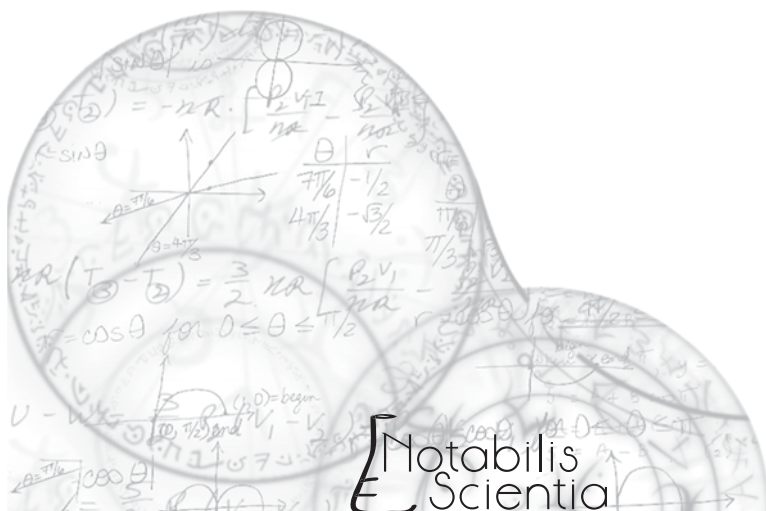








## PROCESOS NO LINEALES EN LA CIENCIA Y LA SOCIEDAD





# PROCESOS NO LINEALES EN LA CIENCIA Y LA SOCIEDAD

Máximo A. Agüero Granados  
[Coordinador]





Dr. en D. Jorge Olvera García  
*Rector*

Dr. en E. Alfredo Barrera Baca  
*Secretario de Docencia*

Dra. en E. L. Ángeles Ma. del Rosario Pérez Bernal  
*Secretaria de Investigación y Estudios Avanzados*

M. en D. José Benjamín Bernal Suárez  
*Secretario de Rectoría*

M. en E. P. y D. Ivett Tinoco García  
*Secretaria de Difusión Cultural*

M. en C. I. Ricardo Joya Cepeda  
*Secretario de Extensión y Vinculación*

M. en E. Javier González Martínez  
*Secretario de Administración*

Dr. en C. P. Manuel Hernández Luna  
*Secretario de Planeación y Desarrollo Institucional*

M. en A. E. D. Yolanda E. Ballesteros Senties  
*Secretaria de Cooperación Internacional*

Dr. en D. Hiram Raúl Piña Libien  
*Abogado General*

Lic. Juan Portilla Estrada  
*Director General de Comunicación Universitaria*

M. en A. Ignacio Gutiérrez Padilla  
*Contralor Universitario*

Profr. Inocente Peñaloza García  
*Cronista*

Mtro. Miguel Ángel López Díaz  
*Director de la Facultad de Ciencias*

M. en H. Blanca Aurora Mondragón Espinoza  
*Directora de Difusión y Promoción de la Investigación y los Estudios Avanzados*

Primera edición: septiembre de 2013

ISBN: 978-607-8289-41-7

D.R. © Universidad Autónoma del Estado de México  
Instituto Literario núm. 100 Ote.  
Colonia Centro  
C.P. 50000. Toluca, México  
<http://www.uaemex.mx>

D. R. © Ediciones y Gráficos Eón, S.A. de C.V.  
Av. México-Coyoacán núm. 421  
Colonia Xoco, Delegación Benito Juárez  
C.P. 03330, México, Distrito Federal  
Tels.: 56 04 12 04 y 56 88 91 12  
[administracion@edicioneseon.com.mx](mailto:administracion@edicioneseon.com.mx)  
[www.edicioneseon.com.mx](http://www.edicioneseon.com.mx)

El contenido total de este libro fue sometido a dictamen en el sistema de pares ciegos.

Prohibida la reproducción total o parcial por cualquier medio, sin la autorización escrita del titular de los derechos patrimoniales.

Impreso y hecho en México  
*Printed and made in Mexico*

# ÍNDICE

<b>Presentación</b>	13
<b>Enigmas de efecto túnel no lineal para solitones ópticos y de ondas de materia</b>	17
<i>Tatyana Belyaeva y Vladimir Serkin</i>	
Introducción	18
I. Efecto túnel para solitones	22
II. Tunelaje solitónico no lineal a través de los potenciales externos: el papel de la energía de enlace solitón	23
III. Tunelaje de los solitones auto-compresivos	30
Conclusiones	35
Referencias	35
<b>Amplificación de pulsos ópticos solitónicos</b>	41
<i>Máximo A. Agüero, Tatyana Belyaeva y Vladimir Serkin</i>	
Parte I	42
Introducción	42
I. Amplificación de solitones en pulsos multisolitónicos	44
II. Amplificación singular de solitones con frecuencia modulada	51
Parte II	59
III. Solitones de color	59
IV. Dinámica no lineal de solitones de la ecuación no lineal de Schrödinger (ENLS) en trampas dependientes del tiempo	62
Conclusiones	68
Referencias	69
<b>El Teorema de Paso de Montaña</b>	73
<i>Alfredo Cano, Fernando A. Ongay Larios y Ma. de Lourdes Nájera-López</i>	
Introducción	73
I. Problemas clásicos de optimización	75
II. Funciones continuas y compacidad	77
III. Mínimos y puntos críticos	78
IV. Existencia de puntos críticos. Teorema de Paso de Montaña	83

V.	El Teorema de Paso de Montaña (TPM) en dimensión infinita	89
	Las sucesiones de Palais-Smale	89
	El caso en dimensión infinita	90
VI.	Aplicaciones del TPM	91
	Referencias	97

### **Inflación cosmológica vista desde la mecánica cuántica supersimétrica**

99

*J. Socorro, Miguel Sabido, W. Guzmán y Máximo A. Agüero*

	Introducción	100
I.	Acción clásica del modelo	105
	a. Esquema clásico	107
	b. Esquema cuántico	109
II.	Formalismo de Witten	111
III.	Soluciones cuánticas supersimétricas para una etapa inflacionaria	112
IV.	Densidad de probabilidad no normalizada $ \Psi ^2$ en las variables de Grassmann	116
	Conclusiones	118
	Referencias	118

### **Estudio dinámico del modelo $\Lambda$ -CDM inhomogéneo**

123

*Roberto A. Sussman y Germán Izquierdo*

	Introducción	124
I.	Espacio-tiempos $\Lambda$ -LTB	125
II.	Escalares cuasilocales (QL)	127
III.	Construcción del sistema dinámico	130
IV.	El espacio de fases, subespacios invariantes y puntos críticos	133
	IV.1 Espacio de fases: caso general	133
	IV.2 Subespacio homogéneo y sus centros de simetría: dinámica del caso FLRW	134
	IV.3 El subespacio espacialmente plano	135
	IV.4 La solución Kottler: vacío Schwarzschild-De Sitter	135
	IV.5 Polvo LTB con $\Lambda = 0$	136
V.	Cinemática de las capas de polvo	137
VI.	Topología de las rebanadas espaciales y elección de coordenada radial	140
VII.	La dinámica del espacio de fases en términos de proyecciones homogéneas e inhomogéneas	141
VIII.	Ejemplos numéricos y gráficos	143
	VIII.1 Configuración con expansión perpetua: patrón cinemático (i)	144

VIII.1.1	Curvatura espacial negativa	144
VIII.1.2	Curvatura espacial positiva con topología abierta	145
VIII.1.3	Curvatura espacial positiva con topología cerrada o de agujero de gusano	147
VIII.2	Configuración en la cual la expansión es revertida o interrumpida para todas las capas de polvo	148
VIII.2.1	Modelos que colapsan: patrón cinemático (ii)	148
VIII.2.2	Modelos con retorno sin colapso: patrón cinemático (iii)	150
VIII.2.3	Modelos de “loitering”: patrón cinemático (iv)	152
VIII.3	Configuraciones con patrones cinemáticos mixtos: patrones cinemáticos (v)	153
VIII.3.1	Formación de estructura	154
VIII.3.2	Modelos de retorno/expansión	156
	Conclusiones	157
	Referencias	160
	<b>Neuronas de integración y disparo</b>	163
	<i>Fernando A. Ongay Larios, Humberto Carrillo Calvet,</i>	
	<i>José A. Aguilar Sánchez, Máximo A. Agüero</i>	
	<i>y Miguel Ángel Mendoza Reyes</i>	
	Oscilaciones no lineales	164
I.	Neuronas de integración y disparo	166
a)	Neurona mecánica	166
II.	Forzamiento y sincronización	168
III.	Modelación de neuronas de integración y disparo	168
a)	El modelo KHR	169
IV.	Funciones de disparo	170
V.	Sincronización y funciones en la circunferencia	171
VI.	Teorías de rotación	172
VII.	Teoremas de regularidad	174
a)	Aplicación al modelo KHR	174
b)	Partición del espacio de parámetros	176
	Referencias	177
	<b>Análisis caótico de las tres variables macroscópicas de tráfico vehicular</b>	179
	<i>Oscar A. Rosas-Jaimes</i>	
	Introducción	179
I.	Diagramas fundamentales	181

II. Modelo de Greenberg	184
III. Análisis de comportamientos no lineales	187
IV. Divergencia de trayectorias cercanas	195
Conclusiones	199
Referencias	201

**Dinámica de los precios forward como proceso estocástico con reversión a la media** 203

*Vladimir G. Makhankov, Ma. de Lourdes Nájera-López*

*y Máximo A. Agüero*

Introducción	204
I. Ecuaciones de la geometría diferencial estocástica	204
II. Precios forward y su modelaje	207
III. Reducción a modelo de un factor	210
IV. Ecuaciones de la geometría diferencial estocástica en el precio forward y volatilidad	212
V. Propiedades dinámicas de las curvas forward	220
Conclusiones	225
Referencias	226

**México, tipo de cambio y crecimiento: un análisis contrafactual, 1950-2010** 227

*Eduardo Loría y Jorge Ramírez*

Introducción	228
I. El papel del tipo de cambio en la economía mexicana	229
II. La restricción externa al crecimiento	232
III. Metodología econométrica	235
III.1 El modelo SVAR	235
III.2 Estimación	236
III.2.1 El modelo base	237
III.2.2 El modelo alternativo	238
Conclusiones y recomendaciones de política	242
Referencias	244
Anexo	246

**Cuantificación de variables para evaluar riesgos de falla en presas de tierra** 249

*David Joaquín Delgado Hernández, David De León Escobedo,*

*Oswaldo Morales Nápoles, Benjamín Pérez Pliego*

*y Juan Carlos Arteaga Arcos*

Introducción	250
I. Conceptos y definiciones: presas	251
II. Conceptos y definiciones: riesgos y confiabilidad	254
III. Juicio estructurado de expertos (JEE)	256
IV. Presas en el Estado de México	260
V. Peligros	262
V.1 Frecuencia sísmica	262
V.2 Precipitación	264
VI. Vulnerabilidad	266
VII. Principales daños en las presas de tierra	266
VII.1. Tubificación	267
VII.2. Desbordamiento	268
VIII. Discusión y conclusiones	270
Referencias	272

**Estudio de los orbitales del electrón en una tapa semiesférica de un nanotubo de carbono usando funcionales de la densidad** 275

*Patricia Rosendo Andrés y Porfirio Rosendo Francisco*

Introducción	276
I. Nanotubos de carbono	278
II. Teoría de funcionales de la densidad (DFT)	281
II.1 Método de combinación lineal de orbitales atómicos (LCAO)	283
II.2 Teoría de orbitales moleculares (OM)	283
III. Resultados	286
Conclusiones	301
Referencias	302

**Algoritmos genéticos aplicados a la optimización de centrales de ciclo combinado y calderas de recuperación de calor** 305

*María Dolores Durán*

Introducción	306
I. Generalidades sobre los ciclos combinados de turbinas de gas y turbinas de vapor	308
I.1 Generalidades de una planta de ciclo combinado	308
I.1.1 Caldera de recuperación de calor (CRC)	310
II. Descripción de la metodología empleada	312
II.1 Generalidades de los algoritmos genéticos	312

II.2 Descripción del algoritmo de optimización empleado	313
III. Descripción del modelo de optimización termoeconómico	314
III.1 Modelo de costo y modelo de flujo anual de caja	314
III.2 Comparación de los modelos de flujo de caja y costo de generación	315
III.3 Modelo de atribución de costos	317
III.3.1 Costo total de combustible	317
III.3.2 Costo de amortización	318
a) Costo de la turbina de gas	318
b) Costo de la turbina de vapor	319
c) Costo de la caldera de recuperación de calor	320
III.3.3 Costo de operación y mantenimiento	321
III.4 Optimización del ciclo combinado de dos niveles de presión con recalentamiento	322
III.4.1 Parámetros de diseño del ciclo de dos niveles de presión con recalentamiento	325
III.4.2 Tendencias de los valores de la presión en los niveles de alta y de baja	325
IV. Descripción de la metodología para el cálculo geométrico de la caldera de recuperación de calor	327
IV.1 Descripción de los parámetros geométricos de la CRC	328
IV.1.1 Restricciones para el diseño de la CRC	330
IV.2 Modelo matemático para el diseño de la CRC	331
IV.2.1 Definición del modelo	331
IV.2.2 Modelo termodinámico para obtener el coeficiente global de transferencia de calor	332
IV.3 Herramienta para el cálculo de las secciones de la CRC	334
IV.3.1 Variables de diseño	334
IV.3.2 Función objetivo del algoritmo genético (función salud)	335
IV.3.2.1 Función objetivo	336
IV.3.2.2 Penalizaciones	336
IV.4 Aplicación del algoritmo para el cálculo geométrico de una caldera de dos niveles de presión con recalentamiento	337
Conclusiones y comentarios finales	340
Referencias	340

## PRESENTACIÓN

La evolución de la sociedad en gran medida está soportada por el avance del desarrollo científico. Por ende, la aplicación de conocimientos obtenidos en esfuerzos indagatorios sobre fenómenos que ocurren en el universo es una de las premisas sobre la cual se basa todo desarrollo tecnológico y científico de la humanidad. La investigación en torno a las diferentes manifestaciones de la naturaleza conduce al análisis de entidades interactuantes que condicionan el comportamiento *a posteriori* de los elementos que la conforman. Es así que el estudio de sistemas interactuantes constituye un tema central en cualquier disciplina científica. Sin embargo, el análisis matemático de estos sistemas no es un problema fácil, de hecho, el estado actual del conocimiento científico y matemático dista aún mucho de comprender a detalle todos los posibles comportamientos de los sistemas complejos. El problema radica principalmente en dos factores: el primero se refiere a la gran cantidad de componentes, variables y parámetros que pueden estar interactuando, lo que obliga a analizar simplificaciones del problema; y el segundo concierne a que las interacciones generalmente no son lineales. Una de las diferencias fundamentales entre procesos lineales y no lineales es que en estos últimos no se cumple el principio de superposición. Cuando el vasto número de procesos naturales no lineales se estudia, comúnmente se modela mediante sistemas de ecuaciones no lineales. Suele ocurrir frecuentemente que la complejidad de fenómenos que ocurren en la naturaleza y sociedad en muchos aspectos es descrita y analizada incluso bajo una misma ecuación diferencial no lineal, incluyendo también en muchos casos la influencia aleatoria de otros agentes externos, por lo que la estocasticidad estará presente de una u otra manera. Esto crea un cuadro relativamente exitoso en vista de que las investigaciones de procesos que ocurren se pueden llevar a cabo en forma aislada, analizarlas abstractamente en el mundo de las matemáticas, para luego volcar todo el arsenal de conocimientos en las respectivas áreas de interés que se realizan.

Este volumen se crea a partir de contribuciones científicas de un sector de profesores-investigadores pertenecientes al Sistema Nacional de Investigadores (SNI) de la Universidad Autónoma del Estado de México (UAEMéx), con sus respectivos coautores. Los autores están principalmente abocados a estudiar la naturaleza de los procesos complejos no lineales que se manifiestan permanentemente en el entorno, reflejando así la complejidad de los fenómenos que se pueden estudiar satisfactoriamente dentro de paradigmas ya establecidos con los métodos

matemáticos. Los manuscritos han sido enviados a comité *ad hoc* pasando por un proceso de arbitraje riguroso correspondiente. Este esfuerzo editorial se realiza bajo los auspicios de la Secretaría de Investigación y Estudios Avanzados (SIEA) de la UAEMéx, con objeto de fortalecer la presencia de los profesores-investigadores de la institución en el SNI. Semejante actividad podría incrementar el número de profesores-investigadores y además ensancharlo de tal manera que pueda permitir la existencia de un núcleo académico científico sólido para garantizar y asegurar los futuros logros académico-científicos en la UAEMéx.

Al principio de la realización de esta obra, la idea fue hacerla con base en colaboraciones propias de las diversas áreas modernas de física-matemática. Pero esto nos cerraba el camino para aceptar contribuciones importantes de otras ramas de la ciencia ligadas al mundo complejo y no lineal. Entonces decidimos ampliar el abanico de recepción de manuscritos, tomando en cuenta la integridad compleja y no lineal del tema que se abordaba en los trabajos. Principalmente los tópicos del libro abarcan problemas de ondas no lineales, cosmología, nanoestructuras, problemas variacionales, econofísica, economía e ingeniería. El nivel matemático del texto supone un buen conocimiento de las matemáticas impartidas en las universidades, como, por ejemplo, en lo referente a ecuaciones diferenciales ordinarias y diferenciales y en tópicos de problemas estocásticos y estadísticos.

Veamos en forma sucinta el contenido de los capítulos del libro que se publican en este trabajo.

La existencia de ondas ocurre como una de las formas de manifestación del movimiento más común en el mundo que nos rodea. Aunque las ondas lineales juegan un papel crucial en el desarrollo de la ciencia en primera aproximación, las ondas de índole no lineal se consideran como las más “realistas” para describir procesos complejos. Una gran posibilidad para el desarrollo de teoría de ondas no lineales se abrió debido a la investigación de solitones, ondas que no cambian de forma y de energía a lo largo de su desplazamiento.

En este contexto, en el capítulo de Tatyana Belyaeva y Vladimir Serkin se estudia el efecto túnel cuántico usando ondas no lineales solitónicas de óptica y ondas de materia. Como es conocido, la ecuación diferencial que gobierna estos pulsos ópticos es de la familia de ecuaciones no lineales de Schrödinger, por su analogía con la famosa ecuación de la mecánica cuántica.

Seguidamente, en el capítulo de Máximo A. Agüero, Tatyana Belyaeva y Vladimir Serkin se discute la posibilidad de obtener amplificación de solitones de manera analítica, considerando la dispersión Raman estimulada al influenciar en la energía de enlace del solitón. La ganancia requerida se puede realizar experimentalmente mediante la fabricación de un amplificador de fibra de dos fotones, que excluye los efectos relacionados con la difracción y el auto-enfocamiento de la radiación.

Posteriormente, Alfredo Cano, Fernando A. Ongay Larios y Ma. de Lourdes Nájera-López analizan la aplicación del Teorema de Paso de Montaña para problemas de optimización, usando sucesiones minimizantes, que podría tener implicaciones técnico-científicas de fenómenos no lineales. Además, se muestra una técnica para resolver la existencia de soluciones de una ecuación diferencial elíptica.

En el siguiente capítulo se aborda el problema de la inflación cosmológica que sigue causando discusiones importantes en el ambiente científico sobre el cosmos. Así, J. Socorro, Miguel Sabido, W. Guzmán y Máximo A. Agüero, en su correspondiente capítulo, discuten la posibilidad de definir los potenciales al momento de ocurrir la inflación, usando elementos supersimétricos y la mecánica cuántica. Se predice que existen múltiples potenciales físicamente aceptables para esta etapa inflacionaria.

Roberto A. Sussman y Germán Izquierdo discuten desde el punto de vista de sistemas dinámicos, modelos de espacio tiempo esféricamente simétricos pero no homogéneos. Ellos encuentran que las trayectorias de fase evolucionan de un atractor común en el pasado y van a parar en otro atractor futuro como un fenómeno determinista de la evolución cósmica.

Seguidamente, Fernando A. Ongay Larios, Humberto Carrillo Calvet, José A. Aguilar Sánchez, Máximo A. Agüero y Miguel Ángel Mendoza Reyes, en la misma línea del método de sistemas dinámicos muestran un análisis en la modelación de neuronas mediante el ejemplo de un oscilador que recibe un estímulo periódico a partir del cual se puede determinar si hay condiciones para que la respuesta esté sincronizada con el estímulo, o si dispara un impulso con implicaciones al sistema nervioso de los seres vivos.

Posteriormente, Oscar A. Rosas-Jaimes nos muestra un análisis sobre el comportamiento del tráfico como un sistema macroscópico no lineal. Este modelo se basa en los denominados diagramas fundamentales que relacionan a las variables por pares, y, como se muestra, se observan *a posteriori* comportamientos caóticos inesperados.

En el área de econofísica, Vladimir G. Makhankov, Ma. de Lourdes Nájera-López y Máximo A. Agüero abordan un problema crucial de la evolución de tasas forward. Estos modelos surgen usando las ideas y ecuaciones de la geometría diferencial estocástica para las tasas forward y sus volatilidades. Se usan métodos de Ito y Stratonovich para determinar las volatilidades en forma analítica.

Siguiendo con los temas económicos, vemos a continuación la investigación de Eduardo Loría y Jorge Ramírez, que se centra en demostrar la importancia crucial del tipo de cambio real para explicar fenómenos de lento crecimiento económico en países endémicos. En economía se considera de alto valor que los países tengan un tipo de cambio estable.

Como ejemplo de fenómenos no lineales en aéreas de geografía, David Joaquín Delgado Hernández, David De León Escobedo, Oswaldo Morales Nápoles, Benjamín Pérez Pliego y Juan Carlos Arteaga Arcos discuten la forma de cuantificar las variables en los daños generados por las fallas en presas construidas en tierra. Esta actividad, en caso de ocurrencia de desastres naturales, podría ayudar en los planes de evacuación y mantenimiento de las presas.

Después, en el capítulo de Patricia Rosendo Andrés y Porfirio Rosendo Francisco se muestra el estudio de la capa electrónica en nanotubos de carbono y sus principales características de la tapa semiesférica para nanotubos zigzags, sobre parámetros energéticos y geométricos. Estas propiedades últimamente han sido objeto de intensa labor de investigación por sus aplicaciones prácticas.

Cierra nuestra obra la contribución de María Dolores Durán sobre algoritmos para optimizar centrales y calderas de calor. Estos procesos pueden analizarse tratándose como problemas de optimización sobre qué diseño presentado para calderas de recuperación de calor sería el óptimo. En el trabajo se presenta una metodología para resolver varios casos.

Finalmente, no se puede dejar de mencionar que en todas estas investigaciones se han involucrado estudiantes en proceso de graduación que están contribuyendo en las investigaciones de sus respectivas aéreas. Se puede decir que esta labor es un tanto difícil de realizar porque requiere que los investigadores pongan más atención al trabajo de asesorías y tutorías. Pero pensamos que es un camino de esperanza para elevar el número y la calidad de científicos en ciencias naturales exactas y sociales para el bien de la sociedad.

Estamos muy agradecidos con todos los autores por su trabajo. Aunque algunos de nuestros requerimientos pudieron haberse mostrado como tediosos o preocupantes, los autores siempre actuaron positivamente para afrontar estas dificultades y eso contribuyó a que nuestra tarea haya sido más llevadera.

Expresamos nuestra sincera gratitud a la Secretaria de Investigación y Estudios Avanzados de la UAEMéx, Ángeles Ma. del Rosario Pérez Bernal, por su invaluable ayuda para la publicación de este material científico que refleja los distintos esfuerzos que realiza una parte del núcleo de investigadores de la UAEMéx, en colaboración con sus colegas de otras instituciones del país y del extranjero. Asimismo, sin el apoyo eficiente de la asistente de investigación, M. en I., Ma. de Lourdes Nájera-López, en su ardua labor de compilación, corrección y redacción de los materiales enviados, el trabajo sobre la culminación del libro no podría haberse llevado a cabo.

Toluca, octubre de 2013  
**Máximo A. Agüero Granados**  
(Coordinador)

# ENIGMAS DE EFECTO TÚNEL NO LINEAL PARA SOLITONES ÓPTICOS Y DE ONDAS DE MATERIA\*

*Tatyana Belyaeva\*\**

*Vladimir Serkin\*\*\**

## **Resumen**

Estudios recientes han arrojado dudas sobre el tratamiento de tunelaje en solitones. Los enigmas más importantes en este campo se pueden formular de la siguiente manera: ¿cuándo el efecto no lineal de tunelaje de solitones se parecerá al de una partícula puntual, como en el caso de partículas clásicas?, ¿o seguirá el comportamiento de la mecánica cuántica en el que la partícula en sí tiene una estructura interna? ¿Cómo los grados de libertad “ocultos” pueden aparecer en el proceso de tunelaje de un solitón? ¿Qué ocurre si la amplitud y duración inicial del solitón varían en el tiempo y el solitón se aproxima a una barrera clásicamente prohibida? En particular, ¿qué sucede en el caso de tunelaje no lineal de un solitón auto-compresivo cuando su energía de enlace se incrementa? ¿Este caso se parecerá más el caso de partícula clásica o el comportamiento de la mecánica cuántica? Estas cuestiones son consideradas en este capítulo.

**Palabras clave:** Efecto túnel solitónico, comportamiento ondulatorio y de partícula de solitones, Ecuación No Lineal de Schrödinger Generalizado.

\* Trabajo parcialmente financiado por la Secretaría de Educación Pública con el proyecto FE01/2012 103.5/12/2126 de redes multidisciplinarias (Promep).

\*\* Doctora en Física. Profesora de Tiempo Completo, Facultad de Ciencias, Universidad Autónoma del Estado de México, Instituto Literario 100, C.P. 50000, Toluca, Estado de México, México. (52) (722) 2965556, <betale01@yahoo.com>.

\*\*\* Doctor en Física. Profesor de Tiempo Completo, Benemérita Universidad Autónoma de Puebla, C.P. 72502, 72001, Puebla, México.

## INTRODUCCIÓN

El efecto túnel cuántico ha sido uno de los conceptos más fructíferos de la física moderna. Este descubrimiento se produjo en un momento en que la nueva mecánica cuántica sólo había estado disfrutando de sus primeros pasos en la explicación de la estructura atómica. En los años transcurridos desde su descubrimiento por Gamow (1928a y 1928b) y Gurney y Condon (1928 y 1929), el concepto de túnel cuántico ha demostrado ser de gran importancia en muchas ramas de la física, tanto desde el punto de vista fundamental como desde el punto de vista de aplicaciones, tal como el microscopio de efecto túnel. Mientras que el estudio de efecto túnel cuántico lineal tiene una larga historia que se remonta a los primeros días de la mecánica cuántica, el campo de efecto túnel cuántico no lineal es relativamente joven y, como hecho, nació sólo después de dos descubrimientos, solitones y el condensado de Bose-Einstein (BEC).

El descubrimiento del BEC en nubes alcalinas de átomos ultrafríos confinados en trampas magnéticas (Cornel y Wieman, 2002) abrió posibilidades únicas para poner a prueba la naturaleza ondulatoria de la materia e investigar los fenómenos cuánticos a escala macroscópica.

El concepto clásico de solitón fue introducido por Zabusky y Kruskal (1965) para caracterizar las ondas solitarias no lineales que no se dispersan y preservan su identidad durante la propagación y después de una colisión. El solitón óptico en las fibras de telecomunicaciones presenta un bello ejemplo en el que un concepto matemático abstracto ha producido un gran impacto en el mundo real de altas tecnologías (Hasegawa, 1989; Hasegawa y Kodama, 1995; Hasegawa y Matsumoto, 2003; Agrawal, 2001; Akhmediev y Ankiewicz, 1997). Cabe destacar que la complementariedad de los tres descubrimientos fundamentales (tunelaje, solitones y el BEC) jugará un papel clave en el desarrollo de futuras aplicaciones tecnológicas de BEC, donde muchas ideas se beneficiarán mutuamente.

Los solitones se comportan como partículas clásicas hasta que su fase comienza a influir en su comportamiento. El efecto de la fase aparece más prominente en interacciones solitón-solitón y en la interacción de solitones con un potencial. Un potencial de confinamiento proporciona nuevas propiedades a la dinámica de solitones y la interacción entre solitones. Por ejemplo, se establece que la diferencia de fase entre solitones vecinos de un BEC atractivo es responsable de la repulsión observada entre solitones (Strecker *et al.*, 2002 y 2003; Khaykovich *et al.*, 2002).

La fase juega un papel importante en la dinámica de solitones en los distintos sistemas físicos no autónomos con diferente dispersión, la no linealidad y la ganancia o absorción (Serkin y Hasegawa, 2000). La formación de ondas solitarias que ajustan magnéticamente la interacción interatómica cerca de una resonancia

de Feshbach es un buen ejemplo de un sistema no autónomo (Strecker *et al.*, 2002 y 2003; Khaykovich *et al.*, 2002).

Los términos científicos son definidos de acuerdo con su primera aparición. El término “tunelaje no lineal de solitón” (NST) se introdujo por primera vez por Newell (Newell, 1978). La propagación de un solitón hacia una barrera de potencial finita (que sólo depende de la coordenada espacial) fue considerada en este trabajo, en el marco del modelo de ecuación no lineal de Schrödinger (NLSE) con barreras lineal y parabólica. Se encontró que en ciertas circunstancias, dependiendo de la relación de amplitud del solitón y la altura de la barrera, el solitón puede atravesar la barrera de una manera sin pérdidas.

Desde el primer descubrimiento experimental de efecto túnel cuántico macroscópico (MQT) en uniones de Josephson (“Josephson junctions”) de Nb por Voss y Webb (Voss y Webb, 1981), este fenómeno ha sido generalmente asociado con el efecto túnel entre los diferentes estados del sistema. Efectos similares han sido descubiertos en sistemas físicos completamente diferentes, tales como helio líquido y nanomagnetos (Tomsovic, 1998; Friedman y Han, 2002; Ankerhold, 2007). Los BEC presentan diferentes fenómenos de tunelaje (Kagan *et al.*, 1996; Shuryak, 1996; Stoof y Stat, 1997; Ueda y Leggett, 1998; Sackett *et al.*, 1998; Anderson y Kasevich, 1999; Salasnich *et al.*, 2001; Moiseyev *et al.*, 2004; Carr *et al.*, 2005; Fleurov y Soffer, 2005; Lee y Brand, 2006; Dekel *et al.*, 2007 y 2009; Khomeriki *et al.*, 2007; Ahufinger *et al.*, 2007). Se ha demostrado que el tunelaje macroscópico cuántico de los BEC es el soterramiento de una función de onda de muchos cuerpos a través de una barrera de potencial, y por lo tanto este efecto proporciona una prueba más rigurosa de la validez de la mecánica cuántica que el caso de una partícula (Carr *et al.*, 2005; Dekel *et al.*, 2007). Este efecto se conoce como efecto túnel cuántico colectivo (CQT) (Khomeriki *et al.*, 2007). Esto se contrasta con tunelaje del condensado entero en un espacio de parámetros variacionales, como se consideró anteriormente en el contexto de colapso de un BEC meta-estable atractivo en tres dimensiones (Kagan *et al.*, 1996).

En su experimento pionero, Anderson y Kasevich (1999) han investigado la dinámica de un BEC que fue acelerado por la gravedad en el potencial periódico formado por dos rayos láser de la propagación vertical contraria. De esta manera, las oscilaciones de Bloch en el condensado fueron inducidas, y cada vez que un punto de inflexión de la oscilación se alcanzó, una fracción de los átomos fue tunelada a una banda continua de Bloch. La salida normal de pulsos atómicos espectacularmente demostró la coherencia macroscópica del condensado inicialmente preparado. En este experimento la influencia de las interacciones atómicas se presentó como la degradación de la interferencia cuando fueron estudiados condensados con altas densidades.

La observación directa de efecto túnel y auto-captura no lineal en una sola unión bosónica de Josephson (“Josephson junctions”), ejecutado por dos BEC débilmente vinculados en un potencial de doble pozo, se informó por Albiez *et al.* (2005). El problema de cómo el efecto túnel se ve afectado por la interacción interatómica fue analizado por varios investigadores (Smerzi *et al.*, 1997; Milburn *et al.*, 1997; Zapata *et al.*, 1998; Raghavan *et al.*, 1999; Salasnich *et al.*, 1999; Cataliotti *et al.*, 2001 y Morsch *et al.*, 2001).

En particular, se ha demostrado que la interacción interatómica tiende a suprimir el tunelaje del condensado en un potencial de doble pozo (Smerzi *et al.*, 1997; Milburn *et al.*, 1997; Zapata *et al.*, 1998; Raghavan *et al.*, 1999; Salasnich *et al.*, 1999; Cataliotti *et al.*, 2001; Morsch *et al.*, 2001). Este efecto puede ser tan fuerte que, bajo ciertas condiciones, el condensado se mantiene auto-atrapado en uno de los mínimos del potencial. La única diferencia con respecto a uniones de Josephson de superconducción es que, en caso de desequilibrio inicial grande, el condensado es principalmente atrapado en uno de los pozos, produciendo lo que se llama auto-atrapamiento macroscópico (MST) (Smerzi *et al.*, 1997; Milburn *et al.*, 1997; Zapata *et al.*, 1998; Raghavan *et al.*, 1999; Salasnich *et al.*, 1999; Cataliotti *et al.*, 2001; Morsch *et al.*, 2001). Efectos resonantes de túnel de los BEC en redes ópticas y, en particular, los efectos de túnel resonante en los potenciales periódicos han sido investigados por Zemesini *et al.* (2008 y 2009).

El sistema específico del condensado que cae por gravedad y se dispersa en una barrera de potencial Gaussiana fue propuesto en los artículos de Salasnich *et al.* (1999, 2001 y 2002). Se demostró que el tunelaje cuántico macroscópico de un BEC en forma de cigarro confinado por dos barreras Gaussianas con cambios periódicos a la altura abre una novedosa opción para generar las ondas solitónicas periódicas en láseres atómicos. Solitones brillantes de ondas de materia son los objetos cuánticos macroscópicos que pueden actuar como objetos clásicos, parecidos a partículas, manteniendo su integridad durante las colisiones o cuando se someten a fuerzas externas. Lee y Brand (2006) también predijeron una mayor reflexión de ondas de materia tipo solitones brillantes de un potencial puramente atractivo. Diferentes sistemas para el control de la dinámica de ondas solitónicas de materia enrejadas, así como las posibilidades extraordinarias para crear conmutadores cuánticos y memorias cuánticas para solitones enrejados de materia (basado en los efectos de túnel y confinación de solitones), fueron presentados por primera vez por Ahufinger *et al.* (2007). En este artículo la naturaleza cuántica de las ondas solitónicas de materia enrejadas se manifiesta de forma explícita en la aparición de la reflexión sobre barrera y tunelaje.

Dekel *et al.* (2007) investigaron fenómenos dinámicos no lineales de tunelaje macroscópico y han demostrado que la similitud entre dos objetos físicos, los soli-

tones ópticos y solitones de ondas de materia, ambos descritos por las ecuaciones matemáticas similares, ha hecho posible el estudio de los dos sistemas en paralelo y arrojado luz sobre un sistema mediante la realización de experimentos sobre la otra. Un estudio analítico de la dinámica macroscópica de tunelaje dentro de la representación hidrodinámica ha sido presentado también en el artículo de Dekel *et al.* (2007).

Debido a la complejidad evidente de los experimentos con solitones de ondas de materia, Demidov *et al.* (2008) demostraron recientemente que las ondas de espín son especialmente atractivas para estudios experimentales del efecto túnel de solitones, a través de una barrera o un pozo de potencial. Como ejemplo notable, Demidov *et al.* (2008) descubrieron experimentalmente que la interacción de solitones magnéticos con los potenciales externos difiere significativamente de la de los paquetes de ondas lineales. En particular, los solitones demostraron un tunelaje mejorado a través de una barrera de potencial y una reflexión mejorada desde un pozo de potencial.

Otro ejemplo es el artículo reciente de Barak y colaboradores (2008), donde efectos no lineales de tunelaje solitónico y las eyecciones de solitones ópticos fueron descubiertos experimentalmente. En este estudio un paquete de ondas se inició dentro de una trampa de potencial. Al aumentar la potencia del paquete de ondas la dinámica fue transformada de tunelaje lineal a tunelaje no lineal, y luego a la eyección del solitón. Además, la captura parcial del solitón fue también observada experimentalmente. Parece muy atractivo utilizar el efecto de tunelaje no lineal de solitones ópticos en el desarrollo de novedosos dispositivos de transferencia óptica de las velocidades de Tera-bits (Serkin, 2000; Sakaguchi y Tamura, 2004 y 2005). Propiedades ondulatorias y de las partículas de los solitones NLSE se han estudiado en el marco de diferentes modelos no lineales (véanse, por ejemplo, los trabajos de Yang y Wu, 2008; Kivshar y Malomed, 1989; Goodman *et al.*, 2004; Stoychev *et al.*, 2004 y sus referencias).

Las preguntas que nos gustaría discutir en el presente capítulo son: ¿qué ocurre si la barrera de potencial no está sólo en función de las coordenadas, sino que también es una función del tiempo y a su vez una función de la amplitud y duración del tunelaje solitónico en sí? Y ¿qué pasa si la amplitud y la duración de solitón inicial varían en el tiempo y el solitón se aproxima a una barrera clásicamente prohibida? En particular, ¿qué sucede en el caso del tunelaje no lineal del solitón auto-compresivo? Llevamos a cabo simulaciones numéricas de la ecuación de Gross-Pitaevskii no autónoma en una dimensión, con la no linealidad y una barrera de potencial externa, variables en el tiempo, y revelamos todos los posibles escenarios macroscópicos de tunelaje de las ondas de materia, incluyendo tunelaje lineal y no lineal, captura y la expulsión de solitones. El crecimiento de

no linealidad en el tiempo proporciona la auto-compresión del solitón durante su tunelaje. En experimentos con ondas de materia de Bose-Einstein esto puede lograrse mediante el uso de las resonancias Feshbach para modular la longitud de dispersión con el tiempo.

## I. EFECTO TÚNEL PARA SOLITONES

El enfoque de un libro de texto estándar describe el efecto túnel de una partícula puntual en un potencial estático externo. La presentación canónica cuántica de túnel se define en términos de una función de onda que pasa a través de una barrera de potencial energético clásicamente prohibida. En otras palabras, cuando un máximo de la barrera de potencial es mayor que la energía de la partícula puntual, la ecuación de Schrödinger tiene soluciones no nulas en la región clásicamente prohibida en el otro lado de la barrera, que es la razón por la cual la probabilidad de transmisión en mecánica cuántica no necesita ser cero (McKagan *et al.*, 2009). Si, por otro lado, la energía de una partícula puntual es mayor que el máximo de barrera de potencial, entonces no hay punto de reflexión; la reflexión está prohibida en la mecánica clásica. Sin embargo, la mecánica cuántica permite el reflejo de una partícula en una región clásicamente permitida, incluso cuando no hay un punto de reflexión clásica. La reflexión cuántica puede ocurrir por encima de una barrera de potencial repulsivo o un potencial atractivo.

Cabe destacar que el efecto de túnel cuántico es un tema de interés siempre renovado, experimental y teórico. En 2011 se celebraron los 100 años del experimento más importante en la historia de la ciencia –la dispersión de Rutherford de partículas alfa y el descubrimiento del núcleo atómico–. Hoy en día el problema de la dispersión en energías bajas y la fusión sub-barrera dieron un nuevo impulso relacionado con la emergencia de los haces de iones radiactivos (RIB) (Lemasson *et al.*, 2009). Estos núcleos exóticos se caracterizan por la baja energía de enlace de un neutrón o un protón y las distribuciones espaciales asimétricas y extendidas (halos).

Intensivos estudios teóricos y experimentales realizados en los últimos años mostraron que las reacciones sub-barreras y fusión (tunelaje) de los núcleos exóticos, como  ${}^6,8\text{He}$  y  ${}^8\text{B}$  en blancos pesados, se distinguen mucho de los núcleos “normales” estables:  ${}^4\text{He}$ ,  ${}^6,7\text{Li}$ ,  ${}^7\text{Be}$ , es decir, muestran un gran incremento de las secciones transversales (Aguilera *et al.*, 2009 y 2011; Belyaeva *et al.*, 2009). El aumento de la sección transversal de reacción está conectado no sólo con la fuerte contribución de los procesos de rompimiento de los proyectiles exóticos, sino con un aumento notable en la sección transversal de fusión en sí. La fusión de los núcleos complejos no puede ser entendida como escenario simple de penetración

de barrera Gamow por un objeto sin estructura. Los problemas más importantes están conectados con efecto túnel y la reflexión de una partícula compuesta, en la que la partícula en sí tiene una estructura interna. Recientemente, Bertulani y colaboradores (2007) predijeron importantes posibilidades de aumento de la probabilidad de tunelaje, debido a la presencia de grados de libertad intrínsecos. Además de las reacciones químicas y nucleares bajo la barrera, los estudios sobre el tunelaje de los objetos extendidos y compuestos son de interés fundamental para las condiciones astrofísicas.

Nótese que la importancia del concepto de solitón en la descripción de las partículas elementales extensas ha sido debatida fuertemente (Rebbi y Soliani, 1984). El solitón no se comporta como una partícula puntual clásica; el solitón, como un objeto no lineal extendido y compuesto, posee un comportamiento de mecánica ondulatoria también.

Históricamente, la primera demostración de que el solitón es una partícula compuesta “que posee un grado de la libertad interior” fue encontrada por Kosevich (1990). Kosevich consideró la dispersión resonante y no resonante de solitón por las impurezas en la aproximación de campo medio y concluyó que “tenemos que interpretar el solitón como un estado ligado de  $N$  cuasi-partículas con sus amplitudes proporcionales a  $N$ , es por eso que, cuando la energía de enlace de cuasi-partículas en el solitón es considerablemente mayor que su energía cinética, es difícil ‘arrancar’ una cuasi-partícula libre del solitón” (Kosevich, 1990). Entonces, existe un mecanismo físico de la descomposición de solitón durante su dispersión por potenciales externos o el decaimiento solitónico. Tal vez Kosevich fue el primero en formular el teorema de Ehrenfest no lineal. Se ha demostrado que si el potencial no depende del tiempo, la ley de movimiento del centro de masa de solitón puede ser considerada como la generalización del teorema de Ehrenfest para el sistema no lineal descrito por la aproximación de campo medio (véanse las ecuaciones [14] y [19] en la artículo [Kosevich, 1990] para más detalles). Para subrayar la complejidad del problema, Kosevich concluyó que “el NLSE solitón es similar a un átomo pero no a una partícula elemental” (1990).

Debe hacerse especial hincapié en que el efecto de energía de enlace en solitones es más pronunciado en el emparejamiento de solitones claros y oscuros de la NLSE (Afanas'ev *et al.*, 1988).

## II. TUNELAJE SOLITÓNICO NO LINEAL A TRAVÉS DE LOS POTENCIALES EXTERNOS: EL PAPEL DE LA ENERGÍA DE ENLACE SOLITÓN

A continuación se consideran nuevos escenarios de tunelaje de solitones que ilustran cómo “ocultos” grados de libertad pueden aparecer en el proceso túnel no

lineal de solitones (Belyaeva *et al.*, 2010). La ecuación de Schrödinger no lineal canónica

$$i \frac{\partial \psi(\xi, \eta)}{\partial \eta} = -\frac{1}{2} \frac{\partial^2 \psi(\xi, \eta)}{\partial \xi^2} - |\psi(\xi, \eta)|^2 \psi(\xi, \eta) + U_{ext}(\xi) \quad (1)$$

con  $U_{ext}(\xi) = 0$  tuvo su comienzo como una descripción analítica de solitones ópticos y de BEC para sistemas autónomos. Sin el potencial de auto-interacción  $|\psi(\xi, \eta)|^2$ , la Ec. (1) es similar a una ecuación de Schrödinger mecánica-cuántica escrita en forma adimensional. Este hecho fue utilizado por Zakharov y Shabat cuando propusieron nombrar a la Ec. (1) como la ecuación no lineal de Schrödinger (Zakharov y Shabat, 1972). Vamos a calcular la energía media como el valor esperado del operador hamiltoniano

$$\langle E \rangle = \frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle} = \frac{\int_{-\infty}^{\infty} [|\psi_{\xi}(\xi, \eta)|^2 - |\psi(\xi, \eta)|^4] d\xi}{2 \int_{-\infty}^{\infty} |\psi(\xi, \eta)|^2 d\xi} \quad (2)$$

El primer término entre corchetes de la Ec. (2) determina la energía cinética media; el segundo se puede considerar como la energía de enlace de las cuasipartículas asociadas con el solitón.

Es conocido que la Ec. (1) con  $U_{ext}(\xi) = 0$  posee la solución llamada el solitón canónico

$$\psi_{sol}(\xi, \eta) = \kappa_0 \operatorname{sech}(\kappa_0 \xi) \exp[i(k\xi - \omega\eta)], \quad (3)$$

donde  $\kappa_0$  es el factor de forma solitónica, el cual determina la distribución de amplitud y está relacionado con el ancho del solitón  $\xi_{sol} = 1/\kappa_0$ , y  $\omega$  es la frecuencia del solitón:  $\omega = (k^2 - \kappa_0^2)/2$ , y  $k$  es el número de onda del solitón. La velocidad del solitón está definida como  $v_0 = \partial\omega/\partial k = k$ , y  $\xi = (\xi - \xi_0 - v_0\eta)$ , donde la posición inicial del centro de masa está definida por  $\xi_0$ .

La ecuación canónica NLSE sin potencial externo  $U_{ext}(\xi) = 0$  es integrable debido al hecho de que tiene un número infinito de integrales conservadas. Las tres primeras cantidades conservadas son

$$\int_{-\infty}^{\infty} |\psi(\xi, \eta)|^2 d\xi = 2\kappa_0 = N, \quad (4)$$

$$\langle P \rangle = \frac{i \int_{-\infty}^{\infty} \psi(\xi, \eta) \psi_{\xi}^*(\xi, \eta) - \psi_{\xi}(\xi, \eta) \psi^*(\xi, \eta) d\xi}{2 \int_{-\infty}^{\infty} |\psi(\xi, \eta)|^2 d\xi} = v_0 \quad (5)$$

$$\langle E \rangle = \frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle} = \langle \hat{T}_{kin} \rangle + \langle \hat{E}_b \rangle = \frac{1}{2} \frac{\int_{-\infty}^{\infty} |\psi_{\xi}(\xi, \eta)|^2 d\xi}{\int_{-\infty}^{\infty} |\psi(\xi, \eta)|^2 d\xi} - \frac{1}{2} \frac{\int_{-\infty}^{\infty} |\psi(\xi, \eta)|^4 d\xi}{\int_{-\infty}^{\infty} |\psi(\xi, \eta)|^2 d\xi}, \quad (6)$$

donde  $N$  en la Ec. (4) es una constante que puede ser interpretada como el número total de átomos en el condensado, o el número total de fotones en el pulso de láser.

En la energía media total de la Ec. (6) separamos especialmente dos términos relacionados con las funciones lineales y no lineales del problema. Cabe destacar que, en contraste con la condición de mecánica cuántica para la densidad de probabilidad  $\int_{-\infty}^{\infty} |\Psi(\xi, \eta)|^2 d\xi = 1$ , la condición de la Ec. (4) se debe aplicar en la Ec. (6). A continuación se obtiene para la energía cinética y de enlace del solitón, correspondientemente

$$\langle \hat{T}_{kin} \rangle = \frac{1}{2} \frac{\int_{-\infty}^{\infty} |\psi_{\xi}(\xi, \eta)|^2 d\xi}{\int_{-\infty}^{\infty} |\psi(\xi, \eta)|^2 d\xi} = \frac{1}{6} \kappa_0^2 + \frac{1}{2} v_0^2, \quad (7)$$

$$\langle \hat{E}_b \rangle = \frac{1}{2} \frac{\int_{-\infty}^{\infty} |\psi(\xi, \eta)|^4 d\xi}{\int_{-\infty}^{\infty} |\psi(\xi, \eta)|^2 d\xi} = \frac{1}{3} \kappa_0^2. \quad (8)$$

Tenemos que recordar que los valores medios para mecánicas clásica y cuántica se interpretan de manera diferente. Es por eso que la energía cinética media del solitón no es igual al caso clásico:  $\langle T_{kin} \rangle \neq v_0^2/2$ . Lo que es más, si fijamos  $v_0 = 0$ , obtenemos  $\langle T_{kin} \rangle = \kappa_0^2/6 > 0$ . La energía cinética media no desaparece para el solitón al estar en reposo, en analogía absoluta con una partícula mecánica-cuántica. Físicamente, esta característica está relacionada con el principio de incertidumbre de Heisenberg, es por eso que la Ec. (7) puede ser considerada como la generalización de este principio aplicada a la dinámica de solitones NLSE.

Hacemos hincapié en que hasta la fecha podemos encontrar muchas interpretaciones erróneas de la energía de enlace de solitón que “vagan de un lugar a otro”, y estas interpretaciones erróneas todavía están presentes en los enigmas fundamentales de los solitones.

En la física de bajas temperaturas, por primera vez la Ec. (1) fue nombrada por Tsuzuki (1971) como la ecuación Gross-Pitaevskii para el condensado a la temperatura del cero absoluto (donde  $U_{ext}(\xi) = \mu$  es el potencial químico asumiendo que es constante). Históricamente, Zakharov y Shabat propusieron el nombre de “ecuación de Schrödinger no lineal” de la Ec. (1) en el límite  $U_{ext}(\xi)$

= 0, para acentuar su analogía con la ecuación de Schrödinger ordinaria de la mecánica cuántica, con el potencial de auto-interacción equivalente  $U_{ext}(\xi, \eta) = |\psi(\xi, \eta)|^2$ . Teniendo en cuenta las leyes de conservación para la NLSE, Zakharov y Shabat (1972) propusieron interpretar las primeras tres integrales “aparte de los coeficientes” como el número de partículas, el impulso y la energía. Después de este trabajo en todas las obras posteriores el impulso y la energía del solitón se han identificado directamente con dos integrales de movimiento de la NLSE:

$$P = \frac{i}{2} \int_{-\infty}^{\infty} [\psi(\xi, \eta)\psi_{\xi}^*(\xi, \eta) - \psi_{\xi}(\xi, \eta)\psi^*(\xi, \eta)] d\xi = C_2, \quad (9)$$

$$E = \frac{1}{2} \int_{-\infty}^{\infty} [|\psi_{\xi}(\xi, \eta)|^2 - |\psi(\xi, \eta)|^4] d\xi = C_3 \quad (10)$$

con la condición ordinaria de la mecánica cuántica para la densidad de probabilidad  $\int_{-\infty}^{\infty} |\Psi(\xi, \eta)|^2 d\xi = C_1 = 1$ .

Pero hacemos énfasis en que en el caso de la NLSE, la condición de normalización principal no es  $C_1 \neq 1$  como en la mecánica cuántica y, obviamente, depende del factor de forma del solitón  $C_1 = \int_{-\infty}^{\infty} |\psi(\xi, \eta)|^2 d\xi = 2\kappa_0 = N$ . Es por eso que en el impulso solitónico la energía cinética y su energía de enlace están dadas por (4-8) y las interpretaciones del factor de forma de solitones, como su masa  $M = 2\kappa_0$ , el impulso  $\langle P \rangle = Mv_0 = 2\kappa_0 v_0$ , la energía cinética  $\langle T_{kin} \rangle = Mv_0^2/2 = \kappa_0 v_0^2$ , y la energía de enlace del solitón  $\langle E_b \rangle = \frac{1}{2} \int_{-\infty}^{\infty} |\psi(\xi, \mu)|^4 d\xi = 2\kappa_0^3/3$ , pueden ser considerados como interpretaciones erróneas vivientes hace mucho tiempo y enigmas de los solitones de Schrödinger.

Similar a la energía de enlace nuclear, la energía de enlace solitónico  $\langle E_b \rangle = -\kappa_0^2/3$ , Ec. (8), denota el grado de la fuerza con la que las “cuasi-partículas” se unen en un solitón. Así como el aumento del factor de forma del solitón  $\kappa_0$  resulta en un crecimiento de la energía de enlace del solitón, así también el creciente de no linealidad resulta en el mismo efecto.

En aplicaciones ópticas el aumento de la energía de enlace de solitones crece con potencia de pico y con la disminución de la velocidad de dispersión de grupo a lo largo de la distancia de propagación (Reeves-Hall *et al.*, 2000; Reeves-Hall y Taylor, 2001; Travers *et al.*, 2007; Liu *et al.*, 2001; Sysoliatin, *et al.*, 2007; Sysoliatin y Belanov, 2008). En las aplicaciones de BEC la energía de enlace del solitón de onda de materia se aumenta en la gestión de resonancia de Feshbach y con crecimiento de la densidad atómica.

En la mecánica clásica, si se considera una partícula en movimiento encontrando una barrera de potencial, la barrera retarda la partícula; y si la energía de

la partícula es demasiado baja en relación con la altura de la barrera, la partícula se refleja. En el escenario clásico la partícula se acelera encontrando el pozo de potencial y se transmite siempre. En el escenario de tunelaje cuántico en ambos casos tenemos la reflexión y transmisión, y el comportamiento de la función de onda  $\psi(\xi, \eta)$  depende de los parámetros del potencial y de la energía de la partícula incidente (Belyaeva *et al.*, 2011 y 2012).

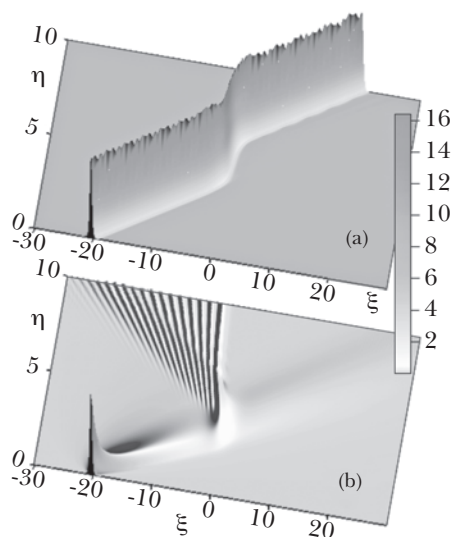
La Figura II.1 muestra la comparación de los escenarios de tunelaje lineal y no lineal del paquete de ondas inicial, tipo solitones (con forma de sech), en una barrera de potencial de repulsión Gaussiano  $U(\xi) = U_0 \exp(-\xi^2/\Delta^2)$ . El paquete lineal (Figura II.1[b]) muestra el comportamiento mecánico-cuántico de la difusión ordinaria y se comporta como una onda plana lentamente modulada que se divide en dos partes: reflejada y transmitida. A diferencia del paquete lineal, el paquete de ondas no lineal, gobernado por la NLSE Ec. (1), puede tener comportamiento dinámico absolutamente diferente.

En la Figura II.1(a) se muestra, por ejemplo, la transmisión del solitón bajo la barrera sin perder la forma inicial y sin producir una onda reflejada. También se observa que el solitón se detiene y sufre un retraso bajo la barrera.

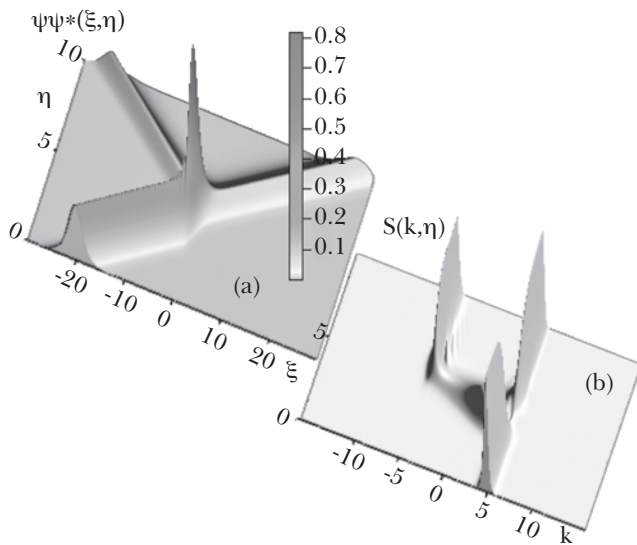
Las Figuras II.2-4 muestran la comparación de tres escenarios de dispersión del paquete de ondas no lineal, tipo solitón (con forma de sech), en una barrera de potencial de repulsión Gaussiano  $U(\xi) = U_0 \exp(-\xi^2/\Delta^2)$ , dependiendo de la energía de enlace solitónica Ec. (8). En los tres casos la energía cinética del solitón es igual y menor que la altura máxima de la barrera,  $T_{kin} < U_0$ . Sin embargo, el comportamiento del solitón es absolutamente diferente y depende del valor de su factor de forma  $\kappa_0$ .

La Figura II.2 muestra que el solitón con  $\kappa_0 = 0.5$  (la altura del potencial es  $U_0 = 12.6$ ) se divide en dos partes: reflejada y transmitida como una onda mecánica-cuántica. En la Figura II.3 se muestra la transmisión total bajo la barrera del solitón con  $\kappa_0 = 6.0$  (la altura del potencial es  $U_0 = 12.55$ ) como una partícula mecánica-cuántica. En la Figura II.4 se muestra que el solitón con  $\kappa_0 = 6.0$  (la altura del potencial es  $U_0 = 12.60$ ) se refleja totalmente a partir de la barrera como una partícula clásica. Todos estos diferentes escenarios dependen de la energía de enlace del solitón (a través de su factor de forma  $\kappa_0$ ). Precisamente, si la energía de enlace es menor que la altura máxima de la barrera  $U_0$ , el paquete de ondas no lineal muestra un comportamiento mecánico-cuántico y tiene una dinámica ondulatoria. En el caso en que la energía de enlace es mayor de  $U_0$ , el solitón tiene un comportamiento clásico, es decir, muestra su naturaleza de partícula.

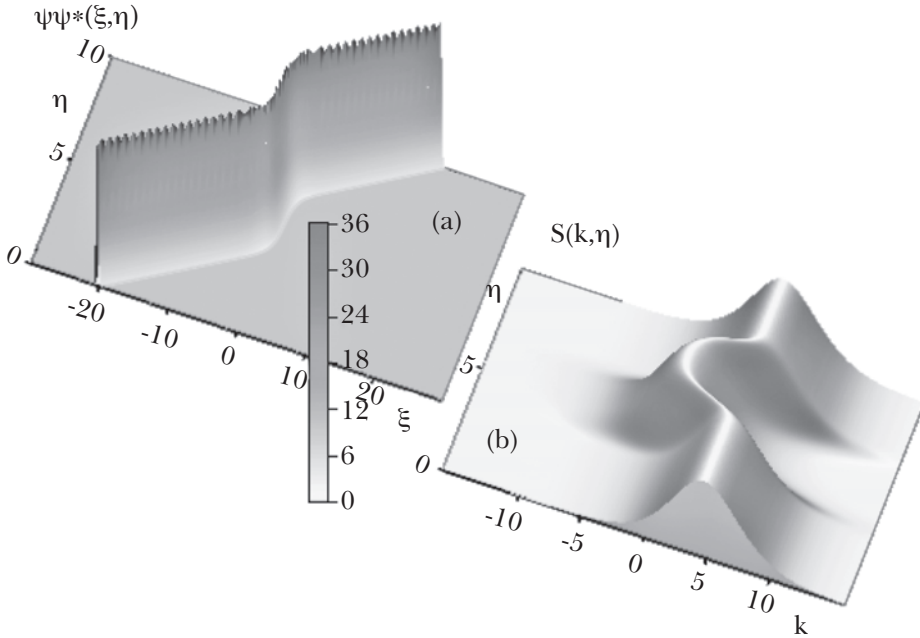
Existe un valor crítico del factor de forma del solitón  $\kappa_{cr}$  a partir del cual todos los solitones con  $\kappa_0 > \kappa_{cr}$  se reflejan y nunca se dividen o transmiten (Figura II.4[a]). Hay un caso muy especial cuando la energía de enlace es casi igual a  $U_0$  y el



**Figura II.1.** Comparación de los escenarios del tunelaje de un paquete de ondas lineal y no lineal con velocidad  $V = -5.0$  a través del potencial de repulsión Gaussiano (barrera)  $U(\xi) = U_0 \exp(-\xi^2/\Delta^2)$  con una altura  $U_0 = 12.6$  y una anchura  $\Delta = 2.0$  : (a) tunelaje del solitón con un factor de forma  $\kappa_0 = 4.0$ ; (b) tunelaje lineal (mecánico-cuántico) de un paquete de ondas dispersivo.



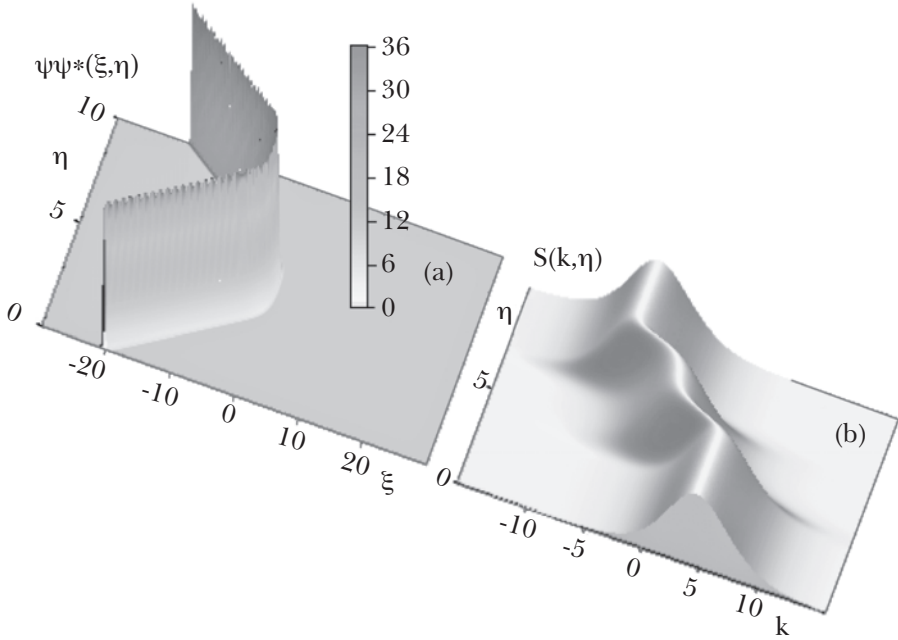
**Figura II.2.** División del solitón con el factor de forma  $\kappa_0 = 0.5$  y la velocidad  $V = -5.0$  durante su tunelaje no lineal a través del potencial Gaussiano de repulsión (la barrera)  $U(\xi) = U_0 \exp(-\xi^2/\Delta^2)$  con la altura  $U_0 = 12.6$  y la anchura  $\Delta = 2.0$  : (a) el comportamiento espacio-temporal de  $|\psi(\xi, \eta)|^2$ ; (b) el comportamiento de su espectro  $S(k, \eta)$ .



**Figura II.3.** Transmisión no lineal bajo la barrera del solitón energético con el factor de forma  $\kappa_0 = 6.0$  y la velocidad  $V = -5.0$  durante su tunelaje no lineal a través del potencial Gaussiano de repulsión (la barrera)  $U(\xi) = U_0 \exp(-\xi^2/\Delta^2)$  con la altura  $U_0 = 12.55$  y la anchura  $\Delta = 2.0$ : (a) el comportamiento espacio-temporal de  $|\psi(\xi, \eta)|^2$ ; (b) el comportamiento de su espectro  $S(k, \eta)$ .

paquete de ondas no lineal se transmite completamente bajo la barrera. Así vemos que la energía de enlace es una medida de cómo las partes internas (“cuasi-partículas”) del paquete de ondas no lineal están fuertemente confinadas en esta estructura localizada.

Aparte de la dinámica espacio-temporal, las Figuras II.2(b)-4(b) muestran la dinámica del espectro del solitón durante su dispersión en el potencial. Cuando el solitón se divide en las partes reflejada y transmitida en su espectro aparece el nuevo vector de onda  $-k$ , correspondiente a la onda reflejada (Figura II.2[b]). La transmisión del solitón bajo la barrera con su parada y retraso se acompaña de la producción del espectro tipo “herradura”; es decir, el vector de onda disminuye



**Figura II.4.** Dinámica newtoniana del solitón energético con el factor de forma  $\kappa_0 = 6.0$  y la velocidad  $V = -5.0$  durante su dispersión en el potencial de repulsión Gaussiano (la barrera)  $U(\xi) = U_0 \exp(-\xi^2/\Delta^2)$  con una altura  $U_0 = 12.60$  y una anchura  $\Delta = 2.0$  : (a) el comportamiento espacio-temporal de  $|\psi(\xi, \eta)|^2$ ; (b) el comportamiento de su espectro  $S(k, \eta)$ .

hasta cero y regresa a su valor inicial (Figura II.3[b]). En el caso de reflexión completa a partir la barrera, en el espectro del solitón, el vector de onda inicial se transforma en el vector de onda  $-k$ , como se muestra en la Figura II.4(b).

### III. TUNELAJE DE LOS SOLITONES AUTO-COMPRESIVOS

En aplicaciones de óptica no lineal y BEC, el NLSE (1) puede escribirse en la forma más general, teniendo en cuenta la variación de la dispersión  $D(\eta)$ , la no linealidad  $G(\eta)$ , la amplificación o la pérdida  $\alpha(\eta)$  y el potencial externo  $U_{ext}(\xi)$

$$i \frac{\partial \psi(\xi, \eta)}{\partial \eta} = \left[ -\frac{1}{2} D(\eta) \frac{\partial^2}{\partial \xi^2} - G(\eta) |\psi(\xi, \eta)|^2 + U_{ext}(\xi) + \frac{i}{2} \alpha(\eta) \right] \psi(\xi, \eta). \quad (11)$$

La Ec. (11) pertenece a la familia de las ecuaciones no autónomas. Podemos aplicar una sustitución simple  $\psi(\xi, \eta) = q(\xi, \eta) \exp(i \int_0^\eta \alpha(\tau) d\tau / 2)$  para transformar la Ec. (11) con la amplificación o la pérdida a la NLSE sin amplificación o pérdida

$$i \frac{\partial q(\xi, \eta)}{\partial \eta} = \left[ -\frac{1}{2} D(\eta) \frac{\partial^2}{\partial \xi^2} - R(\eta) |q(\xi, \eta)|^2 + U_{ext}(\xi) \right] q(\xi, \eta), \quad (12)$$

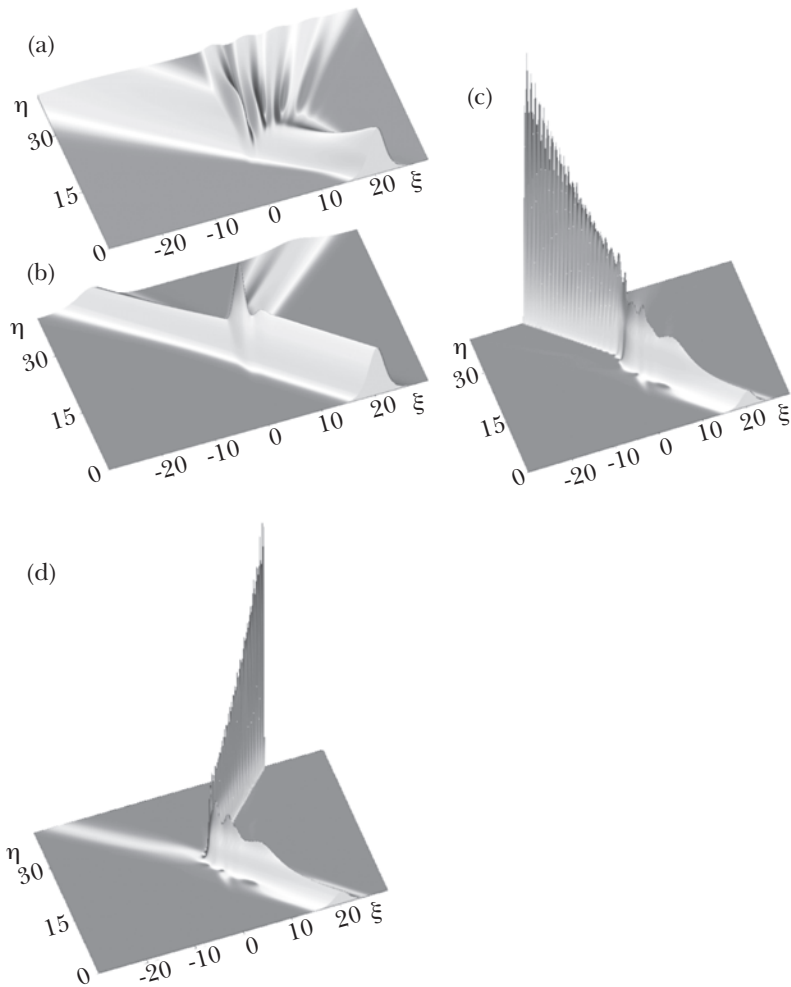
pero con la no linealidad cambiada  $R(\eta) = G(\eta) \exp(i \int_0^\eta \alpha(\tau) d\tau)$ .

Diferentes aspectos de la dinámica de solitones en sistemas no autónomos y dispersivos descritas por la NLSE generalizada se investigaron en varios artículos (Serkin *et al.*, 2000, 2002, 2007; Shin, 2008; Zhao *et al.*, 2008; Li and Luo, 2009). Las soluciones solitónicas exactas de la Ec. (12) no autónoma extienden substancialmente y generalizan el concepto de solitones clásicos. Una multitud de solitones no autónomos interactúan elásticamente y, en general, mueven con amplitudes variables, velocidades y espectros adaptados a los potenciales externos, así como a la dispersión y no linealidad variables.

En nuestros experimentos numéricos el solitón se inicia lejos de la barrera localizada y se propaga hacia el potencial  $U(\xi) = U_0(1 - \varepsilon^2 \xi^2)$  ubicado en la región  $|\xi| > 1$  con diferentes parámetros  $U_0$  y  $\varepsilon$ , bajo la condición de que  $U(\xi) = 0$  en los intervalos  $|\xi| > 1$ . Para efectos de nuestro análisis, es conveniente utilizar el caso de amplificación  $\alpha_0$  constante en la Ec. (11) y  $R(\eta) = G(\eta) \exp(i \int_0^\eta \alpha(\tau) d\tau)$  en la Ec. (12). En la Figura III.1 se ilustra un tunelaje de un solitón auto-compresivo debido a la no linealidad que aumenta exponencialmente:  $R(\eta) = \exp(\alpha_0 \eta)$ .

Los resultados del modelado computacional se muestran a continuación:

1. En no linealidades muy bajas cuando  $R \rightarrow 0$ , el paquete de ondas inicial, tipo solitones (con forma de sech), muestra el comportamiento de la difusión ordinaria y se comporta como una onda plana lentamente modulada que se divide en dos partes: reflejada y transmitida, como se muestra en la Figura III.1(a).
2. Con no linealidad fuerte  $R \rightarrow 1$ ,  $G = 1$  y  $\alpha_0 \approx 0$ , sin efecto de auto-compresión, todos los solitones con energía cinética  $E_{kin} \leq U_0$  se dividen en solitones transmitidos y reflejados como se muestra en la Figura III.1(b).
3. Con grandes parámetros  $\alpha_0$  y el aumento de la no linealidad exponencial  $R(\eta) = \exp(\alpha_0 \eta)$ ; por otro lado, el solitón está casi completamente transmitido a través de una barrera clásicamente prohibida (véase Figura III.1[c]). Este escenario muestra la posibilidad de la transmisión del solitón bajo la barrera en la región intermedia de la fuerza no lineal  $R(\eta) = \exp(\alpha_0 \eta)$ . La trayectoria del solitón en este régimen de transmisión, bajo la barrera, es extremadamente sensible a la fuerza de la no linealidad propuesta por el parámetro  $\alpha_0$ .



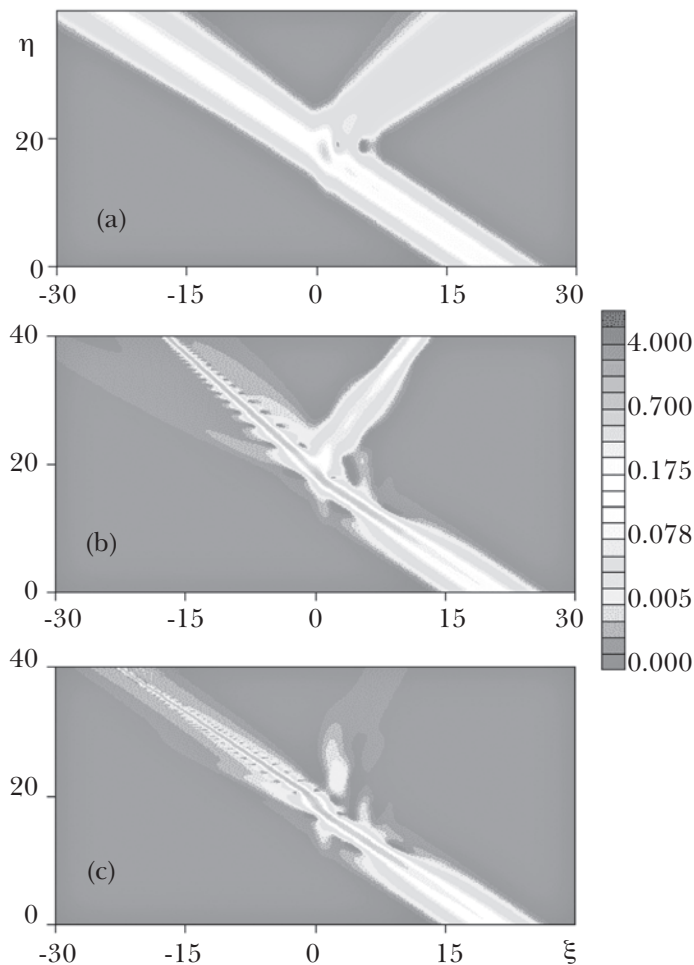
**Figura III.1.** La dependencia de las características principales del efecto túnel no lineal de solitón de la energía de enlace solitónica, calculado en el marco del modelo (12): (a) el túnel lineal de paquete ondas con parámetros:  $R = 0$ ,  $\alpha_0 = 0.0$ ; (b) el decaimiento de solitón inicial a un solitón transmitido y otro reflejado,  $\alpha_0 = 0.0$ ; (c) el efecto de túnel de solitón auto-compresivo bajo la barrera,  $\alpha_0 = 0.075$ ; (d) la reflexión completa de solitón auto-compresivo con propiedades de una partícula;  $\alpha_0 = 0.080$ . El potencial parabólico de repulsión (la barrera)  $U(\xi) = U_0(1-\xi^2)$  se localiza en la región  $-1 \leq \xi \leq 1$  con altura  $U_0 = 0.6$ .

4. Para los parámetros  $\alpha_0$  grandes hay un valor crítico de no linealidad por encima del cual todos los solitones auto-compresivos se reflejan y nunca transmiten (Figura III.1[d]). Hemos encontrado una reflexión de solitón extremadamente elástica en todos los casos de parámetros  $\alpha_0$  grandes. Esencialmente que no se pierde la energía durante la reflexión del solitón; toda la energía cinética del solitón entrante se convierte en la de lo reflejado. Debido a esto, el valor absoluto de su velocidad es inalterado (Figura III.1[d]).

Estos tres escenarios (Figura III.1[b-d]) considerados corresponden a la misma sin dimensional “masa”  $m = 1$  dada por las ecuaciones (1, 11) y por tanto iguales energías cinéticas “clásicas”  $v_0^2/2 = 0.5$ .

Para tener analogía con el efecto cuántico de reflexión sobre barrera, consideremos el caso en que el solitón entrante tiene una energía cinética que es mayor que la altura de la barrera. En la región de la energía cinética  $E_{kin} \geq U_0$  (por ejemplo, en la Figura III.2:  $E_{kin} = v_0^2/2 = 0.605$ ,  $v_0 = 1.1$ , y  $U_0 = 0.6$ ), las simulaciones muestran un proceso no lineal de dispersión más complicada. En este caso el comportamiento del solitón es muy diferente al de la mecánica clásica: el solitón auto-compresivo en el tiempo se divide en solitones reflejados y transmitidos (también auto-compresivos), como se puede ver en la Figura III.2(a, b), donde se muestran los mapas de contorno de  $|\psi(\xi, \eta)|^2$ . Por tanto, el efecto de la reflexión del solitón sobre la barrera se puede observar en este caso. Cuando el parámetro  $\alpha_0$  se incrementa, el comportamiento de túnel solitónico se cambia drásticamente: se ha encontrado que la intensidad del solitón reflejado también se incrementa con el tiempo. Nótese que en este caso el solitón transmitido exhibe oscilaciones con amplitudes y periodos muy regulares, presentados en la Figura III.2(c). Estas oscilaciones pueden explicarse por el efecto bien conocido que está conectado con transformaciones asintóticas de amplitudes solitónicas inducidas por perturbaciones de su factor de forma (Hasegawa, 1989, 1995, 2003; Agrawal, 2001; Akhmediev y Ankiewicz, 1997).

Cuando  $\alpha_0$  se incrementa más y más, el comportamiento muestra en la Figura III.2 se cambia. Para los parámetros grandes de  $\alpha_0$  el solitón parece estar completamente transmitido y nunca reflejado (véase Figura III.2[c]). En otras palabras, un solitón auto-compresivo con gran amplitud se asemeja a una partícula clásica en el sentido de que el solitón mantiene su integridad y sigue una trayectoria clásica bien definida. En este caso, el solitón auto-compresivo permanece principalmente como partícula clásica, con muy poca pérdida de energía a las ondas dispersivas. Existe un valor crítico  $\alpha_0$ , por encima del cual el solitón se transmite clásicamente. Cuanto mayor es el efecto de auto-compresión del solitón dado por



**Figura III.2.** El efecto de reflexión de solitón sobre barrera calculado en la región de parámetros:  $U_0 = 0.6$ ;  $\kappa_0 = 0.5$ ;  $v_0 = -1.1$ ,  $v_0^2/2 = 0.605$  y (a)  $\alpha_0 = 0.0$ ; (b)  $\alpha_0 = 0.07$ ; y (c)  $\alpha_0 = 0.08$ .

el parámetro  $\alpha_0$ , menor es la radiación no solitónica reflejada y transmitida, por consecuencia, es mayor la intensidad del solitón transmitido.

Resumiendo nuestros experimentos numéricos, concluimos que el efecto túnel solitónico tiene muchas propiedades interesantes que deben ser investigadas, por ejemplo, mediante la aplicación directa de los métodos analíticos como la Teoría de Dispersión Inversa.

## CONCLUSIONES

Hemos considerado en este capítulo el efecto no lineal de túnel de los solitones a través de la barrera de potencial, teniendo en cuenta el efecto de auto-compresión del solitón. También se ha demostrado que existieron interpretaciones erróneas de la energía del solitón. Entre los escenarios posibles de tunelaje, incluido el comportamiento lineal y no lineal de paquetes de ondas de solitones, existen la captura, la emisión y la división de los solitones.

La amplificación de un solitón de la ecuación de Schrödinger, que es matemáticamente equivalente al crecimiento de la no linealidad en tiempo, proporciona la auto-compresión del solitón durante su transmisión a través de la barrera parabólica o reflexión de la de ella. En experimentos con ondas de materia de Bose-Einstein, ésta puede lograrse mediante el uso de las resonancias Feshbach, para modular la longitud de dispersión con el tiempo. En óptica no lineal, la auto-compresión de solitón se puede realizar mediante el aumento de la potencia del pulso solitónico o decreciente de la dispersión de la velocidad grupal de ondas a lo largo de su propagación (Reeves-Hall *et al.*, 2000; Reeves-Hall y Taylor, 2001; Travers, 2007; Liu *et al.*, 2001; Sysoliatin, 2007, 2008). Hemos encontrado que, debido a la auto-compresión del solitón, su comportamiento no lineal difiere esencialmente del efecto túnel mecánico-cuántico lineal. En particular, se descubrió la posibilidad de que la transmisión o reflexión completa para solitones baja la barrera del potencial.

## REFERENCIAS

- Afanas'ev, V. V., Dianov, E. M., Prokhorov, A. M. y Serkin, V. N. (1988). "Nonlinear Pairing of Light and Dark Optical Solitons", *JETP Lett.* 48, 638.
- Afanas'ev, V. V., Dianov, E. M. y Serkin, V. N. (1989). "Nonlinear Pairing of Short Bright and Dark Soliton Pulses by Phase Cross Modulation", *IEEE J. Quant. Electron.* 25 (12), 2656-2664.
- Afanas'ev, V. V., Kivshar, Y. S., Konotop, V. V. y Serkin, V. N. (1989). *Opt. Lett.* 14, 805.
- Agrawal, G. P. (2001). *Nonlinear Fiber Optics*. 3a. ed. Academic Press, San Diego.
- (2001). *Applications of Nonlinear Fiber-Optics*. 3a. ed. Academic Press, San Diego.
- Aguilera, E. F., Martínez-Quiroz, E., Lizcano, D., Gómez-Camacho, A., Kolata, J. J., Lamm, L. O., Guimarães, V., Lichtenthäler, R., Camargo, O., Becchetti, F. D., Jiang, H., De Young, P. A., Mears, P. J. y Belyaeva, T. L. (2009). *Phys. Rev. C* 79, 021601 (R).
- Aguilera, E.F. *et al.* (2011). *Phys. Rev. Lett.* 107, 092701.

- Ahufinger, V., Mebrahtu, A., Corbalán, R. y Sanpera, A. (2007). *New Journal of Physics*. 9, 4.
- Akhmediev, N. N. y Ankiewicz, A. (1997). *Solitons. Nonlinear Pulses and Beams*. Charman and Hall, Londres.
- Albiez, M., Gati, R., Fölling, J., Hunsmann, S., Cristiani, M. y Oberthaler, M. K. (2005). *Phys. Rev. Lett.* 95, 010402.
- Anderson, B. P. y Kasevich, M. (1999). *Science*. 282, 1686.
- Ankerhold, J. (2007). *Quantum Tunneling in Complex Systems: The Semiclassical Approach*. Springer.
- Barak, A., Peleg, O., Stucchio, C., Soffer, A. y Segev, M. (2008). “Observation of Soliton Tunneling Phenomena and Soliton Ejection”, *Phys. Rev. Lett.* Vol. 100, 153901.
- Belyaeva, T. L., Aguilera, E. F., Martínez-Quiroz, E., Moro, A. M. y Kolata, J. J. (2009). *Phys. Rev.* C80, 064617.
- Belyaeva, T. L., Serkin, V. N., Hernandez-Tenorio, C. y García-Santibáñez, F. (2010). *J. Mod. Opt.* 57, 1087-1099.
- (2011). *Internet Electronic J. Nanoc. Moletrón.* 9, 1715-1736.
- (2012). *Eur. Phys. J. D* 66, 153.
- Bertulani, C. A., Flambaum, V. V. y Zelevinsky, V. G. (2007). *Phys. G: Nucl. Part. Phys.* 34, 2289.
- Busch, T. y Anglin, J. R. (2001). *Phys. Rev. Lett.* 87, 010401.
- Carr, L. D., Holland, M. J. y Malomed, B. A. (2005). *J. Phys. B* 38, 3217.
- Carr, L. D. y McKagan, S. B. (2009). *American Journal of Physics*. 77, 308.
- Cataliotti, F. S., Burger, S., Fort, C., Maddaloni, P., Trombettoni, A., Smerzi, A. y Inguscio, M. (2001). *Science*. 293, 843.
- Cornell, E. A. y Wieman, C. E. (2002). “Nobel Lecture: Bose-Einstein Condensation in a Dilute Gas, the First 70 Years and Some Recent Experiments”, *Rev. Mod. Phys.* 74, 875-893.
- Dekel, G., Fleurov, V., Soffer, A. y Stucchio, C. (2007). *Phys. Rev. A* 75, 043617.
- Dekel, G., Farberovich, O. V., Soffer, A. y Fleurov, V. (2009). *Physica D: Nonlinear Phenomena*. 238, 1475.
- Demidov, V. E., Hansen, U. H., y Demokritov, S. O. (2008). *Phys. Rev. B* 78, 054410.
- Fleurov, V. y Soffer, A. (2005). *Europhysics Letters*. 72, 287.
- Friedman, J. R. y Han, S. (2002). *Exploring the Quantum/Classical Frontier: Recent Advances in Macroscopic Quantum Phenomena*. Nova Science Publishers, Inc.
- Gamow, G. (1928a). “Zur Quantentheorie des Atomkernes”, *Z. Physik* 51, 204.
- (1928b). *Nature*. 122, 805.
- Goodman, R. H., Holmes, P. J. y Weinstein, M. I. (2004). *Physica. D* 192, 215.

- Gurney, R. W. y Condon, E. U. (1928). "Quantum Mechanics and Radioactive Disintegration", *Nature*. 122, 439.
- (1929). *Phys. Rev.* 33, 127.
- Hasegawa, A. (1989). *Optical Solitons in Fibers*. Springer-Verlag, Berlín.
- y Kodama, Y. (1995). *Solitons in Optical Communications*. Oxford University Press, Nueva York.
- Hasegawa, A. y Matsumoto, M. (2003). *Optical Solitons in Fibers*. Springer-Verlag, Berlín.
- Kagan, Y., Shlyapnikov, G. V. y Walraven, J. T. M. (1996). *Phys. Rev. Lett.* 76, 2670.
- Ketterle, W. (2002). "Nobel Lecture: When Atoms Behave As Waves: Bose-Einstein Condensation and the Atom Laser", *Rev. Mod. Phys.* 74, 1131-1151.
- Khaykovich, L., Schreck, F., Ferrari, G., Bourdel, T., Cubizolles, J., Carr, L. D., Castin, Y. y Salomon, C. (2002). "Formation of Matter Wave Bright Solitons", *Science*. 296, 1290-1293.
- Khomeriki, R., Ruffo, S. y Wimberger, S. (2007). *Europhysics Letters*. 77, 40005.
- Kivshar, Y. S. y Malomed, B. A. (1989). "Dynamics of Solitons in Nearly Integrable Systems", *Rev. Mod. Phys.* 61, 763.
- Kosevich, A. M. (1990). "Particle and Wave Properties of Solitones", *Physica D*. 41, 253.
- Lee, C. y Brand, J. (2006). *Europhysics Letters*. 73, 321.
- Lemasson, A., Shrivastava, A., Rejmund, M., Keeley, N., Zelevinsky, V., Bhattacharyya, S., Chatterjee, A., De France, G., Jacquot, B., Nanal, V., Pillay, R. G., Raabe, R. y Schmitt, C. (2009). *Phys. Rev. Lett.* 103, 232701.
- Li, L. y Luo, H. G. (2009). *Phys. Rev. E*. 79, 056610.
- Liu, X., Xu, C., Knox, W. H., Chandaha, J. K., Eggleton, B. J., Kosinski, S. G. y Windeler, R. S. (2001). *Opt. Lett.* 26, 358-360.
- McKagan, S. B., Perkins, K. K. y Wieman, C. E. (2008). *Phys. Rev. ST Phys. Educ. Res.* 4, 020103.
- Milburn, G. J., Corney, J., Wright, E. y Walls, D. F. (1997). *Phys. Rev. A* 55, 4318.
- Moiseyev, N., Carr, L. D., Malomed, B. A. y Band, Y. B. (2004). *J. Phys. B: At. Mol. Opt. Phys.* 37, L1.
- Morsch, O., Mulle, J. H., Cristiani, M., Ciampini, D. y Arimondo, E. (2001). *Phys. Rev. Lett.* 87, 140402.
- Newell, A. C. (1978). "Nonlinear Tunnelling", *J. Math. Phys.* 19, 1126-1136.
- Raghavan, S., Smerzi, A., Fantoni, S. y Shenoy, S. R. (1999). *Phys. Rev. A* 59, 620.
- Rebbi, C. y Soliani, G. (1984). *Solitons and Particles*. World Scientific, Singapore.
- Reeves-Hall, P. C. y Taylor, J. R. (2001). *Electron. Lett.* 37, 417-419.

- Reeves-Hall, P. C., Lewis, S. A. E., Chernikov, S. V. y Taylor, J. R. (2000). *Electron. Lett.* 36, 622-624.
- Sackett, C. A., Stoof, H. T. C. y Hulet, R. G. (1998). *Phys. Rev. Lett.* 80, 2031.
- Sakaguchi, H. y Tamura, M. (2004). *J. Phys. Soc. Jpn.* 73, 503.
- (2005). *J. Phys. Soc. Jpn.* 74, 292.
- Salasnich, L., Parola, A. y Reatto, L. (1999). *Phys. Rev.* A60, 4171.
- (2001). *Phys. Rev.* A 64 023601.
- (2002). *J. Phys. B: At. Mol. Opt. Phys.* 35, 3205.
- Serkin, V. N. (2000). “Nonlinear Tunneling of Temporal and Spatial Optical Solitons through Organic thin Films and Polymeric Waveguides”, *Proc. SPIE.* 3927, 292.
- Serkin, V. N. y Hasegawa, A. (2000). “Novel Soliton Solutions of the Nonlinear Schrödinger Equation Model”, *Phys. Rev. Lett.* 85, 4502-4505.
- (2002). “Exactly Integrable Nonlinear Schrödinger Equation Models with Varying Dispersion, Nonlinearity and Gain: Application for Soliton Dispersion Management”, *IEEE J. Selected Topics in Quantum Electron.* 8, 418-478.
- Serkin, V. N. y Belyaeva, T. L. (2007). *Phys. Rev. Lett.* 98, 074102.
- Shin, H. J. (2008). *J. Phys. A: Math. Theor.* 41, 285201.
- Shuryak, E. V. (1996). *Phys. Rev. Lett.* 54, 3151.
- Sias, C. (2007). *Phys. Rev. Lett.* 98, 120403.
- Smerzi, A., Fantoni, S., Giovannazzi, S. y Shenoy, S. R. (1997). *Phys. Rev. Lett.* 79, 4950.
- Stoof, H. T. C. y Stat, J. (1997). *Phys.* 87, 1353.
- Stoychev, K. T., Primatarowa, M. T. y Kamburova, R. S. (2004). *Phys. Rev.* E 70, 066622.
- Strecker, K. E., Partridge, G. B., Truscott, A. G. y Hulet, R. G. (2002). “Formation and Propagation of Matter Wave Soliton Trains”, *Nature (Londres).* 417, 150-151.
- (2003). *New J. Phys.* 5, 73.1-73.8
- Sysoliatin, A. A., Senatorov, A. K., Konyukhov, A. I., Melnikov, L. A. y Stasyuk, V. A. (2007). *Opt. Express.* 15, 16302-16307.
- y Belanov, A. (2008). *IEEE J. Sel. Top. Quantum Electron.* 14, 733-738.
- Tomsovic, S. (1998). *Tunneling in Complex Systems.* World Scientific.
- Travers, J. C., Stone, J. M., Rulkov, A. B., Cumberland, B. A., George, A. K., Popov, S. V., Knight, J. C. y Taylor, J. R. (2007). *Opt. Express.* 15, 13203.
- Tsuzuki, T. (1971). *J. Low Temper. Phys.* 4, 441-457.
- Ueda, M. y Leggett, A. J. (1998). *Phys. Rev. Lett.* 80, 1576.
- Voss, R. F. y Webb, R. A. (1981). “Macroscopic Quantum Tunneling in 1- $\mu\text{m}$  Nb Josephson Junctions”, *Phys. Rev. Lett.* 47, 265.

- Yang, R. y Wu, X. (2008). *Opt. Express*. 16, 17759.
- Zabusky, N. J. y Kruskal, M. D. (1965). "Interaction of 'Solitons' in a Collisionless Plasma and the Recurrence of Initial States", *Phys. Rev. Lett.* 15, 240-243.
- Zakharov, V. E. y Shabat, A. B. (1972). *Sov. Phys. JETP* 34, 62-69.
- Zhao, D., Luo, H. G., Chai, H. Y. (2008) *Phys. Lett. A* 372, 5644–5650.
- Zapata, I., Sols, F. y Leggett, A. J. (1998). *Phys. Rev.* A57, R28.
- Zenesini, A. *et al.* (2008). *New J. Phys.* 10, 053038.
- (2009). *Phys. Rev. Lett.* 103, 090403.



# AMPLIFICACIÓN DE PULSOS ÓPTICOS SOLITÓNICOS\*

*Máximo A. Agüero\*\**

*Tatyana Belyaeva\*\*\**

*Vladimir Serkin\*\*\*\**

## Resumen

Se consideran nuevas posibilidades de amplificación casi ideal de solitones ópticos durante la interacción de pulsos de luz incoherente con un medio de amplificación resonante. En la Parte I, se propone un mecanismo de amplificación de solitones ópticos con una modulación óptima de frecuencia. Se estudia la posibilidad de utilizar la modulación cruzada de fase para producir la necesaria fase inicial de solitones amplificados. En la dispersión Raman estimulada (SRSS), los paquetes de onda solitónicos con espectro desplazado viajan continuamente no sólo en el espacio y tiempo, sino también en el espacio espectral, y se les conoce como solitones coloreados femtosegundos. En la Parte II de este trabajo se considera la influencia de la energía de enlace del solitón en la dinámica de la SRSS suponiendo que la amplitud y duración del efecto túnel solitónico varía en el tiempo cuando el espectro del solitón se aproxima a un obstáculo prohibido de dispersión de velocidad de grupo.

**Palabras clave:** Solitones ópticos, modulación de fase, efecto Raman, efecto túnel solitónico, amplificación de pulso.

\* Los autores agradecen a la Secretaría de Educación Pública (SEP) por el financiamiento al proyecto FEO1/2012 103.5/12/2126 de redes multidisciplinarias, a la SIEA de la Universidad Autónoma del Estado de México y al Conacyt por el apoyo importante en las actividades de investigación.

\*\* Ph. D. en Física-Matemática. Profesor de Tiempo Completo, Facultad de Ciencias, Universidad Autónoma del Estado de México, Instituto Literario 100, Toluca 50000, México. (52) (722) 2965556, <maaguerpg@uaemex.mx>.

\*\*\* Doctora en Física. Profesora de Tiempo Completo, Facultad de Ciencias, Universidad Autónoma del Estado de México, Instituto Literario 100, C.P. 50000, Toluca, Estado de México, México.

\*\*\*\* Doctor en Física. Profesor de Tiempo Completo, Benemérita Universidad Autónoma de Puebla, C.P. 72502, 72001, Puebla, México.

## PARTE I

### INTRODUCCIÓN

Los solitones ópticos son ahora considerados como portadores ideales de datos en la fibra óptica en WDM sistemas de comunicación (Hasegawa, 2000; Dianov, 2000). La técnica solitónica abre nuevas posibilidades para la generación de pulsos ultracortos de luz (Krukov y Letokhov, 1969; Agrawal, 1996; Krukov, 2001, Dianov y Krukov, 2001). El problema de una amplificación ideal de solitones ópticos es un problema bastante antiguo en la óptica no lineal. Hasta hace poco se asumió por primera vez (Blow *et al.*, 1987), que la captura no lineal de energía por un solitón se produce de la siguiente manera: el sistema solitónico se divide en dos partes, en una de ellas, llamada onda solitón, se acumula la energía de acuerdo con el concepto de Zakharov y Shabat (1971), mientras que en la otra parte, llamada dispersión de onda, se acumula la energía de olas redundantes, “derribadas” por un solitón (Hasegawa, 2000; Agrawal, 1996; Blow *et al.*, 1987).

Sin embargo, estudios recientes (Moores, 1996; Khasilev, 1996; Khasiliev *et al.*, 1999; Serkin y Hasegawa, 2000b) cambiaron radicalmente este concepto. El problema de la amplificación ideal en óptica de solitones fue, de hecho, formulado por primera vez en los documentos (Moores, 1996; Khasiliev, 1996), donde se descubrió una posibilidad única de preservar la forma del pulso solitónico en un medio con una ganancia espacial no homogénea.

El modelo considerado en los trabajos de Moores (1996) y Khasiliev (1996), que utiliza una ecuación parabólica no lineal, es equivalente a un modelo basado en la ecuación de Schrödinger no lineal con un potencial auto-consistente de interacción producida por la propagación de un pulso debido a los efectos de auto-acción en un medio cúbico no lineal (Hasegawa, 2000; Agrawal, 1996). El modelo matemático (Moores, 1996; Khasiliev, 1996) y sus “inesperadas” soluciones en variables adimensionales estándares solitónicas tienen la forma de la ecuación (1) (Moores, 1996; Khasiliev, 1996; Khasiliev *et al.*, 1999; Serkin y Hasegawa, 2000a, 2000b).

$$i \frac{\partial \Psi}{\partial Z} + \frac{1}{2} \text{sign}(k_{w^2}) \frac{\partial^2 \Psi}{\partial Z^2} + \alpha |\Psi|^2 \Psi = i\Gamma(Z)\Psi \quad (1)$$

$$\Psi_c(Z, T) = -\frac{\eta\alpha^{-1/2}}{1-2\Gamma(0)Z} \text{Sech} \left[ \frac{\eta T}{1-2\Gamma(0)Z} \right] \times \exp \left\{ -i \frac{T^2\Gamma(0)}{1-2\Gamma(0)Z} - i \frac{\eta^2 Z}{2[1-2\Gamma(0)Z]} \right\} \quad (2)$$

$$\Psi_d(Z, T) = -\frac{\eta\alpha^{-1/2}}{1-2\Gamma(0)Z} \text{Tanh} \left[ \frac{\eta T}{1-2\Gamma(0)Z} \right] \times \exp \left\{ i \frac{T^2\Gamma(0)}{1-2\Gamma(0)Z} - i \frac{\eta^2 Z}{1-2\Gamma(0)Z} \right\} \quad (3)$$

Donde  $\eta$  es una constante, que determina el llamado form-factor de la envolvente de los solitones de Schrödinger,  $\alpha > 0$  es un constante arbitraria; y

$$\Gamma(Z) = \frac{\Gamma(0)}{1-2\Gamma(0)Z}, \quad (4)$$

es la ganancia de un medio activo espacialmente heterogénea;  $\Gamma(0) = 1/(2C)$  y  $C$  es una constante arbitraria. La solución  $\Psi_c(2)$  describe el solitón claro (Moore, 1996; Khasiliev, 1996), mientras que la solución de  $\Psi_c(3)$  describe los solitones oscuros. Se puede ver en la Ecuación (2) que la característica principal de la amplificación de los solitones es la presencia de la modulación de fase (frecuencia chirp). El chirp depende de la distancia propagada por el pulso en el medio amplificador. La ganancia (4) que proporciona la amplificación “ideal” de un solitón es una función hiperbólica de la longitud media de amplificación. El área bajo el impulso de un solitón hiperbólicamente amplificado (2) se conserva (Dianov, 2000) debido a una disminución en su duración que está determinada por una función lineal de la longitud media, lo cual es típico de pulsos solitónicos (Hasegawa, 2000; Agrawal, 1996).

El problema de la amplificación ideal de solitones es muy importante para el desarrollo de la información (Hasegawa, 2000; Dianov, 2000). Una solución radicalmente nueva de este problema se propuso en los trabajos, donde el nuevo concepto de quasi-solitón se introdujo para el solitón chirriante en un medio no lineal con un parámetro de dispersión periódico espacialmente heterogénea y una ganancia periódica (Serkin y Hasegawa, 2000a; Serkin, 2000b; Serkin, 2000c; Kumar, 1997). De esta manera, aparece un quasi-solitón de la ecuación de Schrödinger no lineal con un potencial parabólico (Kumar, 1997), y se inicia un nuevo campo de investigación de rápido desarrollo: solitones ópticos en el espacio de sistemas no lineales no homogéneos y las ondas solitónicas de Bloch en las líneas de comunicación no lineal y láseres solitónicos (Khasiliev *et al.*, 1999; Serkin y Hasegawa, 2002). Se ha demostrado que existe un número infinito de soluciones solitónicas de la ecuación de Schrödinger en sistemas solitónicos activos homogéneos (Serkin y Hasegawa, 2000a; Serkin y Hasegawa, 2000b; Serkin y Hasegawa, 2000c). Además, se han demostrado los teoremas de existencia para solitones “chirp” y se ha desarrollado un método para resolver el problema inverso de la teoría de la dispersión con un parámetro variable espectral (Serkin y Hasegawa, 2002b; Balakrishnan, 1985; Burtsev *et al.*, 1987), así como también los pares de Lax construidos para modelos generalizados

completamente integrables de la ecuación no lineal de Schrödinger (Serkin y Belyaeva, 2001a; Serkin y Belyaeva, 2001b; Serkin y Hasegawa, 2002).

Sin embargo, una serie de problemas fundamentales de la teoría sigue sin resolverse. En primer lugar, la ecuación (1) y su solución (2) plantea nuevos problemas, el más importante de ellos es cómo realizar el experimento correspondiente a la solución (2). En segundo lugar, la ganancia en el modelo (1) se hace infinita en la longitud del medio activo  $Z = 1/[2\Gamma(0)] = C$ . La solución (2) tiene la singularidad misma en el punto  $Z = C$ . Los autores de los trabajos (Moore, 1996; Khasiliev, 1996) no proponen ninguna idea sobre la realización experimental de la ganancia (4) que sea homogénea en toda la longitud ni presentan una forma de hallar la ganancia que sea cercana a la óptima.

Nosotros formulamos y resolvemos el problema de amplificación ideal de un solitón en el marco del modelo (1) y mostramos que la ganancia requerida se puede realizar experimentalmente mediante la fabricación de un amplificador de fibra de dos fotones, que excluye los efectos relacionados con la difracción y el auto-enfocamiento de la radiación. También mostramos que el método de modulación de la frecuencia del pulso inicial juega un papel crucial en la realización experimental de la amplificación ideal del solitón. Se estudia la posibilidad de utilizar el cruce de fase de modulación para obtener la fase inicial de solitones amplificados. Esto se puede lograr si un pulso de control en diferentes longitudes de onda se utiliza simultáneamente con la amplificación de solitón.

Así, en la primera parte de este trabajo (sección I) se analizan los conceptos y ecuaciones fundamentales para estudiar la física de la amplificación de solitones ópticos sin tomar en cuenta el efecto Raman. Las secciones II y III están dedicadas al estudio específico de solitones y a la amplificación singular de solitones de frecuencia modulada, respectivamente. En la parte 2, se estudia la física de amplificación de solitones ópticos tomando en cuenta el efecto Raman; está dividida en dos secciones: la primera explica ciertas peculiaridades de los llamados solitones de color; y la segunda muestra los resultados numéricos y analíticos de dinámica no lineal de solitones de la ecuación no lineal de Schrödinger (ENLS) en relación con ciertos aspectos del tunelaje espectral de solitones Raman. Finalmente, en la última sección se presentan las conclusiones pertinentes.

## I. AMPLIFICACIÓN DE SOLITONES EN PULSOS MULTISOLITÓNICOS

Abordaremos en esta primera parte dos métodos básicos para amplificar solitones ópticos: a) una rápida amplificación no adiabática en línea con los amplificado-

res concentrados y b) la amplificación adiabática en un medio activo distribuido (véanse, por ejemplo, libros y compendios (Agrawal, 1996, Dianov *et al.*, 1992; Nakazawa, 1992 y sus referencias).

En el análisis teórico de la amplificación no adiabática, se supone que la longitud de un amplificador medio es infinitesimal en comparación con la característica longitud de la dispersión de propagación de un pulso. Esto permite considerar la amplificación de un solitón como un proceso “puntual” (delta similar), que equivale a una simple multiplicación de la amplitud del solitón por la ganancia en cada punto del medio. Un pulso solitónico amplificado de esa manera se comporta diferente durante su propagación en el medio. Esta solución puede, sin pasar por el proceso intermedio de las oscilaciones en descomposición, transformarse en un solitón nuevo o formar un estado base solitónico que como es conocido tiene complicadas estructuras periódicas temporales (Nakazawa, 1992). Estos dos hechos bien conocidos en la teoría de solitones Schrödinger se utilizaron en uno de los primeros modelos experimentales de una línea de comunicación solitónica, denominado “línea de comunicación del solitón dinámico” (Nakazawa, 1992), en donde se utilizaron pulsos con la amplitud superior a la amplitud fundamental-solitónica en un factor de 1.2 a 2 (Kriukov, 2001).

En el análisis teórico de la amplificación adiabática, la duración de un medio activo ya no puede ser tratado como infinitesimal, y la ganancia es una función de la actividad-medio de duración, por la sencilla razón de la inevitable presencia de las pérdidas de radiación diferente de cero en una fibra. Tal situación se da, por ejemplo, en el bombeo de solitones Raman (Dianov *et al.*, 1992; Mollenauer *et al.*, 1985; Dianov *et al.*, 1985).

Al analizar el desarrollo de estudios teóricos y experimentales sobre la amplificación de solitones ópticos, podemos señalar que en su conjunto la solución de este problema de amplificación óptima de un solitón resultó en una pérdida. Los experimentos en la generación de solitones ópticos son todavía realizados en sistemas homogéneos no conservativos que, sin embargo, se describen por modelos no integrables. Esto, en principio, no permite amplificar un solitón como un todo manteniendo sus propiedades únicas. El análisis teórico se basa enteramente en la perturbación de la teoría (Karpman y Maslov, 1978) que, como veremos a continuación, resulta ser inválida para una amplificación del solitón en un factor de “e”.

Considere la amplificación del solitón de Schrödinger con más detalle. La propagación de un solitón óptico en un medio activo de resonancia bajo la interacción incoherente del pulso con el medio es descrito por el conocido sistema de ecuaciones: la ecuación de Schrödinger no lineal (ecuación parabólica no lineal) y la ecuación de la polarización P de las partículas del medio activo (Afanasiev *et*

*al.*, 1990). Este sistema en la forma estándar adimensional se puede escribir de la siguiente manera

$$i \frac{\partial \Psi}{\partial Z} + \frac{1}{2} \frac{\partial^2 \Psi}{\partial \tau^2} + |\Psi|^2 \Psi = \frac{G}{2} P \quad (5)$$

$$\gamma_a \frac{\partial P}{\partial \tau} + P(1 + i\gamma_a \Delta\Omega) = i\Psi \quad (6)$$

Aquí, el tiempo  $\tau$  está renormalizado con la duración del pulso inicial  $\tau_0$  a la entrada del medio activo, la distancia propagada del solitón se normaliza a la dispersión de longitud de la propagación de impulsos  $L_D = \tau_0^2 / |K_w''|$  y la amplitud de la onda se normaliza a los valores característicos iniciales  $|\Psi_s| = [8\pi k_w'' \times (\tau_0^2 k_0 n_0 c n_0)^{-1}]^{1/2}$ ;  $\gamma_a = T_2 / \tau_0$ ;  $T_2$  el tiempo de relajación transversal del *momentum* del dipolo de alta frecuencia,  $\Delta\Omega = \tau_0 (w_0 - w_{12})$  es la frecuencia de la portadora del pulso, y  $w_{12}$  es la frecuencia de resonancia de la transición. El parámetro de ganancia en la ecuación (5) se determina por la relación  $G = L_d / L_a$  (Dianov *et al.*, 1982), donde  $L_a = 1 / \sigma_0 N_0$  es la longitud característica de ganancia;  $N_0$  es la inversión de población-densidad en la ausencia del campo, y  $\sigma_0$  es la sección transversal de la transición radiactiva en la frecuencia de resonancia.

De la interacción incoherente de un pulso coherente con un medio, la polarización sigue cuasi-estáticamente al campo, y la contribución a la dispersión del índice de refracción es (Kriukov y Letokhov 1969):

$$\delta n_1 = \frac{cN_0}{2n_0 w} \frac{\sigma_0 (T_2 \Delta w + i)}{1 + (T_2 \Delta w)^2} \quad (7)$$

donde  $\Delta w = w_0 - w_{12}$ . Si la frecuencia portadora del pulso coincide con la frecuencia de la resonancia de transición, el índice de refracción cuasi-estática se convierte en imaginario puro. Se sabe que el modelo matemático (5), (6) describe correctamente la dinámica de amplificación de solitones en el rango de picosegundos, cuando el efecto no lineal electrónico de Kerr, en relación con la adición no inercial al índice de refracción  $n_2$  domina. Si  $G \ll 1$  entonces la ecuación (5) puede ser resuelta utilizando el método de perturbación adiabática en teoría de solitones (Karpman y Maslov, 1978; Karpman y Soloviev, 1981).

En el caso de la amplificación incoherente, cuando el ancho del espectro del pulso es pequeño en comparación con el ancho de la línea de ventaja (Kriukov y Letokhov, 1969) el sistema de ecuaciones (5), (6) se puede transformar en

$$i \frac{\partial \Psi}{\partial z} + \frac{1}{2} \frac{\partial^2 \Psi}{\partial \tau^2} + |\Psi|^2 \Psi = i \frac{G}{2} \left( \Psi - \gamma_a \frac{\partial \Psi}{\partial \tau} + \gamma_a^2 \frac{\partial^2 \Psi}{\partial \tau^2} \right) \quad (8)$$

La solución solitónica de la ecuación (8) con parámetros que cambian lentamente tiene las formas

$$\Psi_s(z,t) = \eta(z) \sec h\{\eta(z) [\tau - \tau_c(z) - V(z)z]\} \exp(i\Phi) \quad (9)$$

$$\Phi = -V(z) [\tau - \tau_c(z)] + i[V^2(z) - \eta^2(z)]z/2 + i\Phi_0(z) \quad (10)$$

$$\frac{d\eta}{dz} = G \left(1 - \frac{1}{3} \eta^2 \gamma_a^2 - \gamma_a^2 V^2\right) \eta \quad (11)$$

Siendo  $\tau_c(z)$  la coordenada del centro de gravedad del pulso y  $\Phi_0$  la fase. Para pasar a la ecuación que se acerca al modelo inicial (1), ponemos  $y = 0$ . Entonces, el sistema (5), (6) se transforma en la ecuación conocida no lineal de Schrödinger con amplificación (de absorción)

$$i \frac{\partial \Psi}{\partial z} + \frac{1}{2} \frac{\partial^2 \Psi}{\partial \tau^2} + |\Psi|^2 \Psi = \frac{G}{2} P \quad (13)$$

La teoría de perturbación de solitones (Karpman y Maslov, 1978; Karpman y Soloviev, 1981) da la solución de la ecuación (13) en la forma (Hasegawa, 2000; Agrawal, 1996).

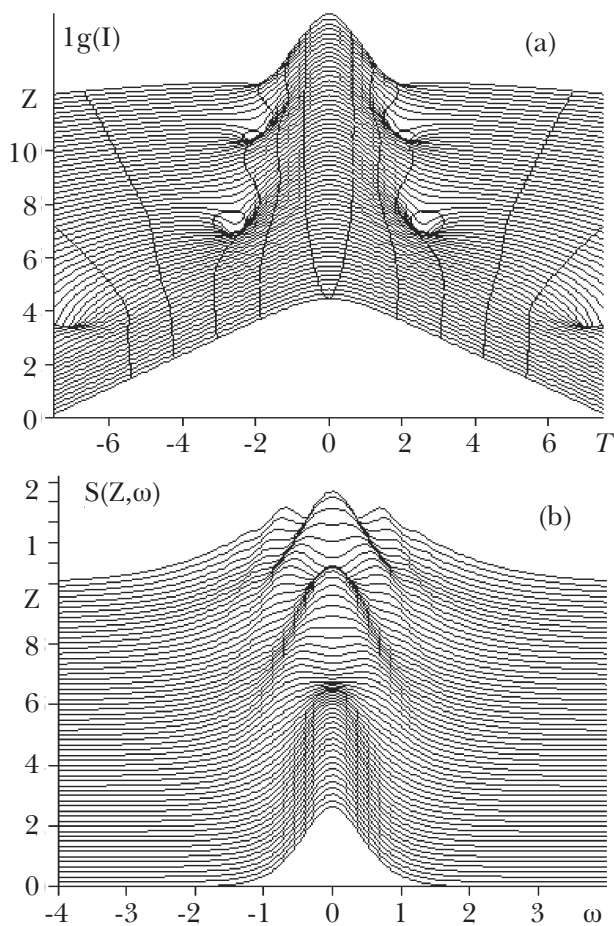
$$\Psi(z,t) = \eta(z) \text{Sech}[\eta(z)\tau] \exp[i\sigma(z)] + O(G) \quad (14)$$

$$\eta(z) = \Psi(0) \exp(2Gz)$$

$$\sigma(z) = \frac{\Psi^2(0)}{8G} [1 - \exp(4Gz)]$$

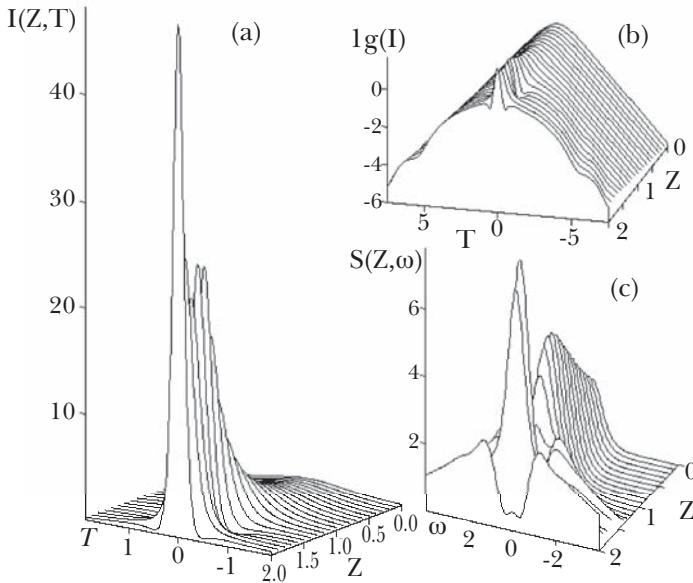
La solución (14) ha sido tratada por muchos años (véase, por ejemplo, el libro de Agrawal [1995] y sus referencias) como una aproximación bastante exacta de la descripción de la amplificación de solitones. Sin embargo, nuestros experimentos computacionales demostraron que este no tiene lugar. Hay que recordar que, como ya se muestra en el artículo (Dianov *et al.*, 1986), las distorsiones de la forma del solitón en un medio lineal de absorción son causadas por la aparición de la modulación de frecuencia del pulso, que sustancialmente restringe la aplicación de los métodos de la teoría de perturbación. Vamos a demostrar que un resultado similar se obtiene para una amplificación lineal de solitones, y que existen diferencias sustanciales cualitativas y cuantitativas con el caso de propagación de un solitón en un medio de absorción lineal (Dianov *et al.*, 1986).

Se realizó el estudio numérico de la propagación y la interacción de solitones en un medio activo, teniendo en cuenta las diversas relaciones entre los parámetros del modelo (13). Cuando los resultados de la teoría de perturbación en solitones y en la expresión (14) llegan a ser inválidos, se encontró que existe una energía límite del solitón amplificado. Un solitón no se amplifica como un todo en el régimen de amplificación incoherente: durante la amplificación de solitones, su forma y espectro cambian sustancialmente.



**Figura 1.1.** (a) La dependencia de la forma  $I(z, t)$  de un impulso solitón (proyección ortográfica en una escala logarítmica, la vista superior en un ángulo de  $60^\circ$ ) y (b) la forma  $S\{Z, \omega\}$  de su espectro en el medio activo longitud  $Z$  en amplificación adiabática de solitones y  $G = 0,1$ . Las curvas gruesas son geodésicas de igual nivel de líneas para  $\log I = 0, -1, -2, -3, -4$ .

Durante la amplificación de un solitón, dos características aparecen en su estructura temporal: un pico intenso angosto y un pedestal de baja intensidad amplia al cual se transfiere una fracción cada vez mayor de energía del pulso en el proceso de amplificación. Un cuadro detallado de la amplificación de solitones se muestra en las Figuras I.1 y I.2, donde los resultados obtenidos en la amplificación en regímenes adiabáticos y no adiabáticos se comparan. El espectro de los solitones amplificados tiene una estructura, lo que demuestra la aparición de una modulación de frecuencia considerable de la radiación.

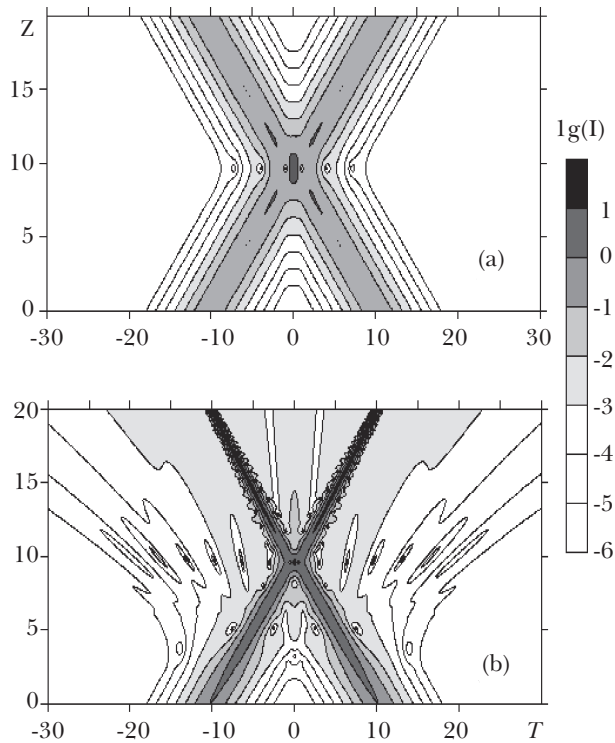


**Figura I.2.** La dependencia de las formas de  $I(Z, T)$  (a, b) de un pulso solitónico y la forma  $S(Z, \omega)$  (c) del espectro del pulso en un medio activo de longitud  $Z$  bajo una amplificación no adiabática y  $G = 1.0$ .

El procesamiento de los resultados de experimentos numéricos sugiere la existencia de una ganancia máxima en la cual los solitones ya no son estables durante la amplificación. Esto ocurre cuando la energía de los solitones aumenta en un factor de “e”, independientemente de la forma como se acumula energía en los solitones [amplificación adiabática (Figura I.1) o no adiabático (Figura I.2)].

Cálculos detallados que se realizan en un amplio rango de ganancia solitónica muestran que la inestabilidad estructural de solitones se explica por el desequilibrio en el proceso de transformación de modulación de frecuencia a modulación de amplitud. Todas estas características se muestran en escala logarítmica en las Figuras I.1 y I.2.

Por tanto, los regímenes adiabático y no adiabático de amplificación incoherente de solitones en un medio activo con parámetros constantes están siempre acompañados por la aparición de la inestabilidad estructural de solitones y el crecimiento del componente no solitónico del campo durante la amplificación. La región de aplicabilidad de la teoría de perturbación está determinada por la desigualdad  $G \ll 1$ , que expresa, de hecho, la condición de reordenamiento adiabático del solitón, que adquiere nuevos valores para su amplitud y su duración correspondientes a una energía mayor.



**Figura I.3.** Interacción de dos solitones opuestos (a) en ausencia de amplificación Y (b) bajo la amplificación adiabática para la desafinación de la velocidad de grupo  $\Delta V = 2.0$  and  $G = 0.1$ .

Debido a la amplificación imperfecta de solitones y al crecimiento del componente no solitónico del campo, los solitones dejan de interactuar elásticamente entre sí, como se muestra en la Figura I.3, donde el mapa de contorno (líneas a la misma altura), se presenta para dos solitones interactuando y acercándose el uno al otro. Los cálculos se realizaron para dos solitones ideales, cuando el parámetro  $G=0$ , y para solitones linealmente amplificados descritos por el modelo más simple (13).

## II. AMPLIFICACIÓN SINGULAR DE SOLITONES CON FRECUENCIA MODULADA

Como se mencionó anteriormente, la opinión aceptada de que los solitones ideales de Schrödinger no pueden ser formados en un medio activo con amplificación incoherente, fue revisado en los trabajos de Moores (1996) y Khasiliev (1996). Los autores de estos trabajos mostraron que existe una posibilidad de amplificar al solitón en su conjunto, si la ganancia es una función singular hiperbólica de longitud media activa, mientras que la fase del solitón en la entrada al medio es una función parabólica del tiempo. La interacción de tales solitones frecuencia-modulados se torna totalmente elástica cuando la fase solitónica y la ganancia se hacen auto-consistentes. En los trabajos (Serkin y Hasegawa, 2000a; 2000b; 2000c; 2002) se reportó un método regular de búsqueda de soluciones en forma de solitones de frecuencia modulada. Se demostró que existen cuatro soluciones solitónicas representantes del solitón oscuro hiperbólicamente amplificado en decaimiento para ganancias iniciales positivos o negativos (4). En sentido estricto, las soluciones presentadas en Moores (1996) y Khasiliev (1996) son casos particulares de la aplicación del segundo teorema expuesto en Serkin y Hasegawa (2000) donde el “wronskiano” de dos funciones constantes: la dispersión [ $D(Z) = 1$ ] y la no linealidad [ $R(Z) = 1$ ] son exactamente iguales a cero [ $W(R, D) = 0$ ].

Seguidamente demostramos que la amplificación singular de los solitones modulados por frecuencia puede ser descrito por un modelo equivalente sin singularidades. Para ello hagamos un cambio de variables en (1):

$$\Psi = a(Z)u, \quad a(Z) = \exp \int_0^Z r(Z') dZ',$$

$$z' = \int_0^z a^2(Z'') dZ''$$

En este caso, la ecuación de Schrödinger tendrá la siguiente forma

$$i \frac{\partial u}{\partial Z'} + \frac{D_{opt}(Z)}{2} \frac{\partial^2 u}{\partial T'^2} + \alpha |u|^2 u = 0, \quad (15)$$

Donde

$$D_{opt}(Z) = \frac{D_2(0)}{a_{opt}^2(\Gamma_{opt}(Z))}$$

Es la función de dispersión óptima. Usando las relaciones

$$\Gamma_{opt}(Z) = \frac{1}{2(C-Z)}, \quad a_{opt}^2(Z) = \frac{C}{2(C-Z)},$$

se obtiene

$$D_{opt}(Z) = D_2(0) \left(1 - \frac{Z}{C}\right), \quad z' = -C \ln \frac{C-Z}{C},$$

$$Z = C \left[1 - \exp\left(-\frac{z'}{C}\right)\right].$$

Después de simples transformaciones, la Ecuación (15) toma la forma

$$i \frac{\partial u}{\partial z'} + \frac{1}{2} D_2(0) \exp\left(\frac{z'}{C}\right) \frac{\partial^2 u}{\partial T'^2} + \alpha |u|^2 u = 0. \quad (16)$$

Por tanto, el modelo (1) se transforma al modelo (15) en nuevas variables y sus soluciones (2) y (3) para solitones oscuros y claros se transforman en las soluciones

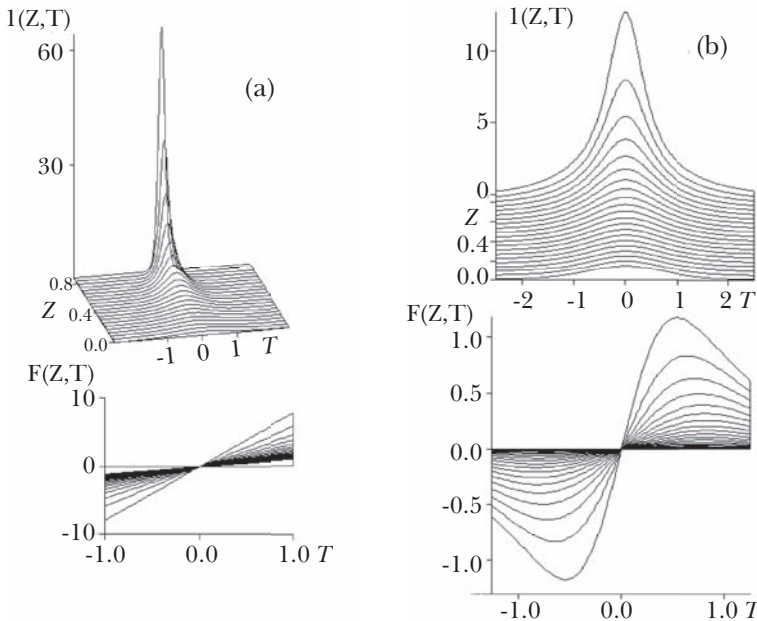
$$u_c(z', T') = -\eta \alpha^{-1/2} \exp[\Gamma(0)z']$$

$$x \operatorname{sech} \left\{ \eta T' \exp[2\Gamma(0)z'] \right\} \exp \left\{ -iT' 2\Gamma(0) \exp[2\Gamma(0)z'] \right.$$

$$\left. - i \frac{1}{2} \eta^2 z' \exp[2\Gamma(0)z'] \right\} \quad (17)$$

$$u_d(Z', T') = \eta \alpha^{-1/2} \exp[\Gamma(0)Z'] \times \tanh \left\{ nT' \exp[2\Gamma(0)Z'] \right\} \exp \left\{ iT'^2 \Gamma(0) \exp [2\Gamma(0)Z'] - in^2 Z' \exp [2\Gamma(0)Z'] \right\} \quad (18)$$

Estas soluciones de gorjeo difieren en dos aspectos de los solitones canónicos del modelo no lineal de la ecuación de Schrödinger sin amplificación o absorción. En primer lugar, los solitones (2), (3) y (17), (18) son pulsos de frecuencia modulada cuya amplitud, duración y frecuencia modulada son descritos por la misma función de  $Z$ —una función hiperbólica con respecto a la longitud de amplificación del medio—. En segundo lugar, los solitones (2) y (3) son soluciones singulares de la ecuación (1) y tienen significado solamente para  $Z < C$ . Cuando la longitud de amplificación media se aproxima a la singularidad  $Z=C$ , la amplitud del soli-



**Figura II.4.** Dinámica de amplificación de un solitón óptico en un medio con una ganancia de crecimiento hiperbólico: (a) un solitón con frecuencia modulada amplificada de la ecuación (1), cuya envolvente  $I(z, t)$  y la modulación de frecuencia  $F(z, t)$  se calculan para  $G = 1,0$  y la longitud activa-medio restringido por la condición de acercamiento a una singularidad por el valor  $Z - C = 0,125$  y (b) un cambio en la envolvente y la aparición de la modulación de frecuencia del pulso en la amplificación en un medio inhomogéneo en la ausencia de modulación inicial del pulso para  $G = 0$ .

tón aumenta infinitamente y su duración disminuye. Sin embargo, las soluciones de (2) y (3) son de cierto interés práctico, debido a la longitud activa del medio puede estar siempre restringido por la condición inalcanzable de la singularidad (Khasiliev, 1996).

Una ley lineal de modulación de frecuencia del pulso en la entrada del medio activo, determina todo el régimen subsecuente de compresión de pulsos en el medio.

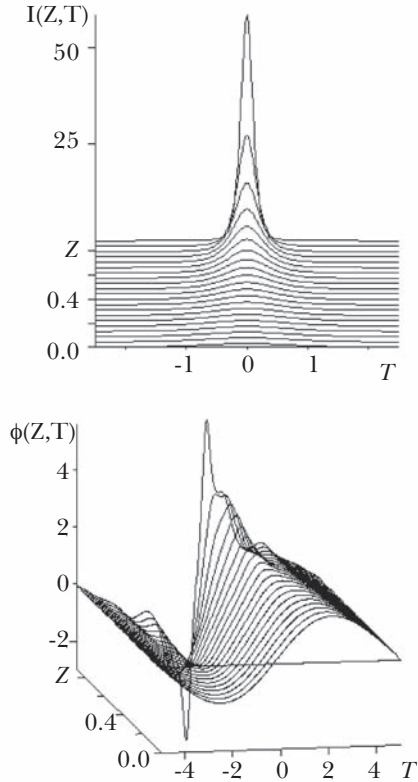
La Figura II.4a muestra las simulaciones por ordenador de la amplificación de un solitón de frecuencia modulada. La compresión de este solitón no cambia la ley de modulación de frecuencia, sólo su valor ha de ser cambiado vía la expresión (2). Por el contrario, la amplificación de un solitón común (sin frecuencia modulada) se acompaña de la aparición de la modulación de frecuencia no lineal, que se describe por la envolvente del pulso amplificado como se muestra en la Figura II.4b.

En la actualidad, se han desarrollado una variedad de métodos para la modulación de la luz (Katys *et al.*, 1967; Mustel y Parygin, 1970) y para diferentes sistemas de fibra óptica, en las que prácticamente cualquier perfil de dispersión puede ser realizado (Lenz y Eggleton, 1998). En este sentido, la principal dificultad de detección experimental de solitones de frecuencia modulada (2) es la necesidad de producir modulación de frecuencia lineal de radiación en la entrada del medio. Vamos a suponer que la modulación de la fase inicial de un pulso se produce debido a la modulación de fase cruzada y tiene la forma

$$\Phi(Z = 0, \tau) = A \exp \left[ - \left( \frac{\tau}{\tau_{\text{mod}}} \right)^{2n} \right] \quad (19)$$

De acuerdo con (19), la frecuencia depende linealmente del tiempo sólo en la parte central del pulso (2). El experimento numérico permite a uno estudiar en detalle la dependencia del grado de compresión del pulso con respecto a la magnitud de modulación de frecuencia en la entrada al medio. Variando la duración del control del pulso super-gaussiano (19), que se propaga en la longitud de onda diferente, podemos ampliar la región de escaneo de la frecuencia lineal en el pulso solitónico. Como es de esperar, el solitón se comprime de manera eficiente sólo en la parte del pulso inicial en la que la dependencia del tiempo con respecto a la frecuencia difiere ligeramente de una ley lineal.

Los resultados típicos de simulaciones por ordenador se presentan en las Figuras II.5 y II.6.

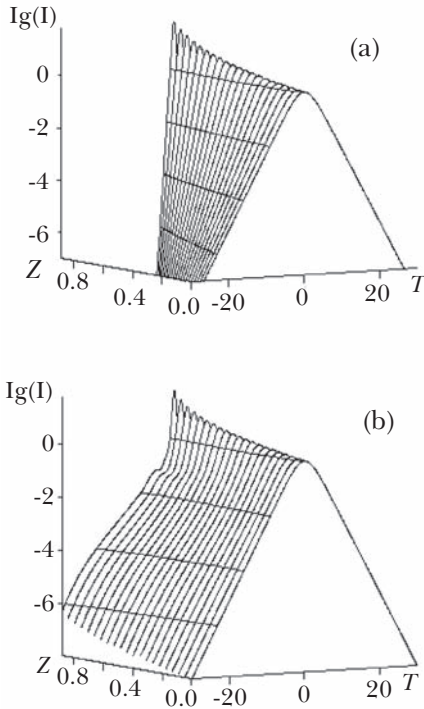


**Figura II.5.** La amplificación de un solitón modulado en fase ( $Z, T$ ) en un medio activo con la ganancia hiperbólica en el caso cuando la fase inicial de modulación de  $\Phi(Z, T)$  se produce debido a la modulación de fase cruzada por el pulso en la longitud de onda no-resonante (19) con una duración de  $\tau_{\text{mod}} = 5.0$ , amplitud  $A = 1.0$ , y el parámetro  $n = 1$ .

Uno puede ver que no puede ocurrir la auto-compresión del pulso como un todo. Este hecho queda claramente demostrado por la dependencia de la forma del pulso con respecto a la longitud de ganancia como se muestra en la Figura II.6. La auto-compresión y la compresión ocurren sólo en la parte superior del pulso, mientras que la duración pedestal no cambia. Vamos a comparar estos resultados con la Figura II.4, donde una solución exacta se presenta a la misma escala, que, por el contrario, demuestra la compresión del solitón de frecuencia modulada como un todo.

Consideremos ahora cómo pueden ser perfectamente ampliados los solitones en un experimento real. Dianov *et al.* (1983) demostraron que la absorción de dos fotones (y la amplificación Raman) puede cambiar la forma del pulso solitónico de tal manera que su amplitud (y por tanto, la duración) variará como

$$\eta^2(Z) = \frac{\eta^2(0)}{1 \pm \left(\frac{8}{3}\right) \eta^2(0)z} \tag{20}$$

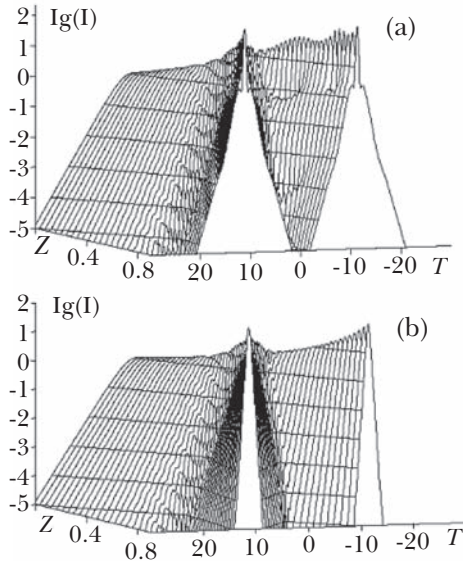


**Figura II.6.** Comparación de la redistribución de energía relativa durante amplificación de solitones por (a) un perfil parabólico perfecto de la fase inicial en (2) y (b) para la fase inicial especificada por el perfil temporal de un impulso de control (19).

Es obvio que, si tomamos el signo menos en el denominador de (20), se obtiene una función singular coincidiendo con la expresión (4) para la ganancia. Por tanto, la respuesta a la pregunta anterior formulada es obvia: es necesario crear un amplificador de fibra óptica de dos fotones.

Para ser coherente hasta el final, vamos a estudiar las propiedades de la interacción de solitones de frecuencia modulada durante su amplificación. El objetivo de los experimentos de nuestro equipo es el de dilucidar la posibilidad de existencia de un solitón estático atractor, lo cual provocaría la interacción inelástica y la fusión de solitones.<sup>1</sup> Tenga en cuenta que la dispersión inelástica es una característica de sistemas no integrables.

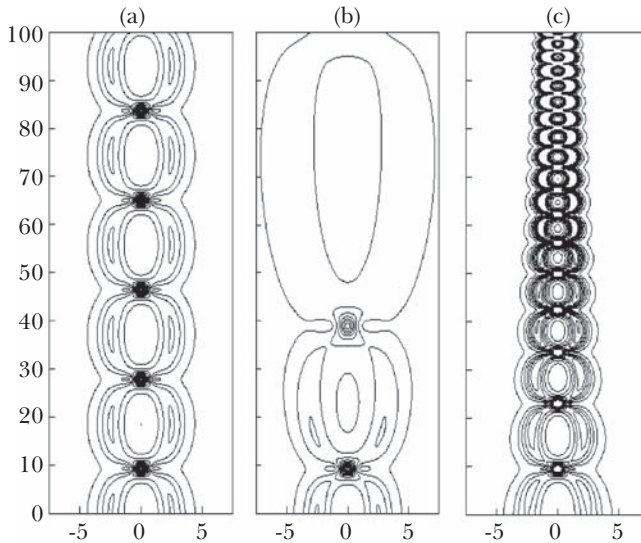
Consideremos, por ejemplo, la situación en la que los solitones amplificados por un factor de 100 (régimen de fuerza no adiabático). Las velocidades de grupo diferentes de los solitones significan que éstos inicialmente se sometieron a WDM. Supongamos que los espectros de los impulsos iniciales no se superponen. Si los solitones están amplificados en un sistema con parámetros constantes, entonces la distribución de energía cambia sustancialmente en el tiempo (Figura II.7).



**Figura II.7.** Comparación de la dinámica de la interacción de solitones por un régimen de amplificación fuertemente no adiabático (amplificación por un factor de 100) en un medio activo con (a) la ganancia que es constante durante la longitud media y (b) una ganancia hiperbólicamente creciente por un espectro no solapado de solitones que interactúan y  $\Delta V = 10.0$ .

<sup>1</sup> Segev y Stegeman (1998) aseguran erróneamente en su manuscrito que Gatz y Herrmann (1992) fueron los primeros en descubrir la dispersión inelástica de solitones en el modelo de Schrödinger no lineal con saturación. Notemos que en realidad el efecto de dispersión inelástica

En este caso, la interacción entre solitones es inelástica. Si por el contrario, los solitones se amplifican en un sistema con ganancia homogénea, su interacción se torna completamente elástica (Figura II.7b). Supongamos ahora que el espectro de solitones iniciales está parcialmente superpuesta (Figura II.8). La interacción de solitones en el medio amplificador no homogénea todavía es elástica (Figuras II.8a y II.8b). Para comparación, la Figura II.8c muestra la interacción entre solitones ideales en la ausencia de la amplificación.



**Figura II.8.** (a, b) Interacción de solitones amplificados hiperbólicamente y (c) solitones ideales con espectros fuertemente superpuestos en la ausencia de amplificación para  $\Delta V = 0.5$ .

Notemos que el problema de la amplificación solitónica en un sistema espacialmente heterogéneo es bien conocido por los experimentadores desde hace tiempo (Hasegawa, 2000; Agrawal, 1996; Dianov *et al.*, 1992; Nakazawa, 1992).

---

de solitones de Schrödinger resultando en su fusión en un solo pulso y llamado solitón atractor en el modelo con saturación fue primeramente descubierto por Zakharov *et al.* (1998) y por Nikