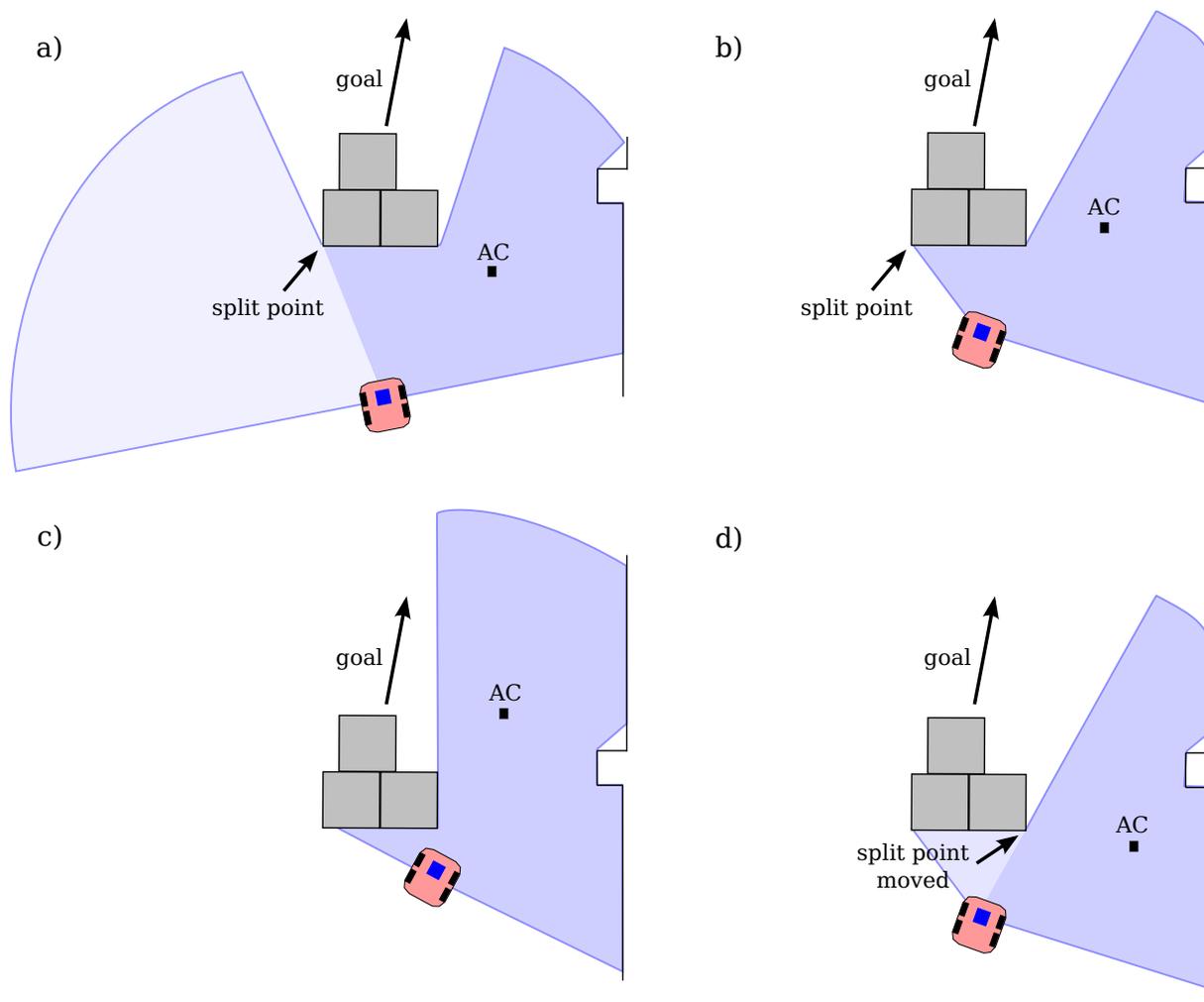


Figura 8.3: Punto de división estáticos y dinámicos. En a) el robot está ligeramente mas cerca del extremo del segmento marcado con una flecha que del otro. Al volverse inaccesible el centro de áreas en ese segmento, se elige el punto marcado como punto de división. Sin embargo la dirección del objetivo hace que se seleccionen el subsector derecho, que queda con un pico en la dirección del movimiento del robot, lo que lleva, b) y d), a una trayectoria insegura como en la figura 8.2. Dado que b) el punto de división ya ha perdido su utilidad sería preferible moverlo hacia el extremo que se encuentre en ese momento más cercano al robot, como se ve en d). El sector seleccionado tiene un color azul más oscuro que el descartado.

el centro de áreas no tiene porqué volverse inaccesible, no se produciría un nuevo proceso de división que podría arreglar el problema. En la figura 8.3 se muestra un ejemplo de esta situación tomado de los experimentos descritos en la sección 7.6. En a) el punto de división, señalado con una flecha se situó en el extremo de segmento más próximo al robot, en el sentido explicado en la sección 8.1. Además en ese momento la dirección de movimiento del robot era hacia dicho punto, por lo que lo constituía como el punto peligroso del recorrido. Sin embargo, debido a la dirección del objetivo, se elige el sector de la derecha. Dado que el punto de división es el extremo izquierdo de un segmento y se ha elegido el sector derecho, dicho segmento queda como lado del sector de avance elegido. La proximidad del segmento al robot, comparado con el resto del entorno, hace que el sector de avance sea poco convexo y se produce entonces una situación como la comentada en la sección 8.1, ver figura 8.3 b) y c).

En la figura 8.3 b) se ve claramente que en ese momento la posición del punto de división ya ha perdido todo su sentido y que sería conveniente que el punto de división estuviera en el pico que realmente está en el camino del robot, señalado con un flecha en la figura 8.3 d). Esta observación nos lleva a dotar a los puntos de división de características dinámicas. De manera que sin tener que mediar un proceso de división, la posición del punto de división se desplace hacia el extremo, de algún segmento, que se encuentre en cada momento más cercano al robot, en el sentido establecido en la sección 8.1. El desplazamiento del punto de división no conlleva un nuevo proceso de división, sino un cambio en los vértices que forman parte del sector de avance y de los descartados.

Al desplazar el punto de división se consiguen áreas más convexas y por lo tanto más seguras para la navegación.

8.3. Experimentos

Para comprobar si las modificaciones introducidas al método del centro de áreas usando segmentación, se procedió a reproducir el restaurante usado en los experimentos de la subsección 7.6. Durante dichos experimentos se había registrado un mapa del restaurante de dicho experimento usando el propio robot Pioneer 3AT, las mediciones láser recogidas durante un paseo conducido del robot, se trataron con el software Mapper3 de MobileRobots para producir un mapa del entorno. Este mapa se usó para producir las figura 7.13. Dicho mapa se usó en el simulador CyberSim para reproducir los experimentos y poder observar detalladamente el comportamiento del robot. Así se pudo descubrir primeramente el origen de los problemas y después testear las soluciones propuestas. No buscábamos problemas concernientes al algoritmo de búsqueda y alcance de objetivos, por lo que la cámara se simuló como un sensor de rango sin errores que solo detectaba el objetivo, recuérdese que en dicho experimentos la altura de los obstáculos no impedía la visión del objetivo por parte de la cámara. El láser se simuló con 181 medidas, que son las que tiene el láser con que está equipado el robot. Se programó el software del simulador de manera que mostrara en la ventana del mundo simulado las posiciones de los puntos de división y poder observar su comportamiento.

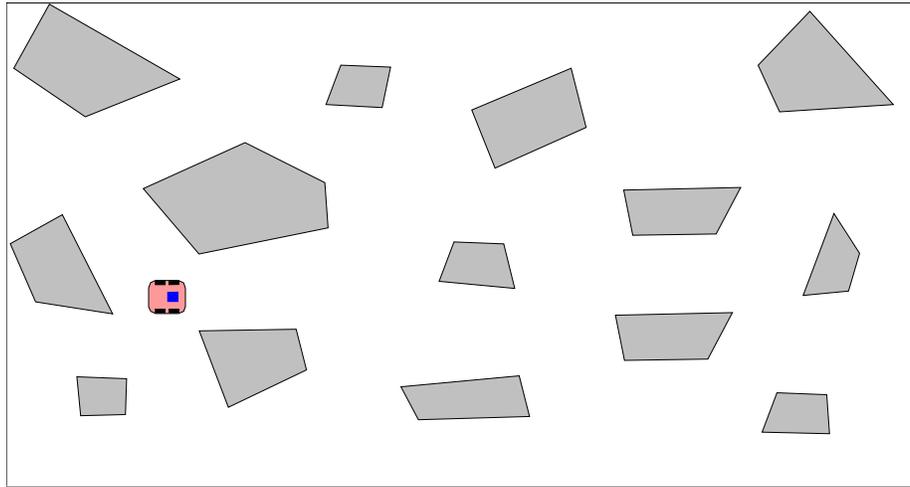


Figura 8.4: Mundo atestado de obstáculos. Si se tiene en cuenta el tamaño del robot se verá que hay pasos realmente estrechos, y aristas peligrosas.

Los resultados que se ofrecen a continuación son cualitativos, a partir de la observación del comportamiento del robot en el simulador frente a las situaciones descritas en las figuras 8.2 y 8.3 y otras en la simulación del restaurante. También se diseñó un mundo con atestado de obstáculos, ver figura 8.4, para ver la capacidad del algoritmo de maniobrar en distancias cortas, así se vió la necesidad de separar los puntos de división de los obstáculos como se comenta en la sección 8.1.

En el restaurante el robot superó un muchas ocasiones las situaciones comentadas sin chocar ni iniciar la maniobra de escape, sino dirigiéndose hacia el objetivo. Los tiempos que el robot tardaba en alcanzar el objetivo se reducían, debido a que se realizaban muchas menos maniobras de escape, que vuelven al robot en la dirección opuesta a la que llevaba. Los puntos de división al estar situados en características relevantes del entorno producían una mejor definición del sector de avance que dirigía al robot de una forma más suave y eficiente para alcanzar el objetivo.

Al observar el mundo atestado de la figura 8.4 conviene comparar el tamaño del robot con el de los pasillos entre obstáculos y se comprenderá que tiene una distribución exigente para un algoritmo de navegación. Los resultados observados al comparar el método de centro de áreas básico con el mejorado muestran, para este último, un incremento notable en la habilidad para sortear distribuciones complicadas de obstáculos, realmente con el método básico el robot caminaba poco entre maniobras de escape consecutivas.

Capítulo 9

Conclusiones y trabajos futuros

9.1. Conclusiones y aportaciones

En el trabajo de Peña (1985) se formula un filtro de Kalman equivalente al procedimiento de Mínimos Cuadrados Lineales Ordinarios, varianza del error constante, y se da un test de detección de outliers. En esta tesis se ha usado esa equivalencia para extender el filtro al procedimiento de Mínimos Cuadrados Lineales Generalizados, varianza del error no constante, donde se descubrió la necesidad de utilizar un factor para la normalización de la ganancia del filtro para su adecuado rendimiento. A continuación el procedimiento se extendió a modelos no lineales, mediante un Filtro de Kalman Extendido, para usar la ecuación de la recta en forma polar. Se han realizado experimentos muy exhaustivos con un modelo de ruido en el que la desviación típica del ruido es proporcional a la distancia medida esperada, con constante de proporcionalidad desconocida. Estos experimentos muestran que aún bajo niveles de ruido altos, el procedimiento proporciona resultados de precisión comparable a los obtenidos mediante el conocido método de Levenberg-Marquardt, que usa todos los datos a la vez por contra del aquí propuesto que es secuencial.

El procedimiento de clustering conocido como scale-space ha sido adaptado al problema de obtención de segmentos en coordenadas polares. Se ha diseñado un filtro para la eliminación del ruido en el proceso de clustering. Este filtro está basado en una caracterización previa de los efectos del ruido en las curvas de nivel obtenidas por el procedimiento de scale-space, la caracterización se obtuvo mediante el estudio estadístico de patrones de alteraciones fijas de diversas intensidades. Esta caracterización se refina, adapta, mediante un procedimiento de filtrado estadístico. El proceso de filtrado tiene un buen rendimiento, eliminando mucha de la fragmentación de características del entorno ocasionada por el ruido.

A partir del test de detección de outliers, se ha dado un criterio para la detección de extremos de segmentos en la estimación secuencial mediante el filtro de Kalman. También se ha establecido un criterio para resolver posibles solapes entre segmentos y se ha usado la equivalencia entre la teoría de la regresión clásica y el filtro utilizado para establecer un criterio de mezcla de segmentos semejantes. Con estas herramientas se ha diseñado un

método de segmentación de medidas de rango. Se han realizado experimentos estadísticos con diversos niveles de ruido, en diversos mundos simulados, usando dos resoluciones angulares para el sensor. Cuando el ruido era bajo, el método de segmentación propuesto ha ofrecido resultados similares a los del popular Split and Merge, superándolo en la longitud total de segmentos clasificada. A niveles de ruido alto el rendimiento del método propuesto es netamente superior, en especial es notable el bajo nivel de fragmentación que produce, el cual proporciona representaciones del mundo consistentes a simple vista, además de estadísticamente aceptables. Estas pruebas también han servido para mostrar que no parece que existan parámetros en el modelo de efecto crítico que el usuario deba ajustar, dentro de los rangos de valores recomendados para los parámetros se observan comportamientos similares del método.

La implementación desarrollada en C++ tiene tiempos de proceso equivalentes a una pequeña fracción del tiempo de muestreo del sensor o del ciclo del robot, lo que hace posible su uso en aplicaciones reactivas o de cualquier otra índole.

El método navegación reactiva basado en el centro de áreas ofrece una potente alternativa a los métodos habituales de navegación reactiva. El procedimiento de división ante obstáculos propuesto es la clave del rendimiento del método en cuanto a evitación, la idea, a pesar de su gran simplicidad, es notablemente efectiva. La fórmula obtenida para las velocidades del robot en función de la posición del centro de áreas, el control, cumple bastante aproximadamente con la idea de seguir la trayectoria del centro de áreas, lo que constituye otro importante factor para explicar los buenos resultados obtenidos. El robot cuando simplemente deambula, lo hace evitando obstáculos, pasando por pasos estrechos y visitando toda la extensión del mundo, describiendo trayectorias sorprendentemente suaves.

El modelado del área libre propuesto, mediante el uso de obstáculos virtuales para elicitar procesos de división en la dirección del objetivo, ha probado ser de gran efectividad para alcanzar dicho objetivo. El método no sufre, o lo hace con bastante menor intensidad, los problemas que aquejan a algunos otros métodos. Experimentos estadísticos en mundos simulados no estructurados, muestran que el objetivo se alcanza recorriendo una distancia un poco superior a la óptima, aunque no hay ninguna planificación previa de la ruta. Experimentos realizados con robots reales muestran resultados concordantes con los obtenidos en simuladores.

En cuanto a velocidad de ejecución, la complejidad del método es $O(n)$, siendo n el número de mediciones, y el cálculo conlleva muy pocas operaciones, lo cual lo hace extremadamente rápido.

Al usar la segmentación para modelar el área libre percibida por el robot, se ha incrementado notablemente la capacidad del método del centro de áreas para conducir la robot de forma segura en entornos complejos. La adecuada situación de los puntos de división del área libre percibida y la posibilidad de que éstos se muevan con el robot, proporciona zonas más convexas, y por lo tanto más seguras, para la determinación del centro de áreas.

9.2. Trabajos futuros

Una de las primeras extensiones del filtro de Kalman diseñado es su aplicación a curvas, sobre el particular hay una tesis de máster en curso. La segunda es la extensión a 3D, para esta cuestión hay dos caminos: trabajar con rodajas 2D e integrar la información o trabajar directamente con planos, la teoría expuesta no limita el número de parámetros del modelo, así que teóricamente es posible la extensión a 3D.

El sólido fundamento teórico en el que descansa el filtro diseñado le hace un interesante candidato para ser usado en aplicaciones de creación de mapas y localización.

Para el procedimiento de clustering se pueden pensar las mismas extensiones que para el filtro de Kalman. No parece que deba haber problemas con ampliar el espectro de tipos de líneas en 2D, pero la extensión a 3D no es directa, ya que en la teoría de scale-space para señales 2D, no se cumplen todas las propiedades que se cumplen en señales 1D. Queda también abierta la cuestión del cálculo de los núcleos gaussianos para resoluciones altas de los sensores, y que es necesario aumentar el máximo del parámetro de escala.

En cuanto al centro de áreas las posibilidades de investigación son muy amplias. Se puede avanzar en el modelado de área libre con propósito de navegación. Se ha modelado el área libre para alcanzar un objetivo introduciendo objetos virtuales puntuales. Pero se podrían usar objetos virtuales de otras formas, por ejemplo, aprovechando que se dispone de segmentación, un segmento. Veamos cómo se puede inducir un comportamiento de seguimiento de paredes. Se elige un lado del robot y para los segmentos percibidos a ese lado se van creando segmentos virtuales paralelos al otro lado del robot. Si esos segmentos se usan para limitar el área libre, el robot, al seguir el centro de áreas, irá paralelo a la pared. Una aplicación mucho más interesante puede ser la siguiente. Si con un sistema de visión es posible extraer los bordes de una carretera, se podría convertirlos en obstáculos virtuales que limiten el área libre. De esta manera el robot, al seguir el centro de áreas, iría por el centro de la carretera evitando los obstáculos que se encontrara. Dando un mayor peso al segmento del lado derecho en el cálculo del centro de áreas, se podría conseguir que el robot circulara por el carril derecho aunque no existiera línea mediana.

La extensión a 3D del método del centro de áreas es otra interesante posibilidad, tanto si el robot posee sensores de rango en 3D y capacidad de movimiento en 3D, como si es un vehículo terrestre equipado con sensores de rango 2D y sensores inerciales con los que fusionar la información de rango.

Bibliografía

- Abdessemed, F., Benmahammed, K., and Monacelli, E. (2004). A fuzzy-based reactive controller for a non-holonomic mobile robot. *Robotics and Autonomous Systems*, 47(1):31–46.
- Alempijevic, A. and Dissanayake, G. (2004). An efficient algorithm for line extraction from laser scans. In *Robotics, Automation and Mechatronics, 2004 IEEE Conference on*, volume 2, pages 970–974 vol.2.
- Álvarez, J. R., de la Paz, F., and Mira, J. (1999). On Virtual Sensory Coding: An Analytical Model of the Endogenous Representation. In Mira, J. and Sánchez-Andrés, J., editors, *Engineering Applications of Bio-Inspired Artificial Neural Networks*, volume 1607 of *LNCS*, pages 526–539. Springer-Verlag.
- Álvarez Sánchez, J., de la Paz López, F., Cuadra Troncoso, J., and de Santos Sierra, D. (2010). Reactive navigation in real environments using partial center of area method. *Robotics and Autonomous Systems*, 58(12):1231–1237.
- Álvarez Sánchez, J. R., de la Paz López, F., Cuadra Troncoso, J. M., and Rosado Sánchez, J. I. (2009). Partial Center of Area Method Used for Reactive Autonomous Robot Navigation. In Mira, J., Ferrández, J. M., Álvarez, J. R., de la Paz, F., and Toledo, F. J., editors, *Bioinspired Applications in Artificial and Natural Computation*, volume 5602 of *LNCS*, pages 408–418. Springer Verlag.
- Anshar, M. and Williams, M.-A. (2007). Extended evolutionary fast learn-to-walk approach for four-legged robots. *Journal of Bionic Engineering*, 4(4):255–263.
- Arkin, R. C. (1998a). *Behaviour-Based Robotics*. MIT Press.
- Arkin, R. C. (1998b). The handbook of brain theory and neural networks. chapter Reactive robotic systems, pages 793–796. MIT Press, Cambridge, MA, USA.
- Arras, K. O. and Siegwart, R. Y. (1997). Feature extraction and scene interpretation for map-based navigation and map building. In *Proc. of SPIE, Mobile Robotics XII*, pages 42–53.

- Baltzakis, H. and Trahanias, P. (2002). An iterative approach for building feature maps in cyclic environments. In *In Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 576–581.
- Barfoot, T., Earon, E., and D’Eleuterio, G. (2006). Experiments in learning distributed control for a hexapod robot. *Robotics and Autonomous Systems*, 54(10):864–872.
- Beal, S. L. and Sheiner, L. B. (1988). Heteroscedastic nonlinear regression. *Technometrics*, 30(3):pp. 327–338.
- Benbrahim, H. and Franklin, J. A. (1997). Biped dynamic walking using reinforcement learning. *Robotics and Autonomous Systems*, 22(3-4):283–302. Robot Learning: The New Wave.
- Borenstein, J. and Koren, Y. (1991). The vector field histogram-fast obstacle avoidance for mobile robots. *Robotics and Automation, IEEE Transactions on*, 7(3):278–288.
- Borges, G. A. and Aldon, M.-J. (2004). Line extraction in 2d range images for mobile robotics. *J. Intell. Robotics Syst.*, 40:267–297.
- Braitenberg, V. (1986). *Vehicles: experiments in synthetic psychology*. MIT.
- Braunstingl, R., Mujika, J., and Uribe, J. (1995). A wall following robot with a fuzzy logic controller optimized by a genetic algorithm. In *Fuzzy Systems, 1995. International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and The Second International Fuzzy Engineering Symposium., Proceedings of 1995 IEEE International Conference on*, volume 5, pages 77–82 vol.5.
- Brooks, R. A. (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2(1):14–23.
- Castellano, G., Attolico, G., and Distanto, A. (1997). Automatic generation of fuzzy rules for reactive robot controllers. *Robotics and Autonomous Systems*, 22(2):133–149.
- Castellanos, J. A. and Tardós, J. D. (1996). *Laser-based segmentation and localization for a mobile robot*, volume 6 of *Robotics and manufacturing: Recent trends in research and applications*. ASME Press, New York.
- Clark, R., Arkin, R. C., and Ram, A. (1992). Learning momentum: On-line performance enhancement for reactive systems. In *Proceedings of the 1992 IEEE International Conference on Robotics and Automation*, pages 111–116. IEEE.
- Cox, D. (1984). Interaction. *International Statistical Review*, 52(1):1–31.
- Cuadra Troncoso, J. (2008). Simulación realista de sensores de rango: un enfoque probabilístico. Technical Report R-01, Dept. Inteligencia Artificial-UNED. <http://www.ia.uned.es/personal/jmcuadra/techreports/simulprobabil-TR-R01.pdf>.

- Cuadra Troncoso, J. M. (2011). Cybersim. <http://www.ia.uned.es/personal/jmcuadra/cybersim/cybersim-qt4/cybersim.tar.gz>.
- Cuadra Troncoso, J. M., Álvarez Sánchez, J. R., and de la Paz, F. (2007). Discretization of iso-learning and ico-learning to be included into reactive neural networks for a robotics simulator. In *IWINAC (2)*, pages 367–378.
- Cuadra Troncoso, J. M., Álvarez Sánchez, J. R., and de la Paz, F. (2009). Discretized iso-learning neural network for obstacle avoidance in reactive robot controllers. *Neuro-computing*, 72(4-6):861–870.
- Cuadra Troncoso, J. M., Álvarez Sánchez, J. R., de la Paz López, F., and Fernández-Caballero, A. (2011). Improving area center robot navigation using a novel range scan segmentation method. In Ferrández, J. M., Álvarez, J. R., de la Paz, F., and Toledo, F. J., editors, *Bioinspired Applications in Artificial and Natural Computation*, volume 6686 of *LNCS*. Springer Verlag. To be published.
- Cuadra Troncoso, J. M. (2011). qscale-space. http://www.ia.uned.es/personal/jmcuadra/tesis-software/qscale_space.tar.gz.
- Davesne, F. and Barret, C. (1999). Reactive navigation of a mobile robot using a hierarchical set of learning agents. In *Intelligent Robots and Systems, 1999. IROS '99. Proceedings. 1999 IEEE/RSJ International Conference on*.
- de la Paz, F., Álvarez, J. R., and Mira, J. (2003). An Analytical Method for Decomposing the External Environment Representation Task for a Robot with Restricted Sensory Information. In Zhou, C., Maravall, D., and Ruan, D., editors, *Autonomous Robotic Systems: Soft Computing and Hard Computing Methodologies and Applications*, volume 116 of *Studies in Fuzziness and Soft Computing*, pages 189–215. Springer Physica-Verlag.
- de la Paz López, F. (2003). *Una arquitectura que integra el modelado endógeno del medio y la navegación para un robot genérico de dos ruedas*. PhD thesis, ETSI Informática-UNED-Spain.
- de la Paz López, F. and Álvarez Sánchez, J. R. (2001). Topological Maps for Robot's Navigation: A Conceptual Approach. In Mira, J. and Prieto, A., editors, *Bio-Inspired Applications of Connectionism*, volume 2085 of *Lecture Notes in Computer Science*, pages 459–467. Springer-Verlag.
- de la Paz López, F., Álvarez Sánchez, J. R., Rosado Sánchez, J. I., and Cuadra Troncoso, J. M. (2009). Mathematical Foundations of the Center of Area Method for Robot Navigation. In Mira, J., Ferrández, J. M., Álvarez, J. R., de la Paz, F., and Toledo, F. J., editors, *Bioinspired Applications in Artificial and Natural Computation*, volume 5602 of *LNCS*, pages 419–428. Springer Verlag.

- de Santos, D., Lorente, V., de la Paz, F., Troncoso, J. M. C., Sánchez, J. R. Á., Fernández, E., and Ferrández, J. M. (2010). A client-server architecture for remotely controlling a robot using a closed-loop system with a biological neuroprocessor. *Robotics and Autonomous Systems*, 58(12):1223–1230.
- del R. Millán, J. and Torras, C. (1995). Efficient reinforcement learning of navigation strategies in an autonomous robot. In Graefe, V., editor, *Intelligent Robots and Systems*, pages 185–199. Elsevier Science B.V., Amsterdam.
- Demirli, K. and Khoshnejad, M. (2009). Autonomous parallel parking of a car-like mobile robot by a neuro-fuzzy sensor-based controller. *Fuzzy Sets and Systems*, 160(19):2876–2891. Theme: Control.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, 39(1):1–38.
- Dietmayer, K. C. J., Sparbert, J., and Streller, D. (2001). Model based object classification and object tracking in traffic scenes from range images. In *IV IEEE Intelligent Vehicles Symposium*, pages 25–30. IEEE Intelligent Vehicles Symposium.
- Dubrawski, A. and Crowley, J. L. (1994). Learning locomotion reflexes: A self-supervised neural system for a mobile robot. *Robotics and Autonomous Systems*, 12(3-4):133–142.
- Duda, R. O. and Hart, P. E. (1972). Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM*, 15:11–15.
- Duda, R. O. and Hart, P. E. (1973). *Pattern Classification and Scene Analysis*. Wiley.
- Erden, M. S. and Leblebicioglu, K. (2008). Free gait generation with reinforcement learning for a six-legged robot. *Robotics and Autonomous Systems*, 56(3):199–212.
- Fahimi, F., Nataraj, C., and Ashrafiuon, H. (2009). Real-time obstacle avoidance for multiple mobile robots. *Robotica*, 27:189–198.
- Fernández, C., Moreno, V., Curto, B., and Vicente, J. A. (2010). Clustering and line detection in laser range measurements. *Robotics and Autonomous Systems*, 58(5):720–726.
- Fernandez-Leon, J. A., Acosta, G. G., and Mayosky, M. A. (2009). Behavioral control through evolutionary neurocontrollers for autonomous mobile robot navigation. *Robotics and Autonomous Systems*, 57(4):411–419.
- Ferrández, J. M., Lorente, V., Cuadra, J. M., de la Paz, F., Sánchez, J. R. Á., and Fernández, E. (2010). A hybrid robotic control system using neuroblastoma cultures. In *HAIS (1)*, pages 245–253.

- Ferrara, A. and Rubagotti, M. (2007). Sliding mode control of a mobile robot for dynamic obstacle avoidance based on a time-varying harmonic potential field. In *ICRA 2007 Workshop: Planning, Perception and Navigation for Intelligent Vehicles*.
- Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24:381–395.
- Floreano, D. and Mondada, F. (1998). Evolutionary neurocontrollers for autonomous mobile robots. *Neural Networks*, 11(7-8):1461–1478.
- Fontanelli, D., Ricciato, L., and Soatto, S. (2007). A fast ransac-based registration algorithm for accurate localization in unknown environments using lidar measurements. In *Automation Science and Engineering, 2007. CASE 2007. IEEE International Conference on*, pages 597–602.
- Forsberg, J., Larsson, U., and Wernersson, A. (1995). Mobile robot navigation using the range-weighted hough transform. *Robotics Automation Magazine, IEEE*, 2(1):18–26.
- Forsyth, D. A. and Ponce, J. (2003). *Computer Vision: A Modern Approach*. Prentice Hall.
- Fox, D., Burgard, W., and Thrun, S. (1997). The dynamic window approach to collision avoidance. *Robotics Automation Magazine, IEEE*, 4(1):23–33.
- Gallant, A. R. (1975). Testing a subset of the parameters of a nonlinear regression model. *Journal of the American Statistical Association*, 70(352):pp. 927–932.
- Gat, E. (1991). *Reliable goal-directed reactive control for real-world autonomous mobile robots*. PhD thesis, Virginia Polytech. State Univ., Blacksburg.
- Gonçalves, V. M., Pimenta, L. C. A., Maia, C. A., Dutra, B. C. O., and Pereira, G. A. S. (2010). Vector fields for robot navigation along time-varying curves in n-dimensions. *Trans. Rob.*, 26:647–659.
- Goodridge, S. G. and Luo, R. C. (1994). Fuzzy behavior fusion for reactive control of an autonomous mobile robot: Marge. In *ICRA*, pages 1622–1627.
- Han, S.-J. and Oh, S.-Y. (2008). An optimized modular neural network controller based on environment classification and selective sensor usage for mobile robot reactive navigation. *Neural Computing and Applications*, 17(2):161–173.
- Harati, A. and Siegwart, R. (2007). A new approach to segmentation of 2d range scans into linear regions. In *Proc. of The IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

- Hui, N. B., Mahendar, V., and Pratihari, D. K. (2006). Time-optimal, collision-free navigation of a car-like mobile robot using neuro-fuzzy approaches. *Fuzzy Sets and Systems*, 157(16):2171–2204.
- Huq, R., Mann, G. K., and Gosine, R. G. (2008). Mobile robot navigation using motor schema and fuzzy context dependent behavior modulation. *Applied Soft Computing*, 8(1):422–436.
- Jacob, H., Fede, S., and Slotine, J. (1997). Real-time path planning using harmonic potential functions in dynamic environment. In *IEEE International Conference on Robotics Automation*, pages 874–881.
- Johnston, J. and DiNardo's, J. (1996). *Econometric Methods*. McGraw-Hill/Irwin, 4th edition.
- Khatib, O. (1986). Real-time obstacle avoidance for manipulators and mobile robots. *Int. J. Rob. Res.*, 5:90–98.
- Khatib, O., Minguez, J., and Montano, L. (2002). Reactive collision avoidance for navigation with dynamic constraint. In *2002 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems*.
- Kim, J. and Khosla, P. K. (1992). Real-time obstacle avoidance using harmonic potential functions. *IEEE Transactions on Robotics and Automation*, 8(3):338–349.
- Kimura, D. K. (1990). Testing nonlinear regression parameters under heteroscedastic, normally distributed errors. *Biometrics*, 46(3):pp. 697–708.
- Kirchner, F. (1998). Q-learning of complex behaviours on a six-legged walking machine. *Robotics and Autonomous Systems*, 25(3-4):253–262. Autonomous Mobile Robots.
- Koren, Y. and Borenstein, J. (1991). Potential field methods and their inherent limitations for mobile robot navigation. In *Proc. of the IEEE international conference on robotics and automation*, volume 2, pages 1398–1404.
- Krogh, B. (1984). A generalized potential field approach to obstacle avoidance control. Technical report.
- Kröse, B. and van Dam, J. (1992). Learning to avoid collisions: A reinforcement learning paradigm for mobile robot navigation. *Annual Review in Automatic Programming*, 17:317–321. Artificial Intelligence in Real-time Control 1992, Selected Papers from the IFAC/IFIP/IMACS Symposium.
- Krystek, M. and Anton, M. (2007). A weighted total least-squares algorithm for fitting a straight line. *Measurement Science and Technology*, 18(11):3438–3442.

- Larbi, M., Fatiha, H., Nasr-eddine, B., and Benyettou, A. (2005). Implementing temporal radial basis function for reactive navigation of mobile robot. *The international Arab Journal of Inf. Tech.*, 2(4).
- Lindeberg, T. (1990). Scale-space for discrete signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:234–254.
- Lindeberg, T. (1991). Discrete scale-space theory and the scale-space primal sketch.
- Lindeberg, T. (1992). Scale-space for n-dimensional discrete signals. In *Proc. NATO ARW on Shape in Picture, Driebergen*, pages 7–11.
- Lindeberg, T. (1993). Discrete derivative approximations with scale-space properties: A basis for low-level feature extraction. *J. Math. Imaging Vision*, pages 349–376.
- Lindeberg, T. (1997). *On the axiomatic foundations of linear scale-space: Combining semi-group structure with causality vs. scale invariance*, chapter 6, pages 75–98. Gaussian Scale-Space Theory: Proc. PhD School on Scale-Space Theory. Kluwer Academic Publishers/Springer,.
- Lindeberg, T. (2009). *Scale space*, pages 2495–2504. Wiley Encyclopedia of Computer Science and Engineering. John Wiley and Sons.
- Lindeberg, T. and Bretzner, L. (2003). Real-time scale selection in hybrid multi-scale representations. In *Proceedings of the 4th international conference on Scale space methods in computer vision, Scale Space'03*, pages 148–163, Berlin, Heidelberg. Springer-Verlag.
- Maaref, H. and Barret, C. (2000). Sensor-based fuzzy navigation of an autonomous mobile robot in an indoor environment. *Control Engineering Practice*, 8(7):757–768.
- Maes, P. and Brooks, R. (1990). Learning to coordinate behaviors. In *8th AAAI*.
- Mahalanobis, P. C. (1936). On the generalised distance in statistics. In *Proceedings National Institute of Science, India*, volume 2, pages 49–55.
- Maravall, D. and de Lope, J. (2003). Autonomous robotic systems. chapter A bio-inspired robotic mechanism for autonomous locomotion in unconventional environments, pages 263–292. Physica-Verlag GmbH, Heidelberg, Germany, Germany.
- Maravall, D., de Lope Asiaín, J., and Serradilla, F. (2000). Combination of model-based and reactive methods in autonomous navigation. In *ICRA*, pages 2328–2333.
- Markovsky, I. and Huffel, S. V. (2007). Overview of total least-squares methods. *Signal Processing*, 87(10):2283–2302. Special Section: Total Least Squares and Errors-in-Variables Modeling.

- Masoud, S. and Masoud, A. (2000). Constrained motion control using vector potential fields. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 30(3):251–272.
- Mastrogiovanni, F., Sgorbissa, A., and Zaccaria, R. (2005). On the tips of one’s toes: self-localization in a dynamic environment. In *Computational Intelligence in Robotics and Automation, 2005. CIRA 2005. Proceedings. 2005 IEEE International Symposium on*, pages 341–346.
- Mataric, M. J. (1994). Reward functions for accelerated learning. In *In Proceedings of the Eleventh International Conference on Machine Learning*, pages 181–189. Morgan Kaufmann.
- Mataric, M. J. (1997). Behaviour-based control: examples from navigation, learning, and group behaviour. *J. Exp. Theor. Artif. Intell.*, 9(2-3):323–336.
- Matellán, V., Fernández, C., and Molina, J. M. (1998). Genetic learning of fuzzy reactive controllers. *Robotics and Autonomous Systems*, 25(1-2):33–41.
- Mikolajczyk, K. and Schmid, C. (2001). Indexing based on scale invariant interest points. In *In Proc. ICCV*, pages 525–531.
- N. Franceschini, J. (1992). From insect vision to robot vision. *Philosophical Transactions Royal Society of London*, 337(B):283–294.
- Nehmzow, U. and McGonigle, B. (1994). Achieving rapid adaptations in robots by means of external tuition. In *From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior*, pages 301–308. MIT Press.
- Nguyen, V., Gächter, S., Martinelli, A., Tomatis, N., and Siegwart, R. (2007). A comparison of line extraction algorithms using 2d range data for indoor mobile robotics. *Auton. Robots*, 23:97–111.
- Nolfi, S. and Floreano, D. (2004). *Evolutionary robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines*. MIT Press.
- Norouzi, M., Yaghibi, M., Siboni, M., and Jadaliha, M. (2009). Recursive line extraction algorithm from 2d laser scanner applied to navigation a mobile robot. In *Robotics and Biomimetics, 2008. ROBIO 2008. IEEE International Conference on*, pages 2127–2132.
- Núñez, P., Vázquez-Martín, R., del Toro, J., Bandera, A., and Sandoval, F. (2008). Natural landmark extraction for mobile robot navigation based on an adaptive curvature estimation. *Robotics and Autonomous Systems*, 56(3):247–264.
- Ogren, P. and Leonard, N. (2005). A convergent dynamic window approach to obstacle avoidance. *Robotics, IEEE Transactions on*, 21(2):188–195.

- Ortega, J. G. and Camacho, E. F. (1996). Mobile robot navigation in a partially structured static environment, using neural predictive control. *Control Engineering Practice*, 4(12):1669–1679.
- Pavlidis, T. and Horowitz, S. (1974). Segmentation of plane curves. *Computers, IEEE Transactions on*, C-23(8):860–870.
- Peña, D. (1985). Sobre la robustificación interna del algoritmo de Plackett-Kalman para la estimación recursiva del modelo de regresión lineal. *Trabajos de Estadística y de Investigación Operativa*, 36(1):93–106.
- Pearce, M., Arkin, R., and Ram, A. (1992). The learning of reactive control parameters through genetic algorithms. <http://hdl.handle.net/1853/22447>.
- Pfister, S. T., Roumeliotis, S. I., and Burdick, J. W. (2003). Weighted line fitting algorithms for mobile robot map building and efficient data representation. In *In ICRA*, pages 14–19.
- Plackett, R. L. (1950). Some theorems in least squares. *Biometrika*, 37((1-2)):149–157.
- Porr, B., von Ferber, C., and Wörgötter, F. (2003). ISO-learning approximates a solution to the inverse-controller problem in an unsupervised behavioural paradigm. *Neural Computation*, (15):865–884.
- Porr, B. and Wörgötter, F. (2006). Strongly improved stability and faster convergence of temporal sequence learning by utilising input correlations only. *Neural Computation*, 18(6):1380–1412.
- Pratihari, D. K., Deb, K., and Ghosh, A. (1999). A genetic-fuzzy approach for mobile robot navigation among moving obstacles. *International Journal of Approximate Reasoning*, 20(2):145–172.
- Ram, A., Arkin, R., Boone, G., and Pearce, M. (1994). Using genetic algorithms to learn reactive control parameters for autonomous robotic navigation. *Adapt. Behav.*, 2:277–305.
- Ram, A., Arkin, R., Moorman, K., and Clark, R. (1997). Case-based reactive navigation: a method for on-line selection and adaptation of reactive robotic control parameters. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 27(3):376–394.
- Reignier, P. (1994). Fuzzy logic techniques for mobile robot obstacle avoidance. *Robotics and Autonomous Systems*, 12(3-4):143–153.
- Rosin, P. L. (1994). Determining local natural scales of curves. *Pattern Recognition Lett*, 19:63–75.

- Rossomando, F. G., Soria, C., and Carelli, R. (2011). Autonomous mobile robots navigation using rbf neural compensator. *Control Engineering Practice*, 19(3):215–222. Special Section: IFAC World Congress Application Paper Prize Papers.
- Ruini, F. and Cangelosi, A. (2009). Extending the evolutionary robotics approach to flying machines: An application to mav teams. *Neural Networks*, 22(5-6):812–821. Advances in Neural Networks Research: IJCNN2009, 2009 International Joint Conference on Neural Networks.
- Saffiotti, A., Ruspini, E., and Konolige, K. (1993). Blending reactivity and goal-directedness in a fuzzy controller. In *Fuzzy Systems, 1993., Second IEEE International Conference on.*
- Santos, S. et al. (2003). Tracking of multi-obstacles with laser range data for autonomous vehicles. In *Proceeding of ROBOTICA 2003 - Scientific ROBMeeting of the 3rd Robotics National Festival*, pages 59–65.
- Selekwa, M. F., Dunlap, D. D., Shi, D., and Jr., E. G. C. (2008). Robot navigation in very cluttered environments by preference-based fuzzy behaviors. *Robotics and Autonomous Systems*, 56(3):231–246.
- SensComp, Inc. (2004). 600 Series Intrument Transducer Specifications. <http://www.senscomp.com/specs/600%20instrument%20spec.pdf>.
- Serradila, F. and Maravall, D. (1996). A navigation system for mobile robots using visual feedback and artificial potential fields. In *30th European Meeting on Cybernetics and System research*, pages 1159–1164.
- Siadat, A., Kaske, A., Klausmann, S., Dufaut, M., and Husson, R. (1997). An optimized segmentation method for a 2d laser-scanner applied to mobile robot navigation. In *Proceedings of the 3rd IFAC symposium on intelligent components and instruments for control applications*.
- SICK AG (2006). Technical Description LMS200/211/221/291 Laser Measurement Systems. <http://www.mysick.com/saqqara/get.aspx?id=IM0012759>.
- Siegel, A. F. (1982). Robust regression using repeated medians. *Biometrika*, 69(1):pp. 242–244.
- Singhal, A. (1997). Issues in autonomous mobile robot navigation. Technical report.
- Smolic-Rocak, N., Bogdan, S., Kovacic, Z., and Petrovic, T. (2010). Time windows based dynamic routing in multi-agv systems. *Automation Science and Engineering, IEEE Transactions on*, 7(1):151–155.
- Song, K.-T. and Sheen, L.-H. (2000). Heuristic fuzzy-neuro network and its application to reactive navigation of a mobile robot. *Fuzzy Sets and Systems*, 110(3):331–340.

- Stollnitz, E. J., DeRose, T. D., and Salesin, D. H. (1996). *Wavelets for Computer Graphics: Theory and Applications*. Morgan Kaufmann, San Francisco.
- Sutton, R. S. (1988). Learning to predict by the method of temporal differences. *Machine learning*, 3:9–44.
- Sutton, R. S. and Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, 88:135–170.
- Tan, F., Yang, J., Huang, J., Jia, T., Chen, W., and Wang, J. (2010). A navigation system for family indoor monitor mobile robot. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 5978–5983.
- Taylor, R. and Probert, P. (1996). Range finding and feature extraction by segmentation of images for mobile robot navigation. In *Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on*, volume 1, pages 95–100 vol.1.
- Tellinghuisen, J. (2010). Least-squares analysis of data with uncertainty in x and y: A monte carlo methods comparison. *Chemometrics and Intelligent Laboratory Systems*, 103(2):160–169.
- Thrun, S., Burgard, W., and Fox, D. (2005). *Probabilistic Robotics*. MIT Press, Cambridge, MA.
- Tianmiao, W. and Bo, Z. (1992). Time-varying potential field based ‘perception-action’ behaviors of mobile robot. In *Robotics and Automation, 1992. Proceedings., 1992 IEEE International Conference on*, pages 2549–2554 vol.3.
- Ulrich, I. and Borenstein, J. (1998). Vfh+: reliable obstacle avoidance for fast mobile robots. In *Robotics and Automation, 1998. Proceedings. 1998 IEEE International Conference on*, volume 2, pages 1572–1577 vol.2.
- Ulrich, I. and Borenstein, J. (2000). Vfh*: local obstacle avoidance with look-ahead verification. In *Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference on*.
- Vandorpe, J., Van Brussel, H., and Xu, H. (1996). Exact dynamic map building for a mobile robot using geometrical primitives produced by a 2d range finder. In *Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on*, volume 1, pages 901–908 vol.1.
- Verschure, P. F. M. J., Wray, J., Sporns, O., Tononi, G., and Edelman, G. M. (1995). Multilevel analysis of classical conditioning in a behaving real world artifact. *Robotics and Autonomous Systems*, 16(2-4):247–265. Moving the Frontiers between Robotics and Biology.
- Walter, W. G. (1950). Electromechanical animal. *Dialectica*, 4:42–49.

- Witkin, A. (1984). Scale-space filtering: A new approach to multi-scale description. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84.*, volume 9, pages 150–153.
- Xu, W. L., Tso, S. K., and Fung, Y. H. (1998). Fuzzy reactive control of a mobile robot incorporating a real/virtual target switching strategy. *Robotics and Autonomous Systems*, 23(3):171–186.
- Yang, H., Borenstein, J., and Wehe, D. (2000). Sonar-based obstacle avoidance for a large, non-point, omni-directional mobile robot. In *Omni-directional Mobile Robot, Spectrum 2000 International Conference on Nuclear and Hazardous Waste Management*, pages 24–28.
- Yen, G. G. and Hickey, T. W. (2004). Reinforcement learning algorithms for robotic navigation in dynamic environments. *ISA Transactions*, 43(2):217–230.