



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

CARACTERIZACIÓN DE EXOPLANETAS MEDIANTE TÉCNICAS DE
PROCESAMIENTO DE SEÑALES Y MÉTODO DE LAS VELOCIDADES RADIALES

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL ELECTRICISTA

RODRIGO MANUEL MAHÚ SINCLAIR

PROFESOR GUÍA:
NÉSTOR BECERRA YOMA

MIEMBROS DE LA COMISIÓN:
JOSÉ NOVOA ILIC
JORGE WUTH SEPÚLVEDA

SANTIAGO DE CHILE
2016

RESUMEN DE LA MEMORIA
PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL ELECTRICISTA
POR: RODRIGO MANUEL MAHÚ SINCLAIR
FECHA: 2016
PROF. GUÍA: NÉSTOR BECERRA YOMA

CARACTERIZACIÓN DE EXOPLANETAS MEDIANTE TÉCNICAS DE PROCESAMIENTO DE SEÑALES Y MÉTODO DE LAS VELOCIDADES RADIALES

El método de las velocidades radiales permite la detección de planetas en otros sistemas solares, al obtener las características y configuración del sistema planetario. Esto es usualmente realizado por métodos secuenciales, que no garantizan obtener una solución global del problema. Una distorsión en la primera señal detectada, en conjunto con la ventana de muestreo no uniforme pueden causar la aparición de componentes artificiales.

En este trabajo se desarrolló un método para encontrar una solución global del problema. Determinando tanto el número de componentes como las características de estas.

Para esto se implementó dos programas en C, que en conjunto permitirán determinar las características y configuración de sistema. Estos se evalúan tanto en la validez de los resultados que entregan, como de la eficiencia computacional de los mismos.

Para la validación se utilizó señales reales proporcionadas por astrónomos y señales artificiales.

Se muestra como el periodograma basado en MMSE entrega resultados comparables con los del periodograma Lomb Scargle. Entregando además los parámetros de las componentes detectadas.

Se utilizó un análisis de grilla para detectar la configuración del sistema y refinar las soluciones.

Dedicado a mis padres.

Agradecimientos

Quisiera agradecer a mis padres Carolina y Francisco por su cariño y sus enseñanzas a lo largo de mi vida y por su apoyo y paciencia estos últimos años sin los cuales esta tarea hubiese sido imposible.

A mi hermano Javier por su compañía, cariño y apoyo.

A mi polola Paulina por darme la energía y motivación que tanta falta me hacían.

A mis amigos de Club de Kung Chung ly Tang por mantenerme cuerdo.

A mi profesor Néstor Becerra Yoma le agradezco la oportunidad de trabajar en el LPTV y por su guía constante.

Y a mis compañeros del LPTV por toda la ayuda y conocimiento que me han brindado.

Tabla de contenido

1. Introducción	1
1.1. Motivación	1
1.2. Objetivos	2
1.2.1. Objetivo general	2
1.2.2. Objetivos específicos	2
1.3. Estructura de la memoria	2
2. Búsqueda de Exo-Planetas	4
2.1. Exoplanetas	4
2.2. Métodos de búsqueda	5
2.2.1. Introducción	5
2.2.2. Observación directa	5
2.2.3. Tránsito	6
2.2.4. Velocidad radial	8
2.2.4.1. Introducción	8
2.2.4.2. Descripción general	8
2.2.4.3. Descripción física	8
2.2.4.4. Adquisición de datos	12
2.2.4.5. Instrumentos y bases de datos, HARPS y Keck	13
2.2.4.6. Causas del muestreo no uniforme	13

2.2.4.7. Consecuencias del muestreo no uniforme	14
2.2.4.8. Manchas solares	15
2.3. Métodos de análisis de señales	15
2.3.1. Introducción	15
2.3.2. Lomb Scargle	15
2.3.3. Lomb Scargle generalizado	17
2.3.4. Pre-Blanqueado	18
2.3.5. Máxima verosimilitud	18
2.3.6. Análisis Bayesiano	18
2.4. Teorema de descomposición de Wold	19
2.5. MMSE	20
3. Metodología	21
3.1. MMSE sin restricciones	21
3.1.1. Introducción	21
3.1.2. Descripción del problema	21
3.1.3. Complejidad computacional	22
3.2. Método propuesto: periodograma MMSE	23
3.2.1. Introducción	23
3.2.2. Periodograma	23
3.2.3. Selección de candidatos	24
3.2.4. Vecindad	24
3.2.5. Grilla	25
3.2.6. Selección de componentes	25
3.3. Complejidad computacional	26
3.3.1. Introducción	26

3.3.2. Algoritmo periodograma	26
3.3.2.1. Introducción	26
3.3.2.2. Análisis	26
3.3.2.3. Optimización	27
3.3.3. Método de la grilla	27
3.3.3.1. Introducción	27
3.3.3.2. Análisis	27
3.3.3.3. Optimizaciones	28
3.4. Usos adicionales método de la grilla	29
3.4.1. Introducción	29
3.4.2. Refinado de resultados	29
3.4.3. Verificación de fase	29
4. Resultados	31
4.1. Datos	31
4.2. Periodogramas MMSE Vs Lomb Scargle	32
4.3. Experimento de validación para el análisis de grilla	35
4.4. Experimento con señales ruidosas	36
4.5. Pruebas de velocidad	37
4.5.1. Introducción	37
4.5.2. Equipos disponibles	38
4.5.3. Periodograma	38
4.5.3.1. Descripción del experimento	38
4.5.3.2. Resultados	39
4.5.4. Grilla	41
4.5.4.1. Descripción del experimento	41

4.5.4.2. Resultados comparación de códigos	43
4.5.4.3. Resultados procesos paralelos	46
5. Conclusiones	50
5.1. Conclusiones	50
5.2. Trabajo futuro	50
Bibliografía	52

Índice de figuras

2.1. Distribución de masa y distancia para planetas detectados por observación directa comparados con los del sistema solar.	6
2.2. Curva de flujo lumínico del tránsito de un exoplaneta en sus distintas etapas.	7
2.3. Diagrama de la relación entre la velocidad aparente V_A , la velocidad de la estrella V_E , la velocidad del planeta V_P y el ángulo de inclinación i	9
2.4. Descripción gráfica del efecto Doppler en la medición de velocidad para la detección de exoplanetas.	12
3.1. Ejemplo del diagrama de grilla para selección de N componentes con vecindades de largo M	25
3.2. Dependencia del número de errores a calcular N_{ErrGr} con los parámetros del experimento.	28
4.1. Comparación diferentes sets de datos de velocidades radiales normalizadas. En (a) estrella GJ876, en (b) ruido blanco muestreado con el patrón de GJ876, en (c) ruido rojo muestreado con el patrón de GJ876, y en (d) datos sintéticos con muestreo aleatorio.	31
4.2. Periodogramas de la señal sintética, en (a) periodograma Lomb Scargle, y en (b) periodograma MMSE.	33
4.3. Periodogramas del ruido blanco, en (a) periodograma Lomb Scargle, y en (b) periodograma MMSE.	34
4.4. Periodogramas del ruido rojo, en (a) periodograma Lomb Scargle, y en (b) periodograma MMSE.	34

4.5. Periodogramas estrella GJ876, en (a) periodograma Lomb Scargle, y en (b) periodograma MMSE.	35
4.6. Detección de componentes señales con distinto nivel de ruido. Un mínimo error indica el número detectando, siendo tres componentes correctamente detectadas para la señal limpia y ruidos de 18dB, 12dB y 6dB.	37
4.7. Tiempo de cómputo de un periodograma versus el número de tareas realizadas por el computador.	39
4.8. Tiempo de cómputo de un periodograma separado en múltiples procesos dentro de un mismo computador.	40
4.9. Comparación de velocidad de los códigos, para la grilla con mínimo MMSE en la primera combinación (configuración 1).	43
4.10. Comparación de velocidad de los códigos, para la grilla con mínimo MMSE en la última combinación (configuración 2).	44
4.11. Comparación de velocidad de los códigos, para la grilla promedio sobre 4 experimentos.	45
4.12. Tiempo de cómputo de la grilla versus número de procesos al calcular 3 componentes de 4 usando la configuración 2 para 1, 2, 3, 4 y 5 partes. . .	46
4.13. Tiempo de cómputo de la grilla versus número de procesos al calcular 4 componentes de 4 usando la configuración 2 para 1, 2, 3, 4 y 5 partes. . .	47
4.14. Tiempo de cómputo de la grilla versus número de procesos al calcular 3 componentes de 4 usando la configuración 2 para 1, 2, 4, 8, 16 y 32 partes. .	48
4.15. Tiempo de cómputo de la grilla versus número de procesos al calcular 4 componentes de 4 usando la configuración 2 para 1, 2, 4, 8, 16 y 32 partes. .	48

Índice de tablas

3.1. Tiempo teórico de cómputo MMSE sin restricciones.	23
4.1. Componentes de la señal sintética.	32
4.2. Componentes detectadas por el periodograma MMSE en la señal sintética.	33
4.3. Candidatos para validación del algoritmo de la grilla.	36
4.4. Componentes y fases detectadas por la grilla en la señal sintética.	36
4.5. Tiempo de cómputo MMSE periodograma y velocidades aparentes para múltiples equipos.	41
4.6. Orden de los candidatos para las diferentes configuraciones utilizadas en las pruebas de velocidad de la grilla. En negritas la componente no relacionada con la señal.	41
4.7. Orden de exploración de las combinaciones de componentes para el caso de 4 candidatos.	42

Capítulo 1

Introducción

1.1. Motivación

La detección de planetas, orbitando estrellas distintas de nuestro Sol, es de gran interés dentro del estudio actual de la astronomía. Ya que la observación directa de la mayoría estos planetas va más allá de la capacidad actual de los instrumentos astronómicos, se han desarrollado diversas técnicas de observación indirecta. Dichos métodos utilizan la medición de diversas características de las estrellas que son alteradas por la presencia de los cuerpos que las orbitan. La observación de estas propiedades en función del tiempo constituye una señal, que codifica la estructura de dichos sistemas planetarios.

El propósito de esta memoria es utilizar los métodos de procesamiento de señales, normalmente utilizados en los diversos ámbitos de la ingeniería eléctrica, para extraer las características de estos sistemas planetarios. Luego, se desea compararlas con las técnicas actualmente utilizadas en astronomía para validar su utilidad y confiabilidad.

En particular se utilizarán las velocidades radiales de la estrella, las cuales pueden ser medidas a través del efecto Doppler en la luz de la estrella. Esta velocidad tiene directa relación con la velocidad a la que rotan los planetas en torno la estrella. Al haber más de un planeta orbitando la misma estrella su velocidad será la resultante del efecto conjunto de todos los planetas que la orbitan. Es por ello que se debe realizar un análisis para descomponer la señal medida en las señales causadas por cada planeta, para así determinar las características de cada uno de ellos.

Se iniciará con un análisis de tipo Fourier para identificar los periodos de rotación de los planetas. Mientras se busca solución para las diversas dificultades propias del

problema, siendo la principal el muestreo no uniforme.

Esta memoria se enmarca dentro de un proyecto para aplicar métodos de procesamientos de señales en diversos ámbitos, incluyendo pero no limitando a la astronomía, minería y vulcanología. Este abarca en la actualidad el trabajo colaborativo entre la Universidad de Chile, Universidad de Santiago de Chile y la Universidad de la Frontera.

Por su naturaleza interdisciplinaria, este tema permite desarrollar el trabajo colaborativo que es la intención del proyecto. Así, se les brindará nuevas herramientas a los astrónomos del proyecto que les permitirán incrementar el alcance de sus investigaciones.

1.2. Objetivos

1.2.1. Objetivo general

Aplicar métodos de procesamiento de señales a la búsqueda y caracterización de exoplanetas.

1.2.2. Objetivos específicos

- Desarrollar un método para identificar número de planetas y sus características.
- Desarrollar un método para refinar la caracterización de los planetas.
- Desarrollar un método para distinguir señales planetarias de componentes periódicas no planetarias.

1.3. Estructura de la memoria

El trabajo aquí contenido se divide en seis partes siendo esta la primera, en la cual se introduce brevemente el tema a tratar y los objetivos.

En el capítulo 2 se entrega una descripción que es un exoplaneta y las razones para estudiarlos. Luego se describe los métodos de detección y sus alcances. A continuación

se muestra el estado del arte de las técnicas para procesar las velocidades radiales. Finalmente, se describe las herramientas matemáticas en las que se fundamenta el método propuesto.

En el capítulo 3 se describe el método propuesto. Luego, se discute la implementación del método en los programas. Se puso énfasis en las técnicas utilizadas para optimizar la velocidad de cómputo.

En el capítulo 4 se analiza los resultados sobre base de datos ficticia y datos reales que permiten demostrar la validez del método propuesto. Luego, se discute las mejoras realizadas a los códigos para mejorar la velocidad de cómputo.

En el capítulo 5 se presentan las conclusiones del trabajo. Se discute si el método cumple las expectativas y posibles formas de mejorar en trabajos futuros.

Capítulo 2

Búsqueda de Exo-Planetas

2.1. Exoplanetas

Un exoplaneta es, por definición, un planeta que orbita una estrella distinta de nuestro Sol. Su estudio es importante para el entendimiento de la formación de los sistemas planetarios y para la búsqueda de planetas habitables entre otros temas de interés científico.

El incremento de la precisión de los instrumentos astronómicos en las últimas décadas ha permitido la búsqueda de estos planetas. Actualmente existe un gran número de cuerpos detectados y candidatos que aún requieren mayor estudio para ser verificados. La “Exoplanet Orbit Database” [1] [2] identifica 1593 planetas confirmados y 3751 posibles candidatos.

Lamentablemente, la observación directa de estos aún no es posible en la mayoría de los casos. Lo que se comprueba al observar que son solo 8 los casos dentro de la lista antes mencionada. Esto se debe a que son significativamente menores a las estrellas que orbitan y no emiten luz propia.

Por estas razones se han desarrollado múltiples métodos de observación indirecta, entre los cuales destacan los del tránsito y las velocidades radiales. Estos métodos permiten la detección en distintas circunstancias. Además, distintos métodos permiten obtener diferentes características, haciéndolos complementarios.

2.2. Métodos de búsqueda

2.2.1. Introducción

A continuación se examina los diversos métodos de búsqueda de exoplanetas explicando en qué condiciones es posible de utilizarlas y que información brindan.

En la sección 2.2.2 se muestra como el método de observación directa está limitado a la observación de planetas masivos lejos de su estrella y como se comparan los planetas detectados con este método con los del sistema solar.

En la sección 2.2.3 se describe en que consiste el tránsito de un planeta y como puede utilizarse para la detección de exoplanetas. Además se explica como múltiples tránsitos de un mismo planeta pueden permitir la detección de planetas secundarios.

Finalmente la sección 2.2.4 describe en detalle el método de las velocidades radiales. Aquí se muestra el modelo en el que se basa esta técnica, como se adquieren los datos, las características de las bases de datos y los problemas que dificultan su uso.

2.2.2. Observación directa

La observación directa de un planeta con las técnicas actuales solo es posible para planetas de gran tamaño y órbita amplia. Ya que un planeta no emite su propia luz y solo refleja la de su estrella, estos son significativamente menos brillantes que éstas. Así, en la mayoría de las observaciones la estrella enmascara la presencia del planeta.

Sin embargo, si el planeta tiene un tamaño considerable y se encuentra a gran distancia de su estrella, es posible detectarlo como un objeto distinto. Ejemplo de un objeto promedio detectado por este método es el de HD95086 b, un planeta de 4 a 5 masas de Júpiter a 52UA de su estrella [3]. El objeto más cercano a su estrella, detectado por este método es β Pictoris b a 8.8UA, con un tamaño de 9 masas de Júpiter [4]. Finalmente el objeto más pequeño, con la mitad de la masa de Júpiter, Fomalhaut b se encuentra a 113UA de su estrella [5].

En contraste, en el sistema solar, el planeta más grande es Júpiter a 5UA del Sol, y el más lejano es Neptuno a 30UA con aproximadamente 0.05 masas de Júpiter, ambos fuera del alcance de este método.

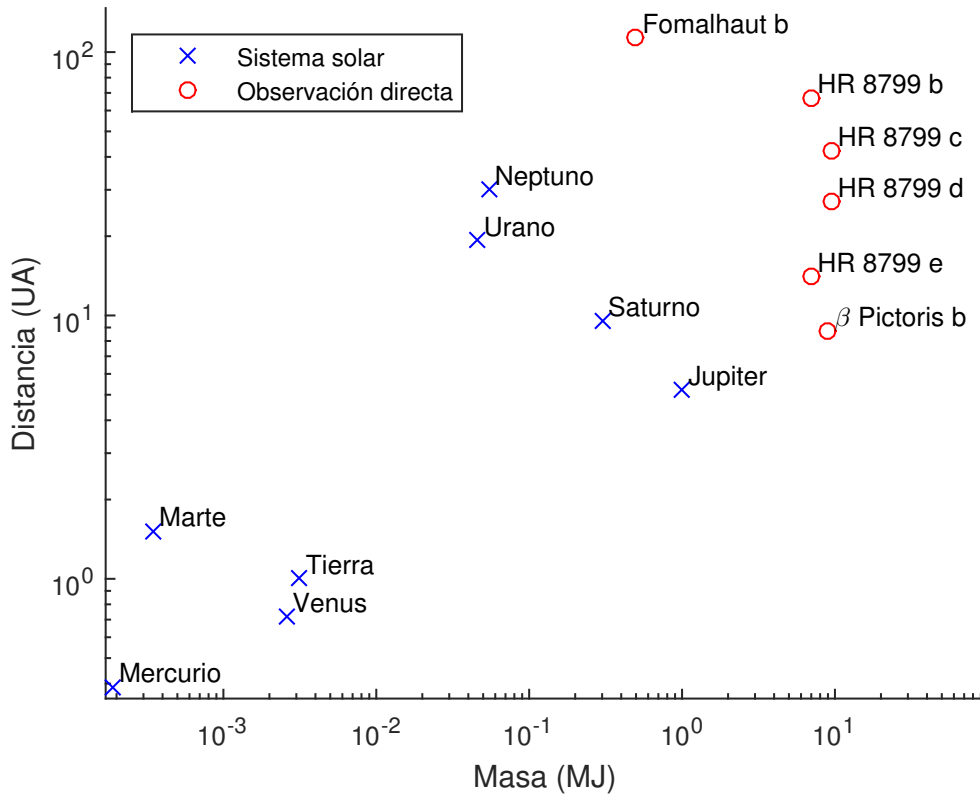


Figura 2.1: Distribución de masa y distancia para planetas detectados por observación directa comparados con los del sistema solar.

En la Figura 2.1 se observa que la distribución de tamaño y distancia de los planetas detectados por observación directa está claramente separada de la distribución de los del sistema solar. Es necesario notar que la figura muestra 6 de los 8 planetas antes mencionados ya que los dos últimos no tienen una estimación de su distancia.

2.2.3. Tránsito

Los tránsitos son posibles en planetas cuyas órbitas pasan entre su estrella y La Tierra, eclipsando parcialmente su estrella. Midiendo la duración de los tránsitos, el intervalo entre ellos y la atenuación de la luminosidad de su estrella se puede caracterizar el planeta. La variación del flujo lumínico en la estrella durante el tránsito permite estimar el radio del planeta, su radio orbital, inclinación y periodo [6]. Lamentablemente los tránsitos solo son posibles en un pequeño número de planetas, y la probabilidad de que un planeta sea detectable por su tránsito decrece con su radio orbital [7]. Probabilidad que se puede aproximar por la ecuación 2.1 para el caso de orbitas circulares.

$$P \approx \frac{R_E}{r} \quad (2.1)$$

Es importante notar que las características que hacen detectables un planeta por este método son precisamente las que hace más difícil la observación directa y aún no ha sido posible detectar un planeta por ambos métodos.

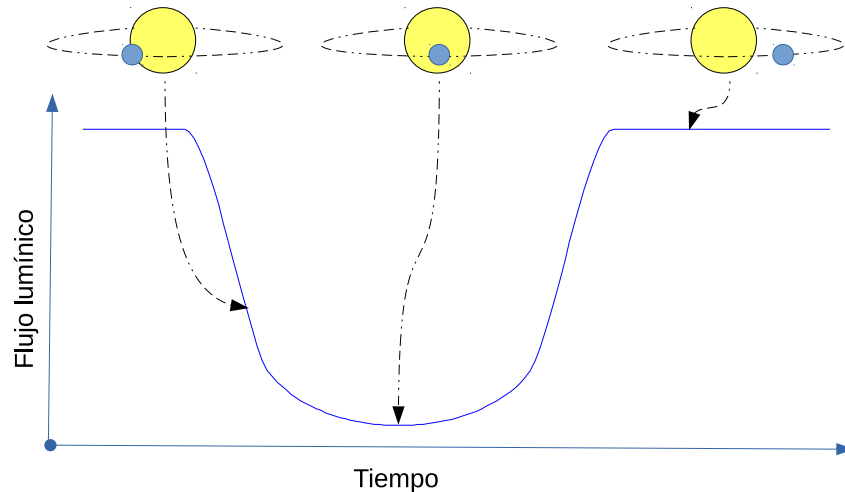


Figura 2.2: Curva de flujo lumínico del tránsito de un exoplaneta en sus distintas etapas.

En la Figura 2.2 se observan las etapas de un tránsito y cómo estas alteran el flujo lumínico de la estrella.

Este método tiene la característica adicional de permitir el análisis de la atmósfera de los planetas detectados por este método. Para esto se analiza el tránsito para diferentes longitudes de onda de luz. Ya que distintos componentes químicos de las atmósferas reaccionan con diferentes longitudes de onda, se detectará un mayor radio planetario cuando la luz sea absorbida que cuando traspase la atmósfera. Así, el patrón de radio planetario en función de la longitud de onda entrega una huella de la composición química de la atmósfera. Ejemplo de esto es la discusión sobre la presencia de TiO en la atmosfera de Wasp 19b [8].

Adicionalmente es posible la detección de planetas secundarios, midiendo el tiempo de ocurrencia de los tránsitos. Ya que la presencia de planetas adicionales causará que su influencia gravitacional los frene y acelere en distintos momentos. Así los tránsitos ya no ocurrirán de manera perfectamente periódica, indicando la presencia de planetas adicionales [9].

2.2.4. Velocidad radial

2.2.4.1. Introducción

En la sección 2.2.4.2 se muestra las características principales del método de las velocidades radiales para luego abordar los detalles en las siguientes secciones.

En la sección 2.2.4.3 se analiza como las características de la señal se relacionan con las del sistema planetario. Para esto se mencionaran las leyes físicas que gobiernan el sistema y los supuestos que permiten establecer un modelo.

La sección 2.2.4.4 describe como se utiliza el efecto Doppler para medir la velocidad radial de la estrella.

Los instrumentos utilizados para la toma de datos son descritos en la sección 2.2.4.5.

En la sección 2.2.4.6 se muestra las causas de que la señal este muestreada no uniformemente mientras las consecuencias de esto se analiza en la sección 2.2.4.7.

La sesión 2.2.4.8 describe como una mancha solar generara una señal de similares características a la de un planeta, pero por un tiempo acotado.

2.2.4.2. Descripción general

El método de velocidades radiales mide las variaciones en el movimiento de las estrellas, el cual es causado por el efecto gravitacional de otros cuerpos. Cada planeta crea un efecto en función de su masa y órbita. Así, la configuración de un sistema planetario queda codificada en la velocidad de su estrella. Luego, es posible utilizar análisis de tipo Fourier para decodificar las características de cada planeta.

Este procedimiento permite detectar simultáneamente múltiples componentes de un mismo sistema, a través de una única señal. De esta forma, el método es limitado por la precisión de las mediciones y la longitud de las series de tiempo. En años recientes ambas han alcanzado niveles suficientes para permitir el descubrimiento de múltiples planetas.

2.2.4.3. Descripción física

Se analiza primero el caso de un único planeta orbitando una estrella. Se tiene que siendo una buena aproximación considerar que el planeta orbita a la estrella, la realidad es que ambos orbitan el centro de masa común. De esta forma, al observar la estrella, se

observa su movimiento en respuesta a la presencia del planeta.

Al observar la estrella se puede determinar la velocidad con la que se acerca y aleja en función del tiempo. Utilizando en una primera aproximación que el planeta tiene una órbita circular, se tendrá que la velocidad observada $V(t)$ será una señal sinusoidal, la cual tendrá magnitud V_A y periodo T , que corresponden a la magnitud de la velocidad aparente de la estrella y el periodo al periodo orbital del planeta respectivamente, como se muestra en la ecuación 2.2. La masa de la estrella M_E se supone conocida, y su estimación escapa del alcance de este trabajo.

$$V(t) = V_A \cos \frac{2\pi t}{T} \quad (2.2)$$

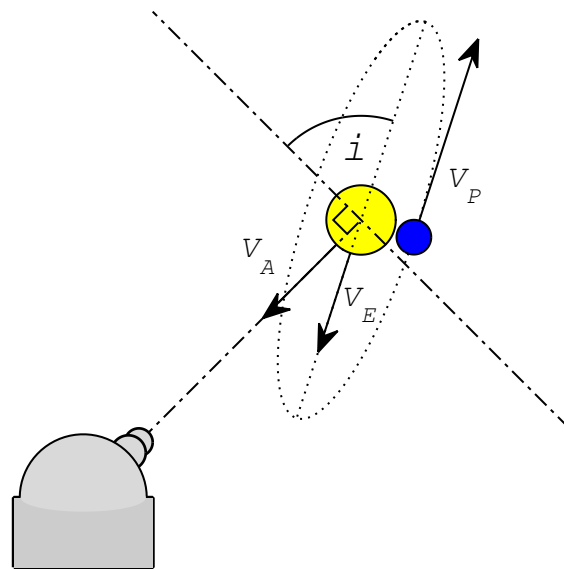


Figura 2.3: Diagrama de la relación entre la velocidad aparente V_A , la velocidad de la estrella V_E , la velocidad del planeta V_P y el ángulo de inclinación i .

En la Figura 2.3 muestra la relación entre las velocidades del sistema. La velocidad aparente de la estrella V_A corresponde a la proyección de la velocidad de la estrella V_E en el eje entre la estrella y La Tierra. El ángulo i corresponde al ángulo entre el plano de traslación del planeta y el plano perpendicular al eje entre La Tierra y la estrella. Luego

ambas velocidades se relacionan según la ecuación 2.3.

$$V_A = V_E \sin i \quad (2.3)$$

La Figura 2.3 también muestra la velocidad del planeta V_P , de dirección opuesta a V_E . Considerando ambas como componentes internas del sistema, es decir eliminando la componente de traslación del sistema, se tiene que el momento total del sistema es 0. Así V_E y V_P se relacionan con la masa de la estrella M_E y la masa del planeta M_P de acuerdo a la ecuación 2.4.

$$V_P M_P = V_E M_E \quad (2.4)$$

Considerando que la masa de la estrella es mucho mayor que la del planeta, se puede aproximar que este gira en torno a la estrella fija con un radio r constante. Utilizando la ecuación 2.5, que corresponde a igualar la fuerza de atracción gravitacional, donde G es la constante de gravitación universal, con la fuerza centrífuga, y la ecuación 2.6 para la velocidad de rotación del planeta, obtenemos expresiones para r en la ecuación 2.7, que corresponde a la tercera ley de Kepler para una órbita circular, y la velocidad del planeta en la ecuación 2.8.

$$\frac{GM_E M_P}{r^2} = M_P \frac{V_P^2}{r} \quad (2.5)$$

$$V_P = \frac{2\pi r}{T} \quad (2.6)$$

$$r^3 = \frac{GM_E T^2}{4\pi^2} \quad (2.7)$$

$$V_P = \sqrt[3]{\frac{2\pi GM_E}{T}} \quad (2.8)$$

Despejado para M_P en la ecuación 2.4 y reemplazando de la ecuación 2.8 se obtiene la masa del planeta en función de V_E y los valores conocidos en la ecuación 2.9 Finalmente en la ecuación 2.10 se tiene la masa mínima del planeta M_M , que es una cota inferior para la masa real del planeta.

$$M_P = 2\pi \sqrt[3]{\frac{T}{G}} V_E M_E^{\frac{2}{3}} \quad (2.9)$$

$$M_M = M_P \sin i = 2\pi \sqrt[3]{\frac{T}{G}} V_A M_E^{\frac{2}{3}} \quad (2.10)$$

Así, con esta primera aproximación se puede caracterizar el planeta respecto a los parámetros de periodo T , radio orbital r y masa mínima M_M .

Para el caso de múltiples planetas se puede considerar que la interacción de la estrella con los planetas es mucho mayor que la interacción entre planetas. Con esto, el problema de múltiples planetas se reduce a la superposición de múltiples instancias de un cuerpo. Así se tiene que la velocidad observada de la estrella corresponde a la suma de las velocidades de la estrella causada por cada uno de los planetas, como se muestra en la ecuación 2.11.

$$V_E(t) = \sum_{i=1}^n V_i(t) \quad (2.11)$$

Adicionalmente, en el caso de que se tenga una detección por tránsito para el exoplaneta, se conoce la inclinación y su radio. Así es posible conocer la masa real y la densidad del planeta.

2.2.4.4. Adquisición de datos

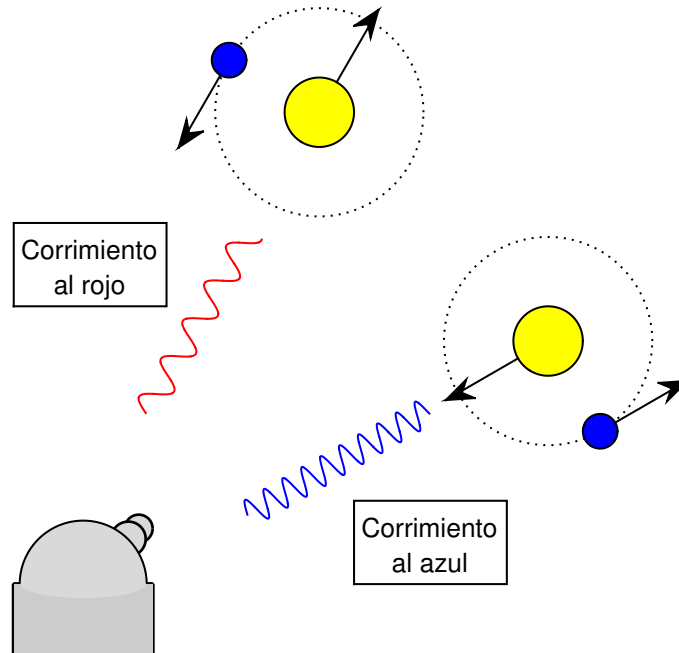


Figura 2.4: Descripción gráfica del efecto Doppler en la medición de velocidad para la detección de exoplanetas.

Para determinar la velocidad radial de las estrellas V se utiliza el efecto Doppler, al medir el cambio de color de la luz emitida por la estrella. Así, un cambio de color hacia el rojo indica que la estrella se aleja y un corrimiento al azul que esta se acerca, como se muestra en la Figura 2.4. Utilizando la ecuación 2.12, donde λ_0 es la longitud de onda de la estrella, λ es la longitud de onda medida y c es la velocidad de la luz, se puede determinar la velocidad radial.

$$V = \frac{\lambda - \lambda_0}{\lambda_0} c \quad (2.12)$$

De esta manera, al observar una variación en el color de la luz, se puede calcular la variación en la velocidad. Si bien con una buena caracterización de la estrella se puede calcular la velocidad exacta a la que se mueve la estrella, en este problema solo interesan las variaciones. Así, la órbita del planeta alrededor de su estrella se manifiesta en una variación periódica de la velocidad de la estrella y esta en una variación periódica del

color.

2.2.4.5. Instrumentos y bases de datos, HARPS y Keck

El HARPS (High Accuracy Radial Velocity Planetary Searcher) es un instrumento para detectar la velocidad radial de las estrellas para la búsqueda de exoplanetas. Este se encuentra instalado en el telescopio de $3.6m$ del observatorio de La Silla. El instrumento consiste en un sensor CCD alimentado por dos fibras ópticas, una de referencia y otra proveniente del telescopio. El equipo se mantiene en una cámara de vacío y su temperatura se regula con aire acondicionado. Con esto se logra determinar la velocidad radial con precisión de $1\frac{m}{s}$ [10].

El HARPS se encuentra en funcionamiento desde enero del 2003, con lo que se tiene a disposición bases de datos de años de duración para la búsqueda de exoplanetas.

Otras bases de datos disponibles son las provenientes del observatorio W. M. Keck, en el volcán inactivo Mauna Kea en Hawai. Este cuenta con dos telescopios de $10m$ [11].

2.2.4.6. Causas del muestreo no uniforme

Idealmente se quisiera realizar la medición de manera frecuente y uniforme. De tener tales mediciones el problema se reduciría a realizar una transformada de Fourier, identificar las distintas componentes periódicas y caracterizar los planetas. Sin embargo, las mediciones disponibles difieren de este ideal por múltiples motivos.

La presencia del Sol en nuestro cielo oculta una gran región del espacio. La traslación de La Tierra en torno al Sol causa que esta región varíe durante el año. Así, las estrellas solo pueden ser observadas durante cierta época del año generando una ventana anual en los datos muestreados.

Del mismo modo, los limitados recursos astronómicos y la abundancia de objetos de interés no permiten observar las estrellas todas las noches. De esta forma, se reserva las noches más oscuras, noches sin luna o escasa luz de ésta, para medir las estrellas más opacas, y las más brillantes se registran cuando la luna es más brillante. Independientemente de la categoría a la cual corresponda nuestro objeto de estudio los datos tendrán una ventana mensual en la toma de datos.

Así mismo habrá una pequeña fracción de días en los cuales por razones climáticas, técnicas o por otras razones imprevistas no se realiza una medición. Esto en conjunto

con lo anterior causa que se tenga una ventana de muestreo no uniforme que es en parte aleatoria y en parte determinística [12].

2.2.4.7. Consecuencias del muestreo no uniforme

El uso de una ventana de muestreo provoca una modificación de la señal, que introduce nuevas frecuencias y modifica las existentes.

Es un resultado conocido que el uso de una ventana de muestreo uniforme puede representar correctamente una señal compuesta por frecuencias que sean inferiores a la mitad de la frecuencia de muestreo [13].

Toda componente que excedan este límite, también conocido como frecuencia de Nyquist, será mal representada y causará interferencia. Esto es ya que la ventana de muestreo replicará cada componente de frecuencia f_0 , para $f = n f_m \pm f_0$, donde f_m es la frecuencia de muestreo y n es un número entero.

Así, cuando una señal está contenida en un ancho de banda que corresponda a la mitad de la frecuencia de muestreo, existirá una relación uno a uno entre las componentes de la señal muestreada y la original. Sin embargo, si hay una componente fuera de este ancho de banda, esta será representada por una componente dentro de éste que satisfaga $f = n f_m \pm f_0$.

Adicionalmente si dos componentes tienen frecuencias que satisfacen que $f_1 = n_1 f_m \pm f_0$ y $f_2 = n_2 f_m \pm f_0$ con $f_1 \neq f_2$, estas serán representadas como una única componente. Este efecto causará interferencia la cual puede ser constructiva o destructiva y evitara la correcta identificación de estas componentes.

En el caso de que se tenga una ventana de muestreo no uniforme esta tendrá sus propias componentes de frecuencia. Esto causará un patrón de interferencias único para cada ventana de muestreo. Esto producirá que las componentes reales varíen en amplitud y nuevas frecuencias aparezcan como suma de la interferencia de múltiples elementos reales de la señal.

Por esto se debe utilizar métodos que consideren esta interferencia, de manera permitan distinguir las señales reales de las generadas por la medición.

2.2.4.8. Manchas solares

Al buscar componentes periódicas dentro de la medición de una estrella es importante considerar la rotación y actividad de la estrella.

Una mancha solar es un efecto de la actividad de la estrella. Estas se ubican en una posición fija dentro de la estrella y tienen un periodo de duración acotado, pero considerable.

Una mancha solar típica para una estrella como nuestro Sol de 0.5% de cobertura generara una señal de $0.5 \frac{m}{s}$. En contraste un planeta como La Tierra genera una señal de $0.09 \frac{m}{s}$ [14].

Asi se tiene una señal real de amplitud considerable que puede ser detectada. Multiples manchas durante la medicion de los datos generaran señales con la misma frecuencia, pero que cambiaran de fase al estar ubicadas en distintas posiciones de la estrella. Por esto es importante considerar si las señales medidas mantienen su fase a lo largo de toda la señal.

2.3. Métodos de análisis de señales

2.3.1. Introducción

A continuación se presenta las diversas técnicas para analizar las señales de velocidad radial. En la sección 2.3.2 se presenta el método clásico del periodograma Lomb Scargle. Luego en 2.3.3 se analiza la versión generalizada del periodograma.

La sección 2.3.4 presenta la técnica alternativa del pre-blanqueado.

Finalmente en las secciones 2.3.5 y 2.3.6 se presentan las técnicas más modernas de la máxima verosimilitud y el análisis Bayesiano respectivamente.

2.3.2. Lomb Scargle

El enfoque clásico para identificar planetas dentro de una señal de velocidad radial es el uso del periodograma Lomb-Scargle [15] [16]. Este realiza un análisis de tipo Fourier en el cual se obtiene la potencia de la señal para cada frecuencia de acuerdo a las ecuaciones 2.13 y 2.14.

$$P_X(\omega) = \frac{1}{2} \left\{ \frac{\left[\sum_{j=1}^n X_j \cos(\omega(t_j - \tau)) \right]^2}{\sum_{j=1}^n X_j \cos^2(\omega(t_j - \tau))} + \frac{\left[\sum_{j=1}^n X_j \sin(\omega(t_j - \tau)) \right]^2}{\sum_{j=1}^n X_j \sin^2(\omega(t_j - \tau))} \right\} \quad (2.13)$$

$$\tan(2\omega\tau) = \frac{\sum_{j=1}^n \sin(2\omega t_j)}{\sum_{j=1}^n \cos(2\omega t_j)} \quad (2.14)$$

Como se puede apreciar estas ecuaciones solo entregan la potencia y no indica las características, de amplitud y fase, de las componentes de esta señal.

El procedimiento habitual consiste en identificar las frecuencias de las señales más prominentes en periodograma y seleccionar solo la más fuerte. En este punto se calcula la órbita Kepleriana que mejor se ajusta a los datos para un planeta de esa frecuencia. Así, se obtiene la caracterización del primer planeta para el sistema estelar en estudio.

Conocidas las características de este planeta se reconstruye la señal generada por este y se sustrae de la data medida. Con esto se obtiene un segundo set de datos a analizar en la cual ya no se encuentra el planeta ya detectado.

El proceso se repite de manera de obtener un nuevo periodograma en el cual se detecta un nuevo planeta. Esto se continúa hasta que ninguna frecuencia sea lo suficientemente prominente, para considerar que corresponda un planeta. En este punto se considera que la señal restante corresponde únicamente al ruido y ya no se puede encontrar ningún otro planeta.

En esta condición final se considera que se han detectado todos los planetas de este sistema y de haber otro su señal es demasiado débil para ser detectada con los datos disponibles.

Este enfoque ha tenido gran éxito en detectar exoplanetas y es altamente utilizado y aceptado en esta área de investigación. Sin embargo no es perfecto y no siempre permitirá detectar todos los planetas de un sistema [17].

El primer problema de este enfoque es que no existe garantía de que la señal del planeta caracterizado sea la misma que se observó en el periodograma. Ya que este solo entrega la información de potencia.

Otro problema es que las señales sinusoidales no son una base ortogonal en

muestreo no uniforme. Por esto al eliminar una señal identificada también se puede estar eliminando parte de otra componente real.

2.3.3. Lomb Scargle generalizado

El periodograma Lomb Scargle generalizado [18] se basa en el mismo formalismo matemático que el periodograma Lomb-Scargel [15] [16] antes descrito, se intenta superar dos debilidades de este. La primera que originalmente se asumía que la media de la señal era igual a la media de la data disponible. Y la segunda es que el método no toma en cuenta los errores de medición disponibles.

La primera la supera al cambiar la función objetivo. Mientras el método original calculaba la potencia de la señal asumiendo que esta era generada por la ecuación 2.15 el método generalizado asume que proviene de la ecuación 2.16.

$$y(t) = A \cos(\omega t + \phi) \quad (2.15)$$

$$y(t) = a \cos(\omega t) + b \sin(\omega t) + c \quad (2.16)$$

La segunda se supera al utilizar sumas ponderadas donde el peso de cada dato está dado por las ecuaciones 2.17 y 2.18, en las cuales ω_i corresponde al peso asociado al dato i , σ_i^2 es la varianza del dato y W el factor para normalizar los pesos.

$$\omega_i = \frac{1}{W\sigma_i^2} \quad (2.17)$$

$$W = \sum \frac{1}{\sigma_i^2} \quad (2.18)$$

Aparte de estas modificaciones que hacen el método más robusto ante señales de gran excentricidad y datos donde se tiene un error de medición distinto y conocido para cada dato. Este método utiliza el periodograma en la misma forma iterativa que el método original. Así mismo como se asume que existe una sola señal en cada iteración se tienen los mismos problemas que en el método original.

2.3.4. Pre-Blanqueado

El método de pre blanqueado filtra la señal para eliminar fuentes de ruido conocido de la señal, para luego pasar al espacio de Fourier. Luego se ocupa el mismo método iterativo de los métodos anteriores.

Este método al igual que los anteriores tiene la ventaja de un rápido cómputo y simpleza de aplicación. Sin embargo, sufre de similares debilidades ya que asume que las señales son sinusoidales y considera una única señal por iteración. Además el uso de este método requiere conocimiento previo de las características del ruido.

Un ejemplo de su utilización es la detección de dos planetas en la estrella CoRot 7 [19].

2.3.5. Máxima verosimilitud

El método de máxima verosimilitud [20] [21] [22] se diferencia de los métodos anteriores es un método global de búsqueda de planetas. Este genera un periodograma donde se grafica la verosimilitud logarítmica de que la data fuera generada por el modelo que se está analizando versus el modelo.

La ecuación 2.19 muestra la función de verosimilitud donde m es la señal medida de largo N , θ es el conjunto de parámetros del modelo, σ_i y σ_l representan las componentes de ruido en la estrella y en los instrumentos respectivamente y v_l es la función del efecto de la órbita Kepleriana del planeta sobre la estrella.

$$L(m|\theta) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi(\sigma_i^2 + \sigma_l^2)}} e^{\left\{ \frac{-[m_i - v_l(t_i)]^2}{2(\sigma_i^2 + \sigma_l^2)} \right\}} \quad (2.19)$$

Este método tiene la ventaja de entregar una solución precisa y confiable para los parámetros que caracterizan a los planetas. Sin embargo, este método requiere mucho más tiempo de cómputo para encontrar una solución.

2.3.6. Análisis Bayesiano

El último y más reciente método es el del Análisis Bayesiano [23] [24], el cual utiliza un análisis global. En este se considera la solución, para el modelo Kepleriano de múltiples

planetas, errores sistemáticos, y fuentes de ruido correlacionado y no correlacionado. Este método utiliza el método adaptativo Metropolis-Hastings con cadenas de Markov [25] para la estimación de parámetros. Utilizando distribución de densidad de probabilidad a posterior para detectar señales en la data.

El método se resume en la ecuación 2.20 en la cual se considera la muestra f del set de datos d con un offset γ_d , una tendencia lineal γ , un modelo Kepleriano F_k , un ruido blanco $\varepsilon_{f,d}$; una estimación de la actividad estelar con parámetros $c_{z,d}$ y $\xi_{z,f,d}$ y una estimación del ruido correlacionado con el modelo de media móvil con parámetros $\phi_{z,d}$ y τ_d .

$$m_{f,d} = \gamma_d + \gamma t_f + F_k(t_f) + \varepsilon_{f,d} + \sum_{z=1}^q c_{z,d} \xi_{z,f,d} + \sum_{z=1}^p \phi_{z,d} \exp \left\{ \frac{t_{f-z} - t_f}{\tau_d} \varepsilon_{z,d} \right\} \quad (2.20)$$

Este método es el menos eficiente de los métodos para el analizar velocidades radiales, ya que requiere realizar largas cadenas de Markov y aun no es ampliamente utilizado. Sin embargo actualmente es el método más flexible al permitir explorar el espacio de parámetros de múltiples maneras. Además, permite la visualización completa del espacio de parámetros cuando las cadenas están completas.

Este método ha demostrado ser el más robusto en cuanto a detección de señales y supresión de falsos positivos hasta el momento.

2.4. Teorema de descomposición de Wold

El teorema de descomposición de Wold [26] [27] enuncia que dado un proceso estacionario X existe:

1. Un proceso puramente determinístico D .
2. Un ruido no correlacionado de media cero R .
3. Un filtro de media móvil C .

De manera que: $X = R * C + D$.

Este teorema indica que es posible descomponer las señales medidas en la señal del proceso que se desea medir más un ruido. Con esto se tiene que una correcta detección de los planetas equivale a una separación del proceso determinístico del ruido,

disminuyendo las componentes remanentes al sustraer las señales detectadas.

2.5. MMSE

El mínimo error cuadrático medio, o MMSE por sus siglas en inglés, es un criterio que permite seleccionar un modelo que mejor represente los datos disponibles.

Para un set de mediciones x y un modelo y con parámetros θ se calcula el error de estimación según la ecuación 2.21. Esto es equivalente a considerar que el error de estimación es la energía en los remanentes. Luego los parámetros se estiman de manera de minimizar este error como se muestra en la ecuación 2.22.

$$E(x, \theta) = \sum_{i=1}^N (x_i + y_i(\theta))^2 \quad (2.21)$$

$$\bar{\theta} = \operatorname{argmin}\{E(x, \theta)\} \quad (2.22)$$

Basado en este esquema se puede caracterizar los planetas de un sistema por un modelo que incluya la medición de la velocidad de la estrella en las múltiples señales periódicas. Cada una de las cuales tendrá una frecuencia característica asociada con el periodo orbital del planeta.

Capítulo 3

Metodología

3.1. MMSE sin restricciones

3.1.1. Introducción

A continuación se analiza la solución del problema de encontrar el menor MMSE sin imponer ninguna restricción, el cual fue originalmente presentado en [28].

En la sección 3.1.2 se presenta la formulación del problema y en la sección 3.1.3 se analiza la factibilidad de obtener una solución para el problema. Se utiliza los mismos valores numéricos supuestos en el paper con el propósito de hacer ambos análisis comparables.

3.1.2. Descripción del problema

Se tiene un conjunto X compuesto por N_T mediciones, en el cual se puede estimar el número de componentes sinusoidales N_C y el set de componentes $S_{N_C} = \{C_1, \dots, C_i, \dots, C_{N_C}\}$, donde la componente C_i queda definida por sus parámetros, como se indica en la ecuación 3.1. De esta manera el conjunto de componentes S_{N_C} puede expresarse, como la colección de parámetros de acuerdo con la ecuación 3.2.

$$C_i(t) = A_i \cos(\omega_i t + \phi_i) \quad (3.1)$$

$$S_{N_C} = \{(\omega_1, A_1, \phi_1), \dots, (\omega_i, A_i, \phi_i), \dots, (\omega_{N_C}, A_{N_C}, \phi_{N_C})\} \quad (3.2)$$

Utilizando un modelo de estimación correspondiente a la suma de las componentes C_i , y el criterio de MMSE descrito en las ecuaciones 2.21 y 2.22, se obtiene que el conjunto S_{N_C} se puede calcular mediante la ecuación 3.3.

$$S_{N_C} = \underset{\text{argmin}}{\left\{ \sum_{j=1}^{N_T} \left(X(t_j) - \sum_{i=1}^{N_C} A_i \cos(\omega_i t_j + \phi_i) \right)^2 \right\}} \quad (3.3)$$

3.1.3. Complejidad computacional

Encontrar la solución de la ecuación 3.3 requiere una exploración exhaustiva de los parámetros N_c , ω , A y ϕ . Aun considerando un valor fijo a priori para N_c y un número limitado de niveles para los parámetros ω , A y ϕ , el número de combinaciones excede rápidamente las capacidades computacionales actualmente disponibles.

Considerando niveles acotados de cuantización K_ω , K_A y K_ϕ , para las frecuencias, amplitudes y fases respectivamente, se puede calcular el número de combinaciones posibles N_{Cb} , para cada valor de N_c utilizando la ecuación 3.4. Esta presenta una exploración más eficiente de la sugerida en [28] al evitar soluciones que contengan dos o más componentes de la misma frecuencia o sean permutaciones de soluciones ya analizadas.

$$N_{Cb} = \binom{K_\omega}{N_C} (K_A K_\phi)^{N_C} \quad (3.4)$$

Suponiendo que la operación más compleja es el coseno y que fuera posible realizarla en un ciclo de máquina, el tiempo de cómputo sin restricciones T_{SR} puede calcularse con la ecuación 3.5 donde V_{CPU} es la velocidad de la CPU.

$$T_{SR} = \frac{N_{Cb} N_T}{V_{CPU}} \quad (3.5)$$

Considerando un nivel alto de precisión para la frecuencia $K_\omega \approx 10^3$ valores moderados para la amplitud $K_A \approx 10^2$ fase $K_\phi \approx 10^2$ y número de muestras $N_T \approx 10^2$ y un valor muy elevado de velocidad de CPU $V_{CPU} \approx 10\text{GHz}$ se puede estimar el número de combinaciones y el tiempo necesario para calcularlas.

Tabla 3.1: Tiempo teórico de cómputo MMSE sin restricciones.

N_C	Tiempo	Nº Combinaciones
1	1.0e-01 Segundos	1.0e+07
2	5.8e+00 Días	5.0e+13
3	5.3e+04 Años	1.7e+20
4	1.3e+11 Años	4.1e+26
5	2.6e+17 Años	8.3e+32
6	4.3e+23 Años	1.4e+39
7	6.2e+29 Años	1.9e+45
8	7.6e+35 Años	2.4e+51

En la Tabla 3.1 se incluyen los valores para hasta 8 planetas, esto equivale a realizar el análisis al sistema solar. Se observa que los cálculos no pueden ser realizados, en un tiempo razonable, para sistemas con más de 2 planetas.

Es útil notar que [28] estima 3×10^{31} años para estimar 7 componentes mientras que aquí se estima 6.2×10^{29} años. Sin embargo a pesar de obtener estimaciones menores estas siguen siendo demasiado grandes para ser factibles por lo que las conclusiones del paper siguen siendo válidas.

3.2. Método propuesto: periodograma MMSE

3.2.1. Introducción

En esta sección se describen los cinco pasos del periodograma MMSE presentado originalmente en [28]. En 3.2.2 se explica el cómputo del periodograma, en 3.2.3 se seleccionan los candidatos a analizar a partir del periodograma, en 3.2.4 se expanden estos candidatos para construir las vecindades, en 3.2.5 se explica cómo se realiza el análisis de la grilla y finalmente en 3.2.6 se determina el resultado final a partir del análisis previo.

3.2.2. Periodograma

En una primera etapa se calcula un periodograma basado en el MMSE. Para esto se establece un set de frecuencias objetivo definido por una frecuencia máxima ω_{max} y un número de niveles K_ω , de acuerdo a la ecuación 3.6

$$\omega_k = \frac{\omega_{max}k}{K_\omega} \quad (3.6)$$

Para cada uno de los valores de ω_k se determinan los valores óptimos de amplitud A_{ω_k} fase ϕ_{ω_k} y error E_{ω_k} de acuerdo con las ecuaciones 3.7 y 3.8 . Con esto es posible generar un set de candidatos S_P según a la ecuación 3.9 que corresponde al periodograma MMSE.

$$E(\omega, A, \phi) = \sum_{j=1}^{N_T} \left(X(t_j) - A \cos(\omega t_j + \phi) \right)^2 \quad (3.7)$$

$$(A_\omega, \phi_\omega) = \operatorname{argmin}\{E(\omega, A, \phi)\} \quad (3.8)$$

$$S_P = \left\{ (\omega_k, A_{\omega_k}, \phi_{\omega_k}, E_{\omega_k}) \right\}_{k=1}^{K_\omega} \quad (3.9)$$

3.2.3. Selección de candidatos

Teniendo el periodograma se seleccionan los candidatos a analizar. Para esto se puede realizar por inspección visual en busca de mínimos MMSE o máximas amplitudes en el periodograma. Adicionalmente se puede utilizar un criterio automático que considere un umbral, de MMSE o amplitud, definido por el usuario o por un factor multiplicado por el valor promedio del periodograma como se sugiere en [29]. Así, a partir del conjunto S_P se construye un subconjunto $S_{candidato} = \{(\omega_i, A_{\omega_i}, \phi_{\omega_i}, E_{\omega_i})\}$ de componentes que minimicen el error, maximicen la amplitud o sean propuestas por otro método.

3.2.4. Vecindad

En torno a cada elemento de $S_{candidato}$ genera una vecindad de M elementos de acuerdo a la ecuación 3.10 donde δ se elige de manera de obtener M elementos.

$$V_i = \left\{ (\omega_k, A_{\omega_k}, \phi_{\omega_k}) \in | \omega_k \in [\omega_i - \delta, \omega_i + \delta] \right\} \quad (3.10)$$

Alternativamente es posible generar vecindades mediante una perturbación de los valores de $S_{candidato}$, sin embargo en este trabajo no se realizará un estudio detallado de

esta posibilidad.

3.2.5. Grilla

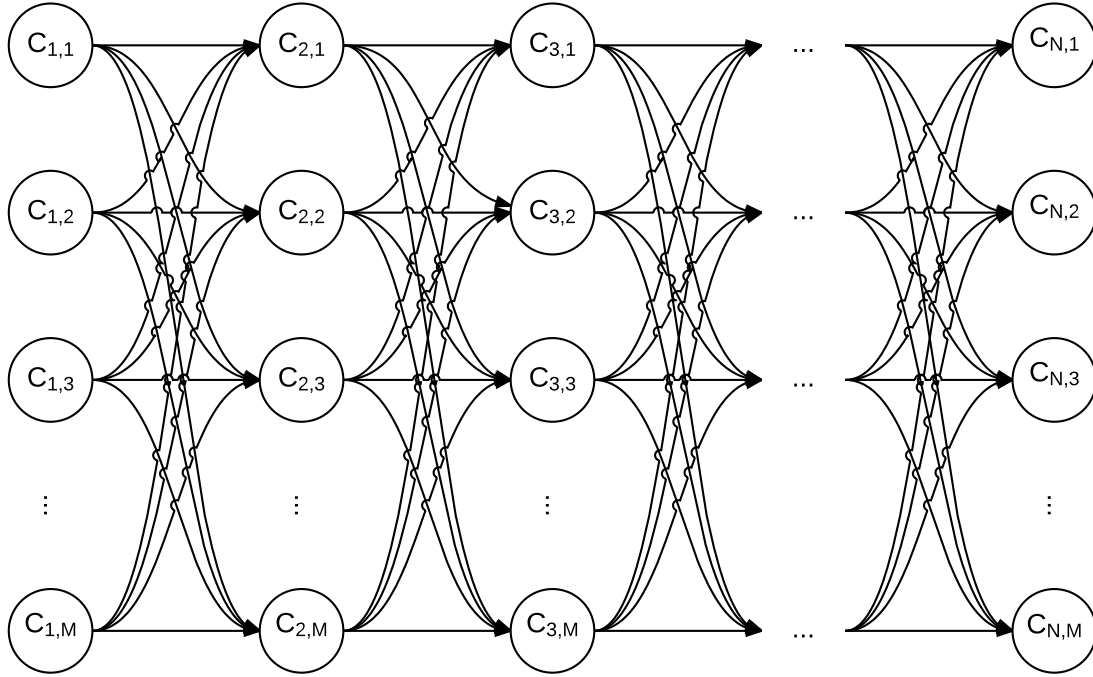


Figura 3.1: Ejemplo del diagrama de grilla para selección de N componentes con vecindades de largo M .

Para cada sub conjunto s de N elementos con $1 \leq s \leq \binom{N_C}{N}$ de $S_{candidatose}$ determina el conjunto óptimo de componentes SO_N según la ecuación 3.11 donde $1 \leq i \leq N$ y $1 \leq k_i \leq M_i$ siendo M_i el largo de las vecindades V_i .

$$SO_N = \underset{\cdot}{\operatorname{argmin}} \left\{ \sum_{j=1}^{N_T} \left(X(t_j) - \sum_{i=1}^N A_{i,k_i}^s \cos(\omega_{i,k_i}^s t_j + \phi_{i,k_i}^s) \right)^2 \right\} \quad (3.11)$$

3.2.6. Selección de componentes

Finalmente se determina el número óptimo de componentes N_o tal que el conjunto SO_N asociado tenga un error mínimo, como se muestra en la ecuación 3.12.

$$N_o = \underset{\cdot}{\operatorname{argmin}} \left[E\{SO_N\} \right] \quad (3.12)$$

3.3. Complejidad computacional

3.3.1. Introducción

En esta sección se analizarán los dos algoritmos que componen la mayor carga computacional del método. En 3.3.2 se analizará el periodograma previamente explicado en 3.2.2 y en 3.3.3 se analizará la grilla mencionada en 3.2.5.

Para ambos algoritmos se realizara un análisis de la carga computacional seguido por la optimizaciones que se realizaran para lograr códigos más eficientes que los originalmente implementados en [28].

3.3.2. Algoritmo periodograma

3.3.2.1. Introducción

A continuación se analiza la carga computacional del periodograma y como el uso del descarte de los cálculos innecesarios permitirá optimizar el algoritmo.

3.3.2.2. Análisis

El cómputo del periodograma requiere el cálculo de K_ω set de componentes cada uno de los cuales requiere una exploración exhaustiva sobre las K_A amplitudes y las K_ϕ fases. Con lo que se puede determinar el número total de errores a calcular N_{ErrP} según la ecuación 3.13.

$$N_{ErrP} = K_\omega K_A K_\phi \quad (3.13)$$

El tiempo de cómputo para el periodograma T_{Per} puede calcularse con la ecuación 3.14 donde V_{aCPU} es la velocidad aparente de la CPU en cosenos por segundo y N_T es el largo de la señal.

$$T_{Per} = \frac{N_{ErrP} N_T}{V_{aCPU}} \quad (3.14)$$

3.3.2.3. Optimización

Considerando que el tiempo de cálculo de una comparación es despreciable respecto del cálculo de un coseno. Es posible reducir el número de cálculos necesarios usando el descarte de errores, este consiste en verificar si el error acumulado hasta la muestra i con $1 \leq i \leq N_T$ es menor al error mínimo actual. El error actual se debe reiniciar cada $K_A K_\phi$ cuando se inicia el cálculo para una nueva frecuencia. Con esto el tiempo se reduce en un factor F , como se muestra en la ecuación 3.15.

$$T_{Per} = \frac{N_{ErrP} N_T}{Va_{CPU}} F; 1 < F \leq 1 \quad (3.15)$$

Ya que el valor del factor F depende tanto de los valores obtenidos por el periodograma y del orden en que se explore el espacio de parámetros este tiene un comportamiento estadístico cuyo cálculo escapa del interés de este trabajo y solo se calculará empíricamente.

3.3.3. Método de la grilla

3.3.3.1. Introducción

A continuación se analiza la carga computacional del algoritmo de la grilla y como el uso del descarte de los cálculos innecesarios y el precálculo de las componentes sinusoidales permitirá reducir la carga computacional.

3.3.3.2. Análisis

El cálculo de MMSE utilizando la grilla requiere hacer un barrido sobre todos los elementos de cada vecindad de la combinación que se está evaluando. Además se deben explorar todas las combinaciones posibles.

$$E_{ErrGr} = \binom{T_C}{N_C} M^{N_C} \quad (3.16)$$

El número de errores a calcular N_{ErrGr} queda definido por la ecuación 3.16. donde M es el tamaño de las vecindades N_C el número de componentes y T_C el total de candidatos.

Luego el tiempo de cómputo se calcula con la ecuación 3.17 análoga a la ecuación

3.14, pero considerando que cada señal debe calcularse por separado.

$$T_{Per} = \frac{N_{ErrGr} N_T N_C}{V a_{CPU}} \quad (3.17)$$

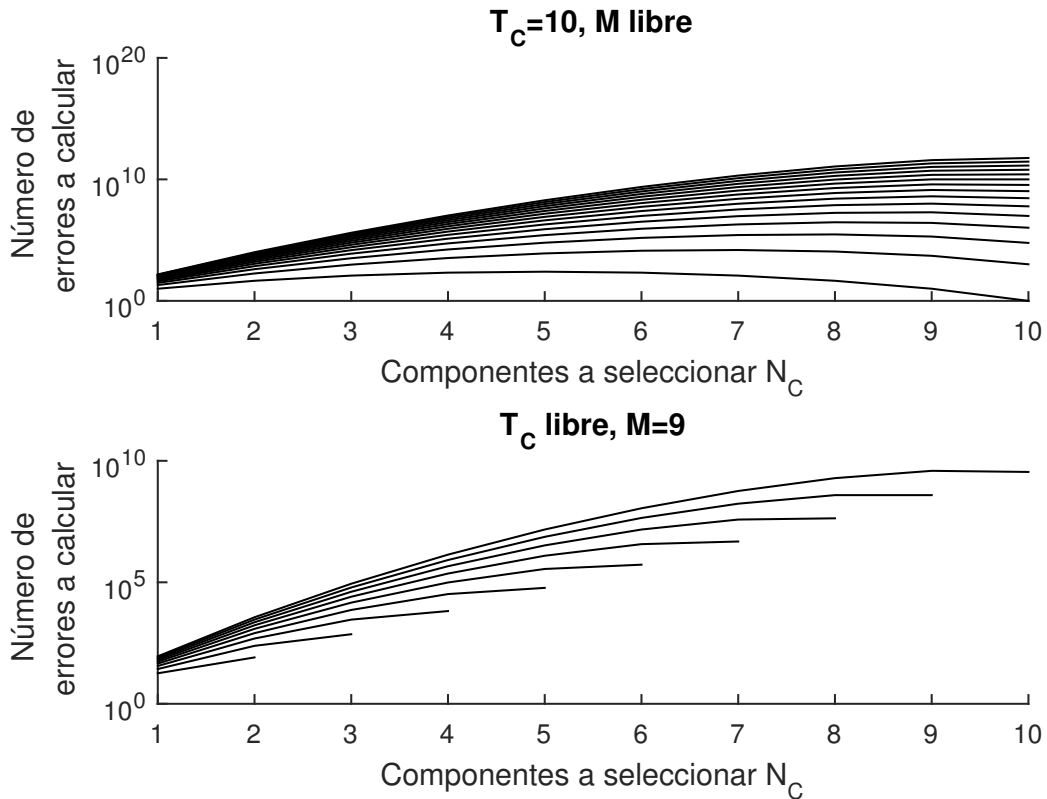


Figura 3.2: Dependencia del número de errores a calcular N_{ErrGr} con los parámetros del experimento.

En la Figura 3.2 se muestra el número de errores a calcular de acuerdo a la ecuación 3.16 y cómo la carga computacional aumenta exponencialmente al incrementar el espacio de búsqueda.

3.3.3.3. Optimizaciones

Para este código se consideran dos optimizaciones. La primera es el descarte de errores ya implementado para el periodograma. Es importante notar que en este caso el error actual no se resetea ya que el código solo calcula una solución óptima.

La segunda mejora es el precálculo de todas las componentes sinusoidales de manera de que en el cálculo de velocidad se reemplaza la velocidad aparente en cos/seg

V_{aCPU} por la velocidad aparente de la suma V_{iCPU} medida en iteraciones por segundo. Esta nueva velocidad corresponde al de la suma de operaciones realizadas en una iteración del programa y se habían considerado despreciables respecto a la del coseno. Con esto la ecuación 3.18 muestra el nuevo tiempo de cómputo.

$$T_{Per} = \frac{N_{ErrGr} N_T N_C}{V_{iCPU}} F; 1 < F \leq 1 \quad (3.18)$$

3.4. Usos adicionales método de la grilla

3.4.1. Introducción

En esta sección se explorara cómo es posible utilizar el análisis de grilla originalmente descrito en 3.2.5 para realizar tareas adicionales a la antes descrita. En 3.4.2 se analiza el uso de la grilla para refinar los resultados y en 3.4.3 se utiliza para discernir si la señal pertenece a un planeta mediante el análisis de fase.

3.4.2. Refinado de resultados

Una vez determinado el número de componentes y sus características estas pueden ser refinadas.

Para esto se crea una nueva vecindad para cada componente. En esta se incluyen además de la solución ya detectada, perturbaciones en uno o más de sus parámetros.

Al incluir las soluciones previamente detectadas y optimizar sus parámetros se obtendrá soluciones globalmente superiores o iguales a las ya detectadas.

3.4.3. Verificación de fase

Se sabe que un de los falsos positivos detectados por el método de radio velocidad es el causado por las manchas solares. Estas señales son consistentes en frecuencia pero no en amplitud y fase.

Considerando que estas señales varían en fase, se sugiere dividir la señal en dos set de datos y recalculer la fase de estas componentes al generar vecindades que contengan un set de fases uniformemente distribuido en el rango completo de estimación.

En el caso de señales reales la fase detectada en ambos set de datos debiese ser consistente mientras en el caso de manchas solares esté cambiará.

Capítulo 4

Resultados

4.1. Datos

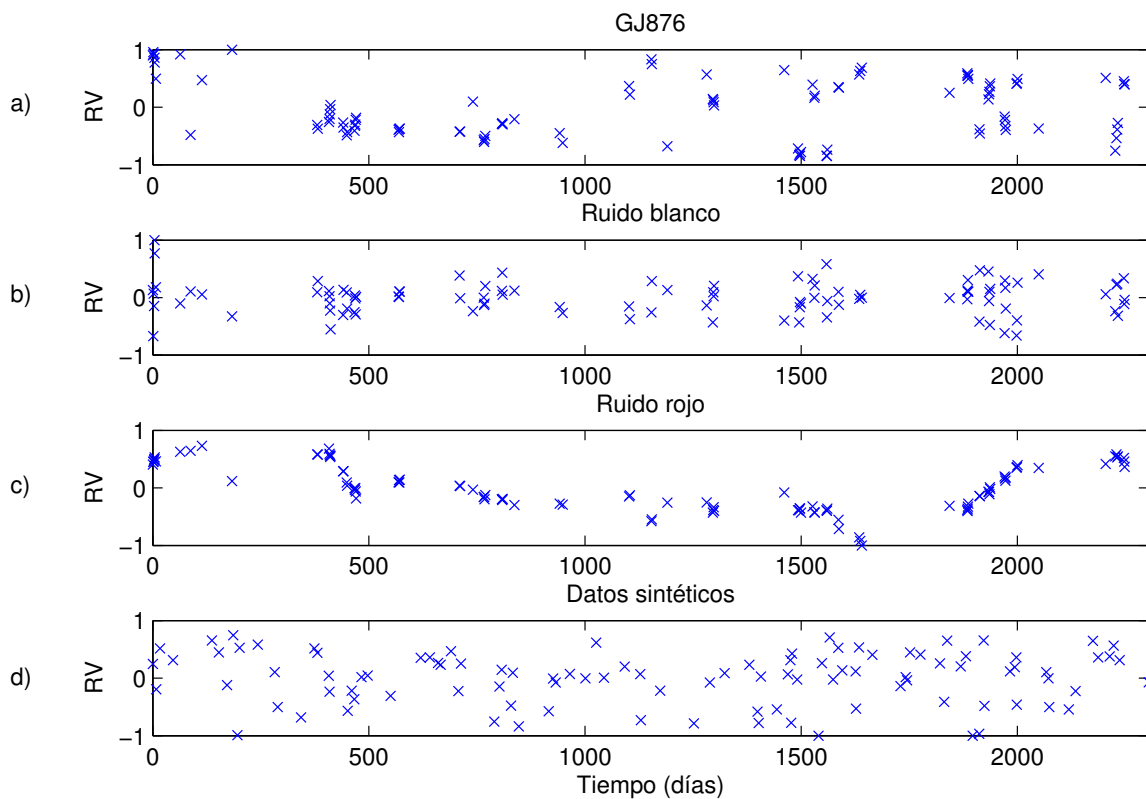


Figura 4.1: Comparación diferentes sets de datos de velocidades radiales normalizadas. En (a) estrella GJ876, en (b) ruido blanco muestreado con el patrón de GJ876, en (c) ruido rojo muestreado con el patrón de GJ876, y en (d) datos sintéticos con muestreo aleatorio.

En la Figura 4.1 se muestran algunos de los conjuntos de datos utilizados. Para simplificar la comparación se muestran cien puntos de cada uno de los conjuntos de datos, normalizados a la unidad.

El primer set de datos corresponde a los datos medidos de la estrella GJ876 [28]. Luego se muestran dos sets correspondientes a ruido blanco y rojo.

Por las características multidisciplinarias de este trabajo es útil notar que el ruido rojo utilizado decae $20 \frac{dB}{dec}$. Esto corresponde a la definición usual utilizada en audio, comunicaciones o ingeniería en general, sin embargo algunos autores del área de astronomía utilizan el término para cualquier ruido correlacionado.

Ambos ruidos fueron generados para un tiempo uniforme y luego sub muestreados de acuerdo al patrón de muestreo de GJ876.

Tabla 4.1: Componentes de la señal sintética.

Frecuencia	Amplitud	Fase
2,0944	1,0000	0,8915
0,8976	0,5000	2,6500
0,3696	0,3000	5,7537

La última señal corresponde a una señal sintética compuesta por tres señales sinusoidales con un patrón no uniforme de muestreo en el tiempo. La Tabla 4.1 indica los parámetros de las tres componentes de esta señal.

Se observa que en el tiempo es difícil distinguir si estas corresponden a ruido o a señales que contienen información.

4.2. Periodogramas MMSE Vs Lomb Scargle

En esta sección se comparan los periodogramas Lomb Scargle con los periodogramas MMSE. Para esto se muestran en paralelo ambos periodogramas en cuatro casos utilizando iguales conjuntos de frecuencias.

Todos los periodogramas MMSE de esta sección fueron calculados con 25.000 niveles de frecuencia, 500 niveles de amplitud y 200 de fase.

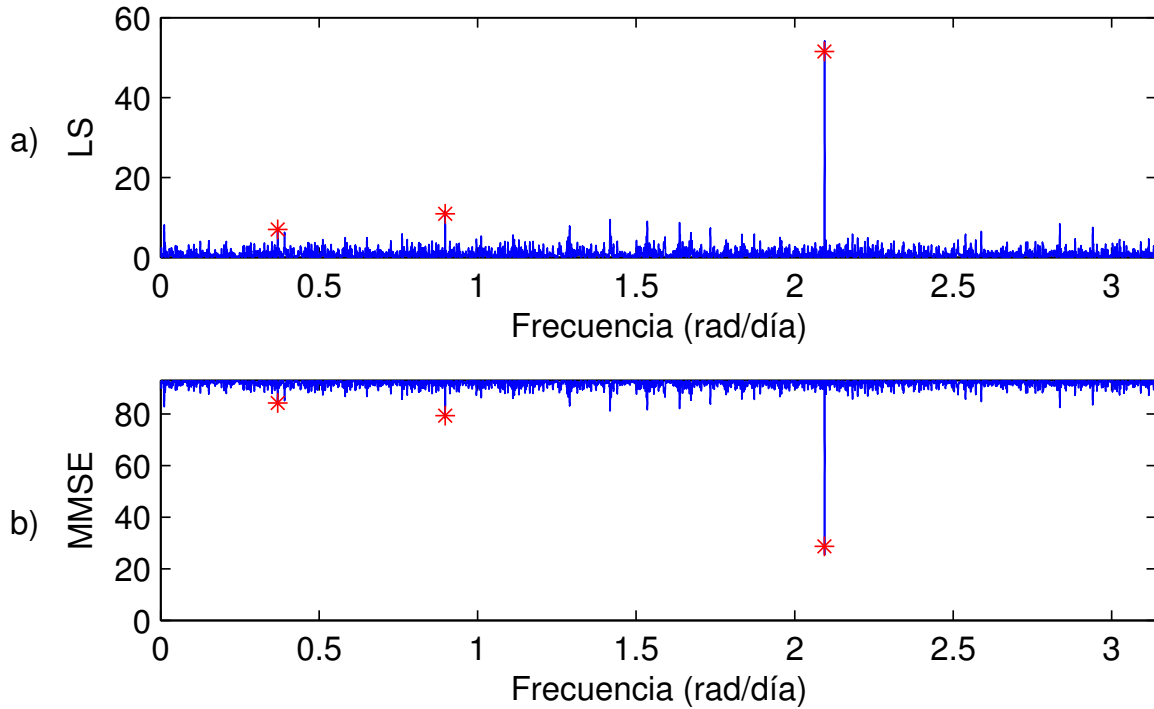


Figura 4.2: Periodogramas de la señal sintética, en (a) periodograma Lomb Scargle, y en (b) periodograma MMSE.

En la Figura 4.2 se muestra el periodograma sobre de 150 datos de la señal sintética. Se marca en la señal las frecuencias de las componentes utilizadas para crear esta señal.

Tabla 4.2: Componentes detectadas por el periodograma MMSE en la señal sintética.

Frecuencia	Error ω	Amplitud	Error A	Fase	Error ϕ	MMSE
2,0944	0,00 %	0,972	2,80 %	0,8478	0,70 %	25,3501
0,8976	0,00 %	0,453	9,40 %	2,7318	1,30 %	78,3433
0,3692	0,11 %	0,339	13,00 %	0	8,43 %	83,6349

En la Tabla 4.2 se muestra las tres componentes más prominentes en el periodograma MMSE, se incluye el error respecto de los verdaderos valores de las señales. Los porcentajes de frecuencia y amplitud se calculan respecto del valor original y los de fase respecto a un ciclo (2π).

Se observa como las frecuencias son detectadas con gran precisión mientras las amplitudes y fases tienen una variación perceptible. Llegando a un 13 % para la amplitud de la componente más pequeña.

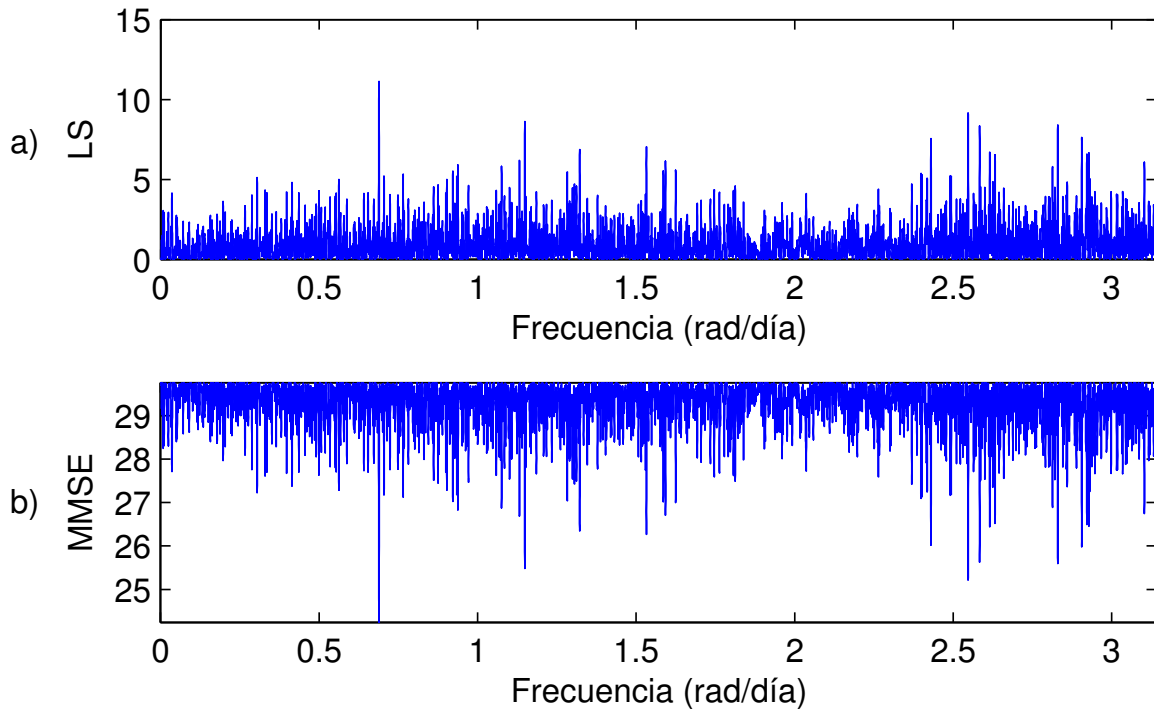


Figura 4.3: Periodogramas del ruido blanco, en (a) periodograma Lomb Scargle, y en (b) periodograma MMSE.

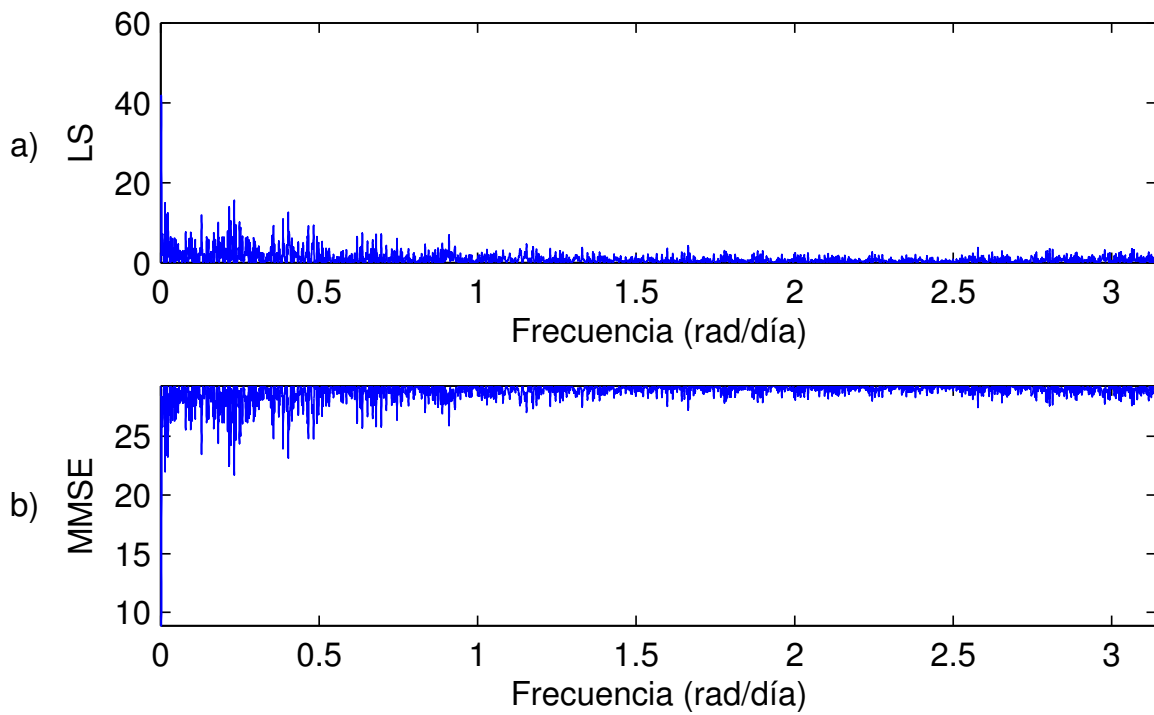


Figura 4.4: Periodogramas del ruido rojo, en (a) periodograma Lomb Scargle, y en (b) periodograma MMSE.

En las Figura 4.3 y Figura 4.4 se muestra el caso de ruidos y se obtiene resultados equivalentes. Es útil notar que algunas componentes del ruido se encuentran a niveles comparables a los valores de las señales reales y artificiales.

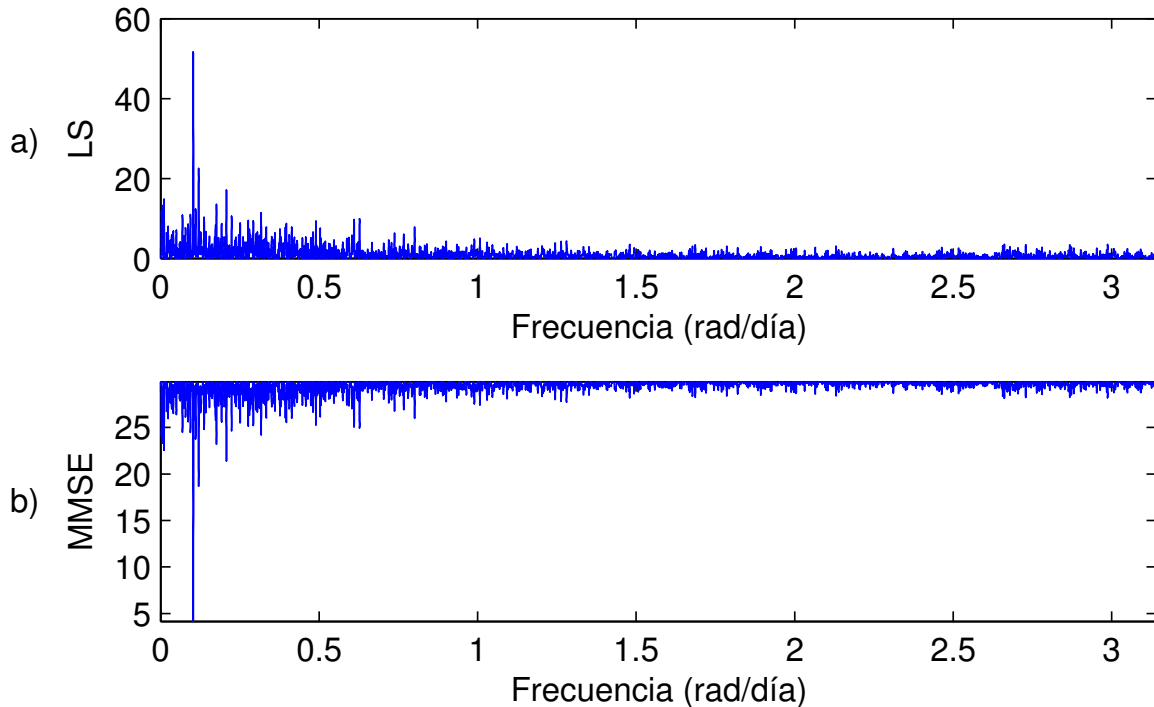


Figura 4.5: Periodogramas estrella GJ876, en (a) periodograma Lomb Scargle, y en (b) periodograma MMSE.

Finalmente en la Figura 4.5 se muestra el periodograma de la señal de la estrella GJ876 donde nuevamente ámbos periodogramas muestran similares características.

4.3. Experimento de validación para el análisis de grilla

Para verificar el comportamiento del algoritmo de la grilla este se utilizó para calcular la fase de las tres señales presentes en la señal sintética. Adicionalmente se agregó una cuarta componente no relacionada, para realizar una detección del número de componentes.

En la Tabla 4.3 se muestran las componentes utilizadas para la medición. Se utilizaron 200 fases por componente lo cual implica una resolución de 0.0314 radianes.

En la Tabla 4.4 se muestran las fases y errores detectados. En estos se observa

Tabla 4.3: Candidatos para validación del algoritmo de la grilla.

Candidato	1	2	3	4
Frecuencia	2,0944	0,8976	0,3696	0,5712
Amplitud	1.0000	0,5000	0,3000	0,6000

Tabla 4.4: Componentes y fases detectadas por la grilla en la señal sintética.

Nº Componentes	1	2	3	4	Real
Fase 1	0,9425	0,8796	0,8796	0,9739	0,8915
Fase 2		2,6389	2,6389	2,6075	2,6500
Fase 3			5,7491	5,9376	5,7537
Fase 4				2,4190	
Error	25,3457	7,2582	0,1310	24,4308	

claramente que la solución correcta incluye tres componentes y que las fases fueron detectadas correctamente con un margen de error correspondiente al nivel de resolución utilizado.

4.4. Experimento con señales ruidosas

Se utilizaron 200 datos de la señal sintética antes descrita y se adicionó ruido a distintos niveles de SNR para detectar el número de componentes. Para esto se utilizó un tamaño de vecindad de 15 elementos.

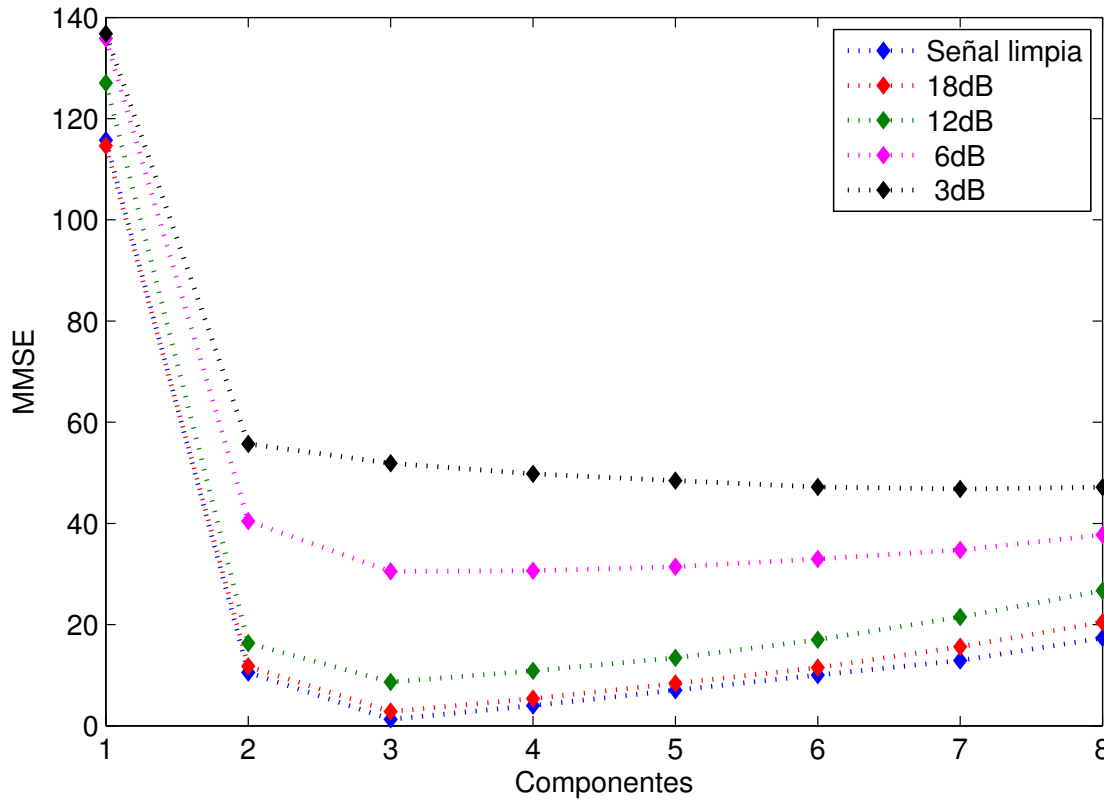


Figura 4.6: Detección de componentes señales con distinto nivel de ruido. Un mínimo error indica el número detectando, siendo tres componentes correctamente detectadas para la señal limpia y ruidos de 18dB, 12dB y 6dB.

En la Figura 4.6 se muestra cómo el método es capaz de identificar el número de componentes, que viene dado por el menor error, para la señal de hasta 6dB de SNR.

4.5. Pruebas de velocidad

4.5.1. Introducción

Para verificar la utilidad de las optimizaciones propuestas en 3.3.2.3 y 3.3.3.3 se procede a realizar diversas pruebas de velocidad. En la sección 4.5.2 se describen los equipos que se utilizarán para realizar estas pruebas. En 4.5.3 se analiza el periodograma y se estima la capacidad computacional disponible para compararla con la capacidad supuesta en la sección 3.1.3 y en [28]. En 4.5.4 se analiza el algoritmo de la grilla, en 4.5.4.1 se describen las configuraciones y estrategias para realizar los experimentos.

Luego en 4.5.4.2 y 4.5.4.3 se analizan los resultados de la comparación de códigos con y sin optimizaciones y los resultados de dividir la carga computacional en múltiples procesos respectivamente.

4.5.2. Equipos disponibles

Para realizar los experimentos se utilizaron múltiples equipos los que se describirán a continuación.

Servidor Wasabi: este es un servidor destinado a prestar servicios de red al conjunto de servidores utilizados a continuación. Este equipo no se utiliza en la mayoría de los experimentos, ya que este tiene una capacidad limitada y su uso puede limitar el uso de los otros equipos. Este cuenta con un procesador Intel Xeon E5-2407 de 2.2GHz de 4 núcleos.

El servidor Naga es un equipo destinado a realizar experimentos y es utilizado a través de Wasabi. Este cuenta con 64 núcleos del procesador AMD Opteron 6272.

El servidor Yeti es un equipo destinado a realizar experimentos y es utilizado a través de Wasabi. Este cuenta con 64 núcleos del procesador AMD Opteron 6276.

Adicionalmente se utilizan dos computadores de escritorio con procesadores Intel Core i7 4770 de 3.4GHz y Intel Core i7 2700K de 3.5GHz.

4.5.3. Periodograma

4.5.3.1. Descripción del experimento

Se utilizó de 150 muestras de la señal sintética descrita en la sección 4.1 y Tabla 4.1, para las cuales se calculó un periodograma de 25.000 niveles de frecuencia, 500 niveles de amplitud y 200 niveles de fase. Utilizando la ecuación 3.13 se tiene que se requiere calcular $2.5e9$ errores lo que equivale a $3.75e11$ sinusoides.

Para tener una referencia de la velocidad de los computadores disponibles se calcularon múltiples instancias del mismo periodograma, sin descarte, en las diversas maquinas disponibles. En experimentos sucesivos se incrementó el número de instancias simultáneas del programa, lo que permite simular simultáneamente el caso en que se divide la carga computacional en múltiples procesos y el caso en que se realiza el experimento simultaneo al trabajo de otros usuarios. Esto último es especialmente

relevante para el caso de los servidores.

Para los computadores de escritorio se realizaron los experimentos con un número de procesos que se incrementa linealmente de 1 a 10. Para el servidor Wasabi que no está destinado a experimentos se realizó el mismo esquema lineal pero limitado al rango de 1 a 8 procesos. Finalmente para los servidores destinados a experimentación se realizó un esquema exponencial utilizando: 1, 2, 4, 8, 16, 32 y 64 procesos.

Adicionalmente se realizó una ejecución del mismo periodograma, ejecutando la versión del código con descarte, para determinar el incremento de velocidad obtenido gracias a esta optimización.

4.5.3.2. Resultados

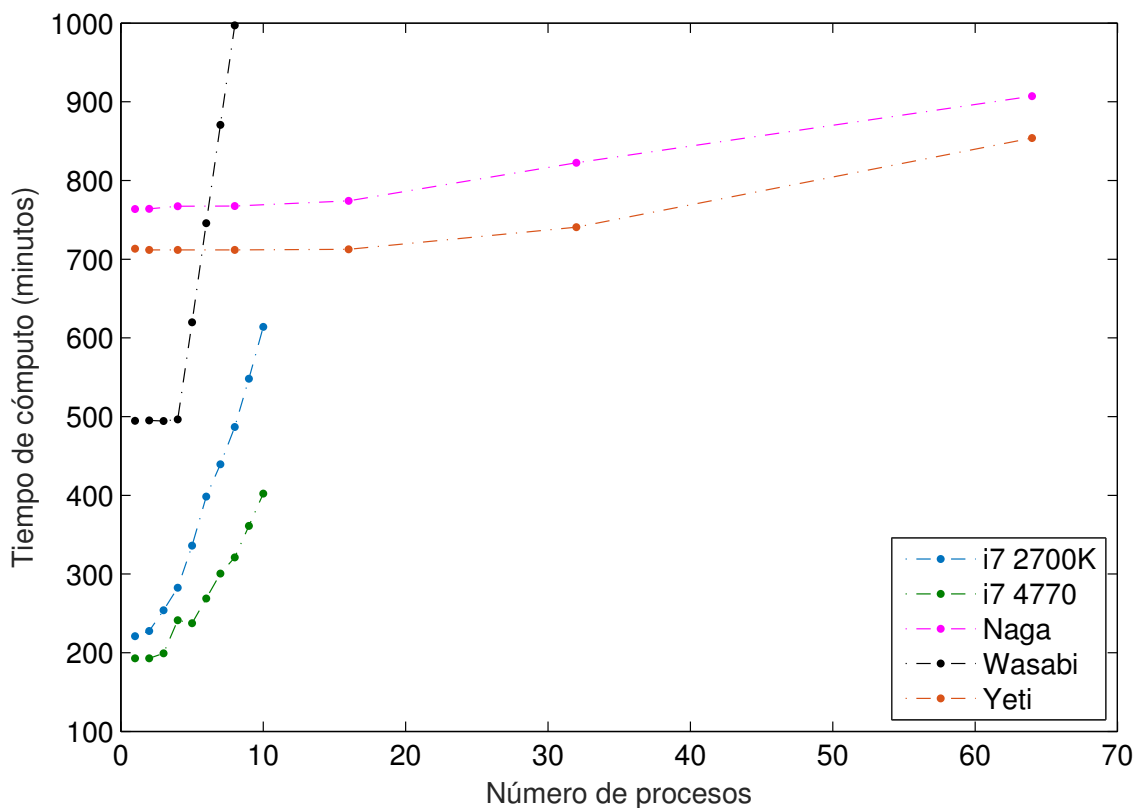


Figura 4.7: Tiempo de cómputo de un periodograma versus el número de tareas realizadas por el computador.

En la Figura 4.7 se observa como la adición de tareas simultáneas incrementa el tiempo necesario para la ejecución del cálculo. Además se observa que los servidores

tienen procesadores más lentos, pero la velocidad de estos es más estable ante la presencia de trabajos adicionales.

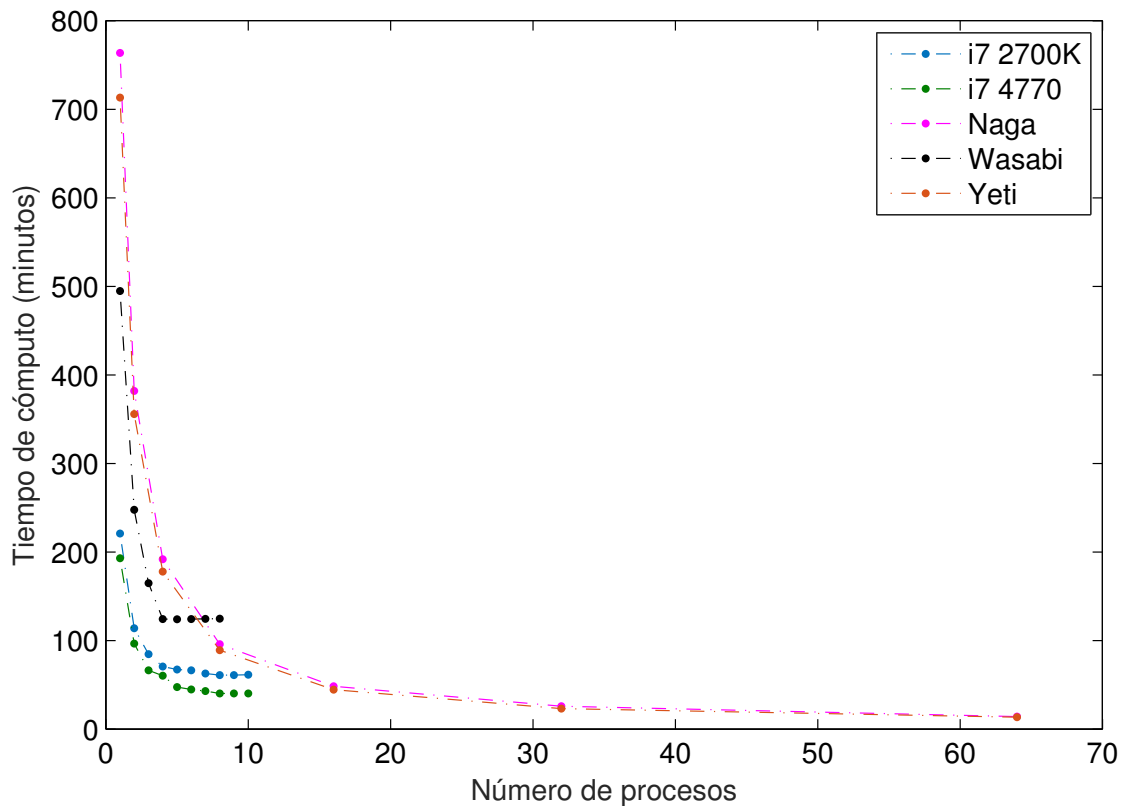


Figura 4.8: Tiempo de cómputo de un periodograma separado en múltiples procesos dentro de un mismo computador.

En la Figura 4.7 se aprecia que al dividir el cálculo en múltiples procesos se reduce el tiempo necesario para completar la tarea, hasta alcanzar el número máximo de núcleos. Lamentablemente el incremento de velocidad no es lineal, ya que los procesadores disminuyen su velocidad al usar múltiples núcleos u otros recursos compartidos, por los procesos, y crean retrasos en la ejecución de estos. Aun así, la ganancia de velocidad es suficiente, para justificar el uso de múltiples procesos para realizar los cálculos.

En la Tabla 4.5 se lista el tiempo necesario para completar la tarea utilizando al máximo los recursos disponibles. Además, sabiendo que el número de errores calculados en el programa es $2,5e9$, se puede calcular una velocidad aparente para cada máquina utilizando la ecuación 3.14. Con esta información se tiene que la capacidad computacional disponible al momento de realizar los experimentos es de $1,2e9$ cos/seg.

En la sección 3.1.3 se determinó que una velocidad de $1e10$ cos/seg era insuficiente para realizar el MMSE sin restricciones en un tiempo razonable, estas pruebas muestran

Tabla 4.5: Tiempo de cómputo MMSE periodograma y velocidades aparentes para múltiples equipos.

Equipo	Tiempo mínimo	Velocidad Aparente
i7 2700K	60,845 min	1.0e+8 cos/seg
i7 4770	40,123 min	1.6e+8 cos/seg
Naga	14,174 min	4.4e+8 cos/seg
Wasabi	123,981 min	5.0e+7 cos/seg
Yeti	13,344 min	4.7e+8 cos/seg

que esa afirmación se sostiene al considerar los recursos disponibles.

Al realizar el cálculo del periodograma con descarte se obtiene el resultado en 2:27:40 horas en el computador i7 2700K mientras el código sin descarte lo realiza en 4:48:53 horas en similares condiciones. Esto es una reducción del 48.88 % en el tiempo de cómputo.

Este es un único experimento, pero considerando que está compuesto por el cálculo sucesivo de 25.000 sets de resultados y que el incremento de velocidad es significativo, se puede determinar que la adición de esta modificación es útil.

4.5.4. Grilla

4.5.4.1. Descripción del experimento

Se utilizó el mismo set de 150 datos del experimento anterior y 4 candidatos de 200 elementos cada uno, tres de los cuales corresponden a las frecuencias y amplitudes de las señales presentes y con una distribución homogénea de fases.

Tabla 4.6: Orden de los candidatos para las diferentes configuraciones utilizadas en las pruebas de velocidad de la grilla. En negritas la componente no relacionada con la señal.

	Candidato	1	2	3	4
Configuración 1	Frecuencia	2,0944	0,8976	0,3696	0,5712
	Amplitud	1.0000	0,5000	0,3000	0,6000
Configuración 2	Frecuencia	0,5712	0,3696	0,8976	2,0944
	Amplitud	0,6000	0,3000	0,5000	1.0000
Configuración 3	Frecuencia	2,0944	0,8976	0,3696	0,8976
	Amplitud	1.0000	0,5000	0,3000	0,6000
Configuración 4	Frecuencia	0,8976	0,3696	0,8976	2,0944
	Amplitud	0,6000	0,3000	0,5000	1.0000

En la Tabla 4.6 se muestran cuatro configuraciones a para realizar los experimentos.

Configuraciones 1 y 2 corresponden a permutaciones de los candidatos utilizados en la sección 4.3. Configuraciones 3 y 4 incluyen dos componentes con igual frecuencia lo que escapa de los usos propuestos para el algoritmo y solo se ocuparan para calcular velocidades promedio del programa.

Tabla 4.7: Orden de exploración de las combinaciones de componentes para el caso de 4 candidatos.

Combinaciones 1 de 4				Combinaciones 2 de 4					
1 ^o	2 ^o	3 ^o	4 ^o	1 ^o	2 ^o	3 ^o	4 ^o	5 ^o	6 ^o
1	2	3	4	1	1	1	2	2	2
				2	3	4	3	4	4
Combinaciones 3 de 4				Combinaciones 4 de 4					
1 ^o	2 ^o	3 ^o	4 ^o	1 ^o					
1	1	1	2	1					
2	2	3	3	2					
3	4	4	4	3					
				4					

En la Tabla 4.7 se muestra el orden en que el programa explora las combinaciones. Al contrastarla con los resultados de la Tabla 4.4 se observa que en la configuración 1 el mínimo MMSE se obtendrá siempre en la primera combinación explorada, mientras en la configuración 2 siempre será en la última.

En la sección 4.5.4.2 se ejecutan 4 versiones del código, las cuales contienen distintas combinaciones de las dos mejoras a evaluar. Estas son el uso de descarte y el precálculo de las señales generadas por cada candidato.

En cada uno de los métodos se realiza la totalidad de los cálculos para encontrar la configuración óptima de candidatos, ejecutando todos los procesos simultáneamente en el computador i7 2700K. Esta modalidad no permite determinar la velocidad exacta de cada programa, pero garantiza que todos los procesos se enfrenten a similares condiciones de carga computacional hasta que el término de los procesos libere recursos. En consecuencia se garantiza que los programas que terminan primero sean más rápidos, pero no garantiza que las velocidades se mantengan al terminar los procesos adicionales. Por esto solo se considerarán comparaciones relativas entre estos programas.

En la sección 4.5.4.3 se prueba cómo se comporta el código optimizado al separar las tareas en varios procesos paralelos. Esto se realiza ejecutando el código en la configuración 2 para seleccionar 3 componentes de 4 y 4 componentes de 4.

Similar a lo realizado para la comparación de códigos cada set de experimentos se ejecuta simultáneamente en el mismo equipo. En el computador i7 2700K se divide el

trabajo en 1, 2 ,3 ,4 y 5 partes y en los servidores en 1, 2, 4, 8, 16 y 32 partes.

4.5.4.2. Resultados comparación de códigos

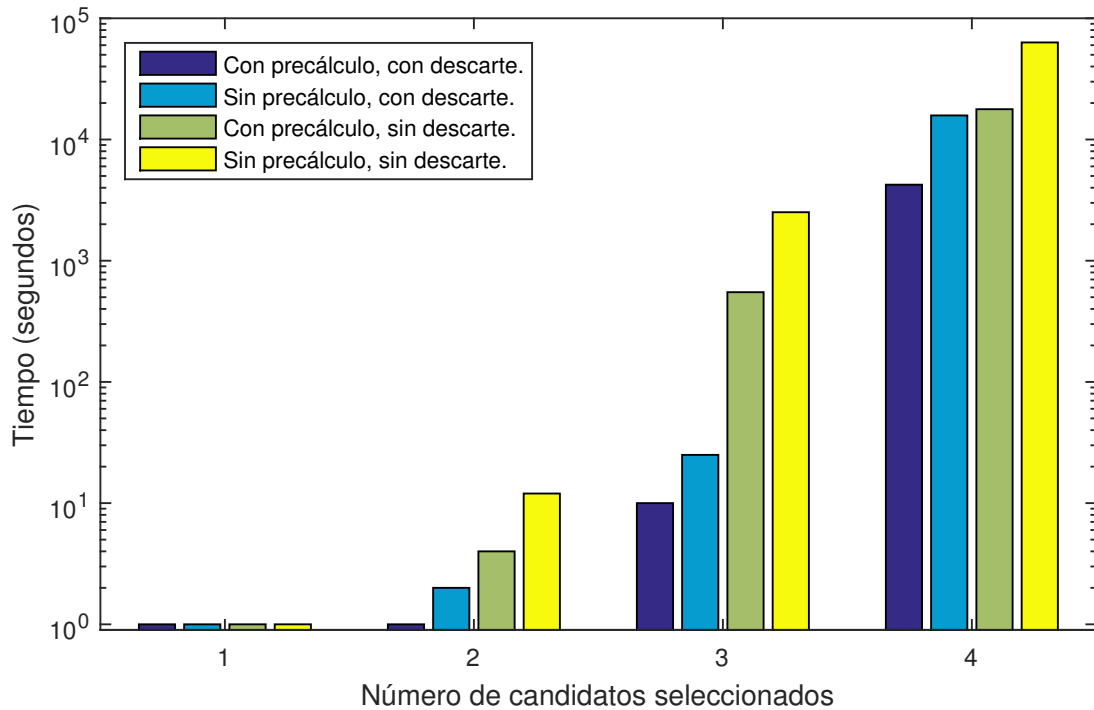


Figura 4.9: Comparación de velocidad de los códigos, para la grilla con mínimo MMSE en la primera combinación (configuración 1).

En la Figura 4.9 se muestra el caso en el cual el mínimo MMSE se obtiene para la primera combinación. Esto es una condición muy favorable para el uso del descarte ya que al encontrar errores bajos al inicio de la búsqueda gran parte de los cálculos siguientes pueden ser descartados.

En estas condiciones se observa que el uso de descarte genera una mejora mucho más significativa que el uso del precálculo de señales. Sin embargo es el uso de ambas técnicas en conjunto lo que permite tener la mayor mejora.

Se observa además que en el caso de la selección de tres componentes la mejora total es ligeramente superior a dos órdenes de magnitud. Aun sobre el código que incluye precálculo. Esto se debe a que en este caso en particular la solución exacta del problema se detecta antes de completar el primer cuarto de la ejecución.

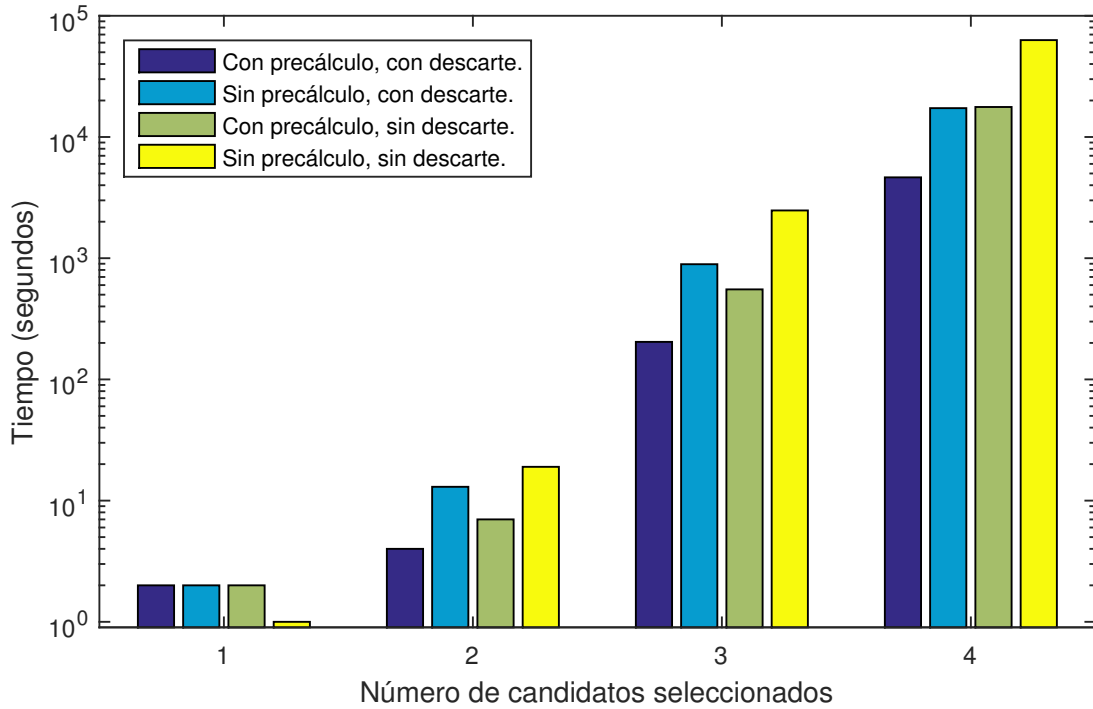


Figura 4.10: Comparación de velocidad de los códigos, para la grilla con mínimo MMSE en la última combinación (configuración 2).

En la Figura 4.10 se observa un caso menos favorable en el cual las componentes reales corresponden a las últimas en ser evaluadas. En este caso se reduce la ganancia obtenida al utilizar el descarte. Aun así, se tiene que gracias al uso de precálculo y descarte se obtiene una mejora total en torno a un orden de magnitud.

Un resultado aparente en este caso es que para una componente el método sin optimizaciones es el más rápido. Sin embargo los otros métodos registran un tiempo de ejecución de un segundo con un margen de error de un segundo, por lo que este resultado no es significativo.

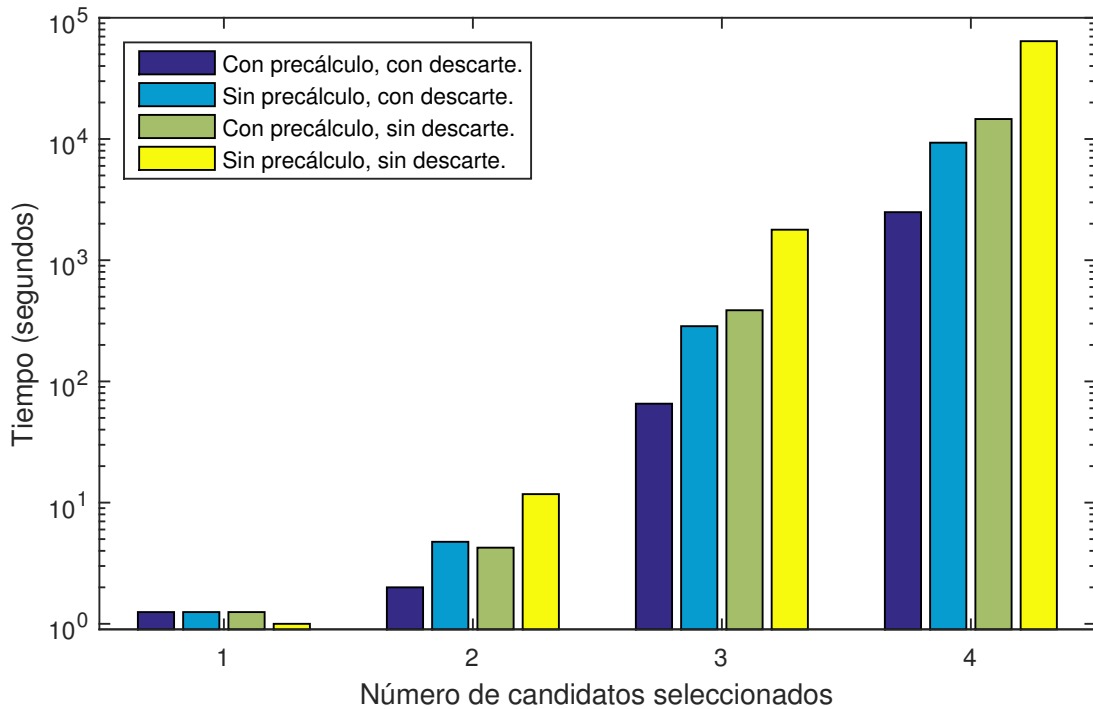


Figura 4.11: Comparación de velocidad de los códigos, para la grilla promedio sobre 4 experimentos.

En la Figura 4.11 se tiene el resultado promedio de los experimentos con las cuatro configuraciones. En este caso se observa una mejora significativa gracias a ambas técnicas. El resultado para una componente es una consecuencia del experimento con la configuración 2 ya que los otros tres experimentos tuvieron tiempos de ejecución inferiores a un segundo y no fueron registrados.

Se observa caso de la configuración 1, donde los candidatos están ordenados de la componente más fuerte en la señal a la que no está presente, se observa un mayor efecto del descarte que en la configuración 2. Estos resultados indican que podría existir una ventaja en el tiempo de cómputo al forzar las componentes más probables al inicio del cálculo. En la implementación actual esto corresponde a utilizar un índice más bajo para las componentes más probables.

4.5.4.3. Resultados procesos paralelos

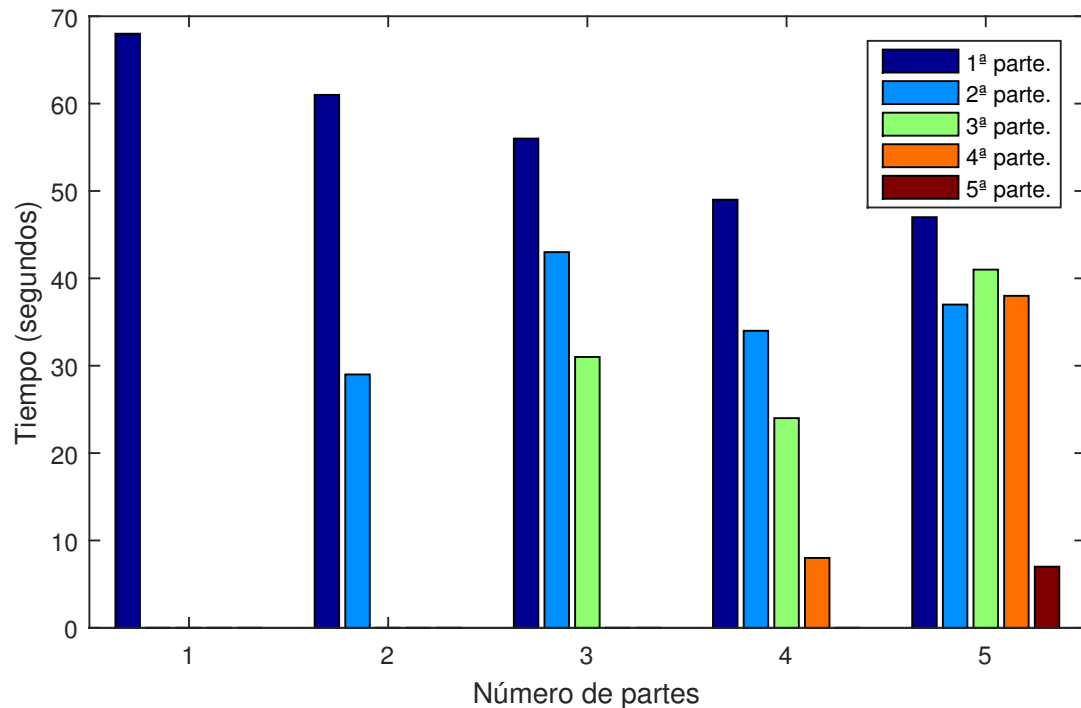


Figura 4.12: Tiempo de cómputo de la grilla versus número de procesos al calcular 3 componentes de 4 usando la configuración 2 para 1, 2, 3, 4 y 5 partes.

En la Figura 4.12 se muestra el tiempo de cómputo al dividir el cálculo en múltiples procesos. Se observa que a pesar de que cada caso el número de soluciones a analizar se reparte uniformemente el tiempo difiere considerablemente. La explicación más probable es que el número de cálculos realizados por cada proceso difiera debido al descarte.

Al dividir el trabajo en múltiples procesos cada uno operara con un resultado parcial para el MMSE distinto. En particular a partir del segundo proceso paralelo este inicia con un resultado parcial mayor al que tendría de realizarse los cálculos secuencialmente. Esto causa que al dividir el trabajo en paralelo se realicen cálculos adicionales con lo cual es natural que la suma de los tiempos de los procesos paralelos crezca con el número de procesos. Sin embargo, es posible reducir el tiempo requerido para obtener los resultados si el ahorro de tiempo por dividir el trabajo supera el trabajo extra que debe realizarse.

Como se vio en la sección anterior, la velocidad del código de la grilla es vulnerable al orden de ejecución de las componentes. Así, al realizar múltiples procesos es posible que algunos se encuentren en situaciones favorables mientras otros en situaciones desfavorables. En particular en la Figura 4.12 el primer proceso de cada caso se muestra

síntomas de un caso negativo, mientras el caso final se encuentra en una situación favorable.

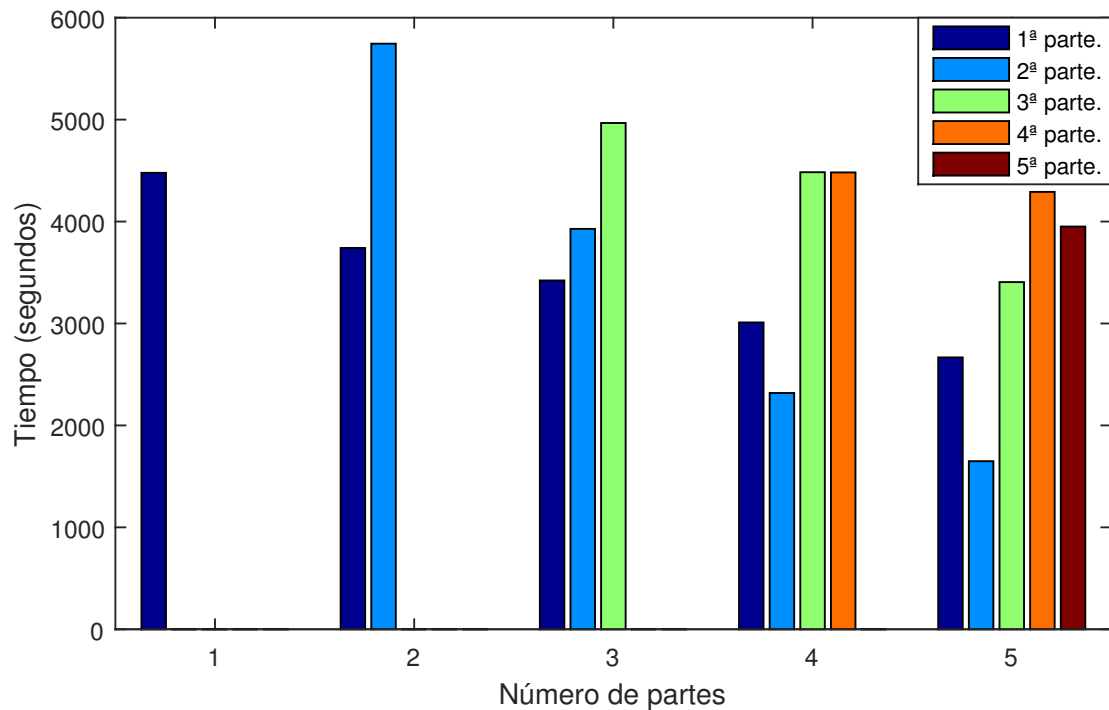


Figura 4.13: Tiempo de cómputo de la grilla versus número de procesos al calcular 4 componentes de 4 usando la configuración 2 para 1, 2, 3, 4 y 5 partes.

En la Figura 4.13 se presenta el peor escenario para el uso de procesos paralelos. En este un MMSE parcial bajo es encontrado en los primeros procesos. Esto causa que si bien el primer programa reduce su duración al aumentar el número de procesos algunas instancias adicionales del programa toman más tiempo, que la ejecución del programa original.

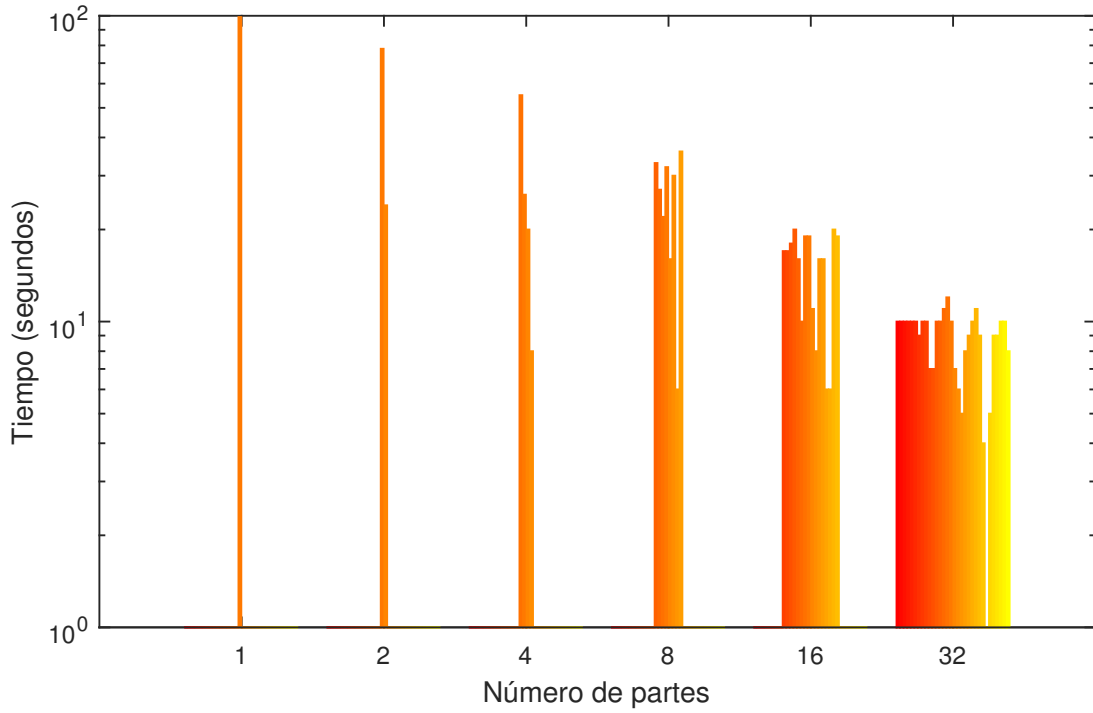


Figura 4.14: Tiempo de cómputo de la grilla versus número de procesos al calcular 3 componentes de 4 usando la configuración 2 para 1, 2, 4, 8, 16 y 32 partes.

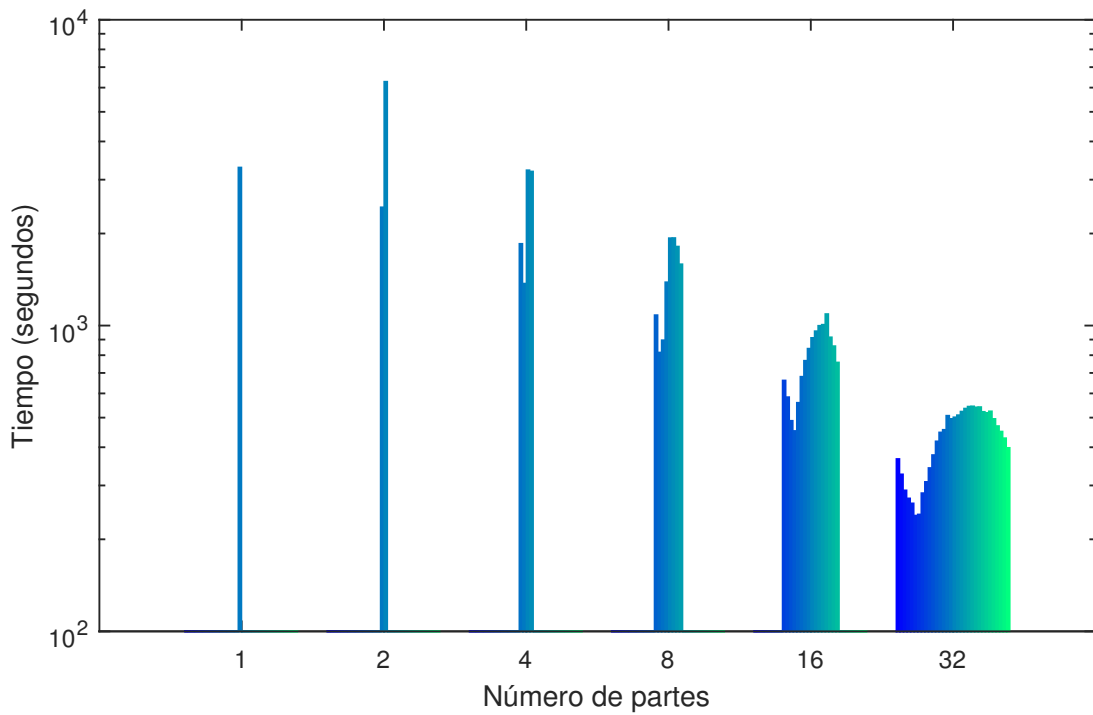


Figura 4.15: Tiempo de cómputo de la grilla versus número de procesos al calcular 4 componentes de 4 usando la configuración 2 para 1, 2, 4, 8, 16 y 32 partes.

En las Figuras 4.14 y 4.15 se muestran los mismos casos presentados en las Figuras 4.12 y 4.13 duplicando el número de procesos en cada paso hasta llegar 32. Aquí se observa como al aumentar el número de procesos cada proceso es dividido en dos. Así, si bien no todas las divisiones son favorables se observa una tendencia favorable llegando a reducir el tiempo total en un orden de magnitud al dividir en 32 procesos.

Basado en estos resultados se observa que el uso de procesos paralelos puede acelerar la ejecución del programa. Sin embargo un leve incremento del número de procesos implica el riesgo de que el incremento en la carga computacional sea mayor al beneficio de repartir esta carga.

Capítulo 5

Conclusiones

5.1. Conclusiones

- En este trabajo se logró desarrollar un método para calcular una solución global de un sistema compuesto por componentes sinusoidales el cual puede ser aplicado a la búsqueda de exoplanetas.
- Se logró implementar el mismo método de manera de obtener resultados en tiempos razonables, además de indicar algunas consideraciones que pueden ayudar a obtener mejor rendimiento. Con esto es posible que el método sea adoptado para realizar análisis por otros investigadores.
- Se mostró cómo el algoritmo puede ser utilizado para la estimación de fase y cómo este análisis puede ser utilizado para distinguir planetas de manchas solares.
- Se analizó cómo el método reacciona ante la presencia de ruido en la señal, indicando las limitaciones actuales del método.

5.2. Trabajo futuro

Los datos astronómicos vienen acompañados con una medida del error de medición para cada dato, el cual indica que datos son más confiables. Adicionalmente, los dos códigos que componen el método fueron implementados con la opción de utilizar un peso para determinar cuánto impacta cada dato el valor del MMSE. Queda pendiente el desarrollar un esquema para definir el valor de los pesos que sea apropiado, para darle

el mejor uso a la información disponible.

Los códigos implementados asumen que el valor medio de la señal medida y real es igual. Este supuesto es válido si el patrón de muestreo es uniformemente distribuido, sin embargo los patrones de muestreo real tienen patrones anuales y mensuales. Esto implica que si un exoplaneta tiene un periodo múltiplo del ciclo terrestre o lunar el valor medio será distinto. Para abordar este inconveniente es posible agregar una componente de estimación del valor medio. Esto puede realizarse para los códigos del periodograma la grilla o ambos y se propone estudiar el mejor esquema para implementarlo.

El uso de sinusoides implica asumir que los planetas tienen órbitas circulares. Esto es válido en múltiples casos, sin embargo ser capaz de detectar objetos con gran excentricidad permitiría ampliar el alcance de la técnica.

Finalmente se propone evaluar métodos para optimizar la técnica en el caso de ruido correlacionado.

Bibliografía

- [1] J. T. Wright, O. Fakhouri, G. W. Marcy, E. Han, Y. Feng, J. A. Johnson, A. W. Howard, D. A. Fischer, J. A. Valenti, J. Anderson y N. Piskunov, «The exoplanet orbit database», *Publications of the Astronomical Society of the Pacific*, vol. 123, n.º 902, págs. 412-422, 2011.
- [2] *Exoplanet orbit database [En línea]*, <http://exoplanets.org/>, Último acceso: 24 Agosto 2015.
- [3] J. Rameau, G. Chauvin, A.-M. Lagrange, A. Boccaletti, S. Quanz, M. Bonnefoy, J. Girard, P. Delorme, S. Desidera, H. Klahr y col., «Discovery of a probable 4-5 Jupiter-mass exoplanet to HD 95086 by direct imaging», *The Astrophysical Journal Letters*, vol. 772, n.º 2, pág. L15, 2013.
- [4] G. Chauvin, A.-M. Lagrange, H. Beust, M. Bonnefoy, A. Boccaletti, D. Apai, F. Allard, D. Ehrenreich, J. Girard, D. Mouillet y col., «Orbital characterization of the β Pictoris b giant planet», *Astronomy & Astrophysics*, vol. 542, A41, 2012.
- [5] P. Kalas, J. R. Graham, E. Chiang, M. P. Fitzgerald, M. Clampin, E. S. Kite, K. Stapelfeldt, C. Marois y J. Krist, «Optical images of an exosolar planet 25 light-years from Earth», *Science*, vol. 322, n.º 5906, págs. 1345-1348, 2008.
- [6] K. Mandel y E. Agol, «Analytic light curves for planetary transit searches», *The Astrophysical Journal Letters*, vol. 580, n.º 2, pág. L171, 2002.
- [7] T. G. Beatty y S. Seager, «Transit probabilities for stars with stellar inclination constraints», *The Astrophysical Journal*, vol. 712, n.º 2, pág. 1433, 2010.
- [8] E. Sedaghati, H. Boffin, S. Csizmadia, N. Gibson, P. Kabath, M. Mallonn y M. Van den Ancker, «Regaining the FORS: Optical ground-based transmission spectroscopy of the exoplanet WASP-19b with VLT+ FORS2», *Astronomy & Astrophysics*, vol. 576, pág. L11, 2015.
- [9] E. Agol, J. Steffen, W. Clarkson y col., «On detecting terrestrial planets with timing of giant planet transits», *Monthly Notices of the Royal Astronomical Society*, vol. 359, n.º 2, págs. 567-579, 2005.

- [10] M. Mayor, F. Pepe, D. Queloz, F. Bouchy, G. Rupperecht, G. Lo Curto, G. Avila, W. Benz, J.-L. Bertaux, X. Bonfils y col., «Setting new standards with HARPS», *The Messenger*, vol. 114, pág. 20, 2003.
- [11] *W. M. Keck Observatory [En línea]*, <http://www.keckobservatory.org/>, Último acceso: 24 Marzo 2015.
- [12] R. I. Dawson y D. Fabrycky, «Radial velocity planets de-aliased», en *Bulletin of the American Astronomical Society*, vol. 41, 2010, pág. 838.
- [13] C. E. Shannon, «Communication in the presence of noise», *Proceedings of the IRE*, vol. 37, n.º 1, págs. 10-21, 1949.
- [14] A. P. Hatzes, «The radial velocity detection of Earth-mass planets in the presence of activity noise: The case of α Centauri Bb», *The Astrophysical Journal*, vol. 770, n.º 2, pág. 133, 2013.
- [15] N. R. Lomb, «Least-squares frequency analysis of unequally spaced data», *Astrophysics and space science*, vol. 39, n.º 2, págs. 447-462, 1976.
- [16] J. D. Scargle, «Studies in astronomical time series analysis. II-Statistical aspects of spectral analysis of unevenly spaced data», *The Astrophysical Journal*, vol. 263, págs. 835-853, 1982.
- [17] M. Tuomi y J. S. Jenkins, «Counting the number of planets around GJ581. false positive rate of bayesian signal detection methods», *ArXiv preprint arXiv:1211.1280*, 2012.
- [18] M. Zechmeister y M. Kürster, «The generalised Lomb-Scargle periodogram-A new formalism for the floating-mean and Keplerian periodograms», *Astronomy & Astrophysics*, vol. 496, n.º 2, págs. 577-584, 2009.
- [19] D. Queloz, F. Bouchy, C. Moutou, A. Hatzes, G. Hébrard, R. Alonso, M. Auvergne, A. Baglin, M. Barbieri, P. Barge y col., «The CoRoT-7 planetary system: Two orbiting super-Earths», *Astronomy & Astrophysics*, vol. 506, n.º 1, págs. 303-319, 2009.
- [20] R. V. Baluev, «Detecting multiple periodicities in observational data with the multifrequency periodogram—I. Analytic assessment of the statistical significance», *Monthly Notices of the Royal Astronomical Society*, stt1617, 2013.
- [21] —, «Detecting multiple periodicities in observational data with the multifrequency periodogram—II. Frequency Decomposer, a parallelized time-series analysis algorithm», *Astronomy and Computing*, vol. 3, págs. 50-57, 2013.

- [22] G. Anglada-Escudé, M. Tuomi, E. Gerlach, R. Barnes, R. Heller, J. S. Jenkins, S. Wende, S. S. Vogt, R. P. Butler, A. Reiners y col., «A dynamically-packed planetary system around gj 667c with three super-earths in its habitable zone», *Astronomy & Astrophysics*, vol. 556, A126, 2013.
- [23] M. Tuomi, «A new cold sub-Saturnian candidate planet orbiting GJ 221», *Monthly Notices of the Royal Astronomical Society: Letters*, slu014, 2014.
- [24] J. S. Jenkins y M. Tuomi, «The curious case of HD 41248. A pair of static signals buried behind red noise», *The Astrophysical Journal*, vol. 794, n.º 2, pág. 110, 2014.
- [25] H. Haario, E. Saksman y J. Tamminen, «An adaptive metropolis algorithm», *Bernoulli*, págs. 223-242, 2001.
- [26] H. Wold, *A study in the analysis of stationary time series*. Almqvist & Wiksell Book Co, 1938.
- [27] J. D. Scargle, «Studies in astronomical time series analysis. I-Modeling random processes in the time domain», *The Astrophysical Journal Supplement Series*, vol. 45, págs. 1-71, 1981.
- [28] J. S. Jenkins, N. B. Yoma, P. Rojo, R. Mahu y J. Wuth, «Improved signal detection algorithms for unevenly sampled data. Six signals in the radial velocity data for GJ876», *Monthly Notices of the Royal Astronomical Society*, vol. 441, n.º 3, págs. 2253-2265, 2014.
- [29] R. Kuschnig, W. Weiss, R. Gruber, P. Bely y H. Jenkner, «Microvariability survey with the hubble space telescope fine guidance sensors. Exploring the instrumental properties», *Astronomy and Astrophysics*, vol. 328, págs. 544-550, 1997.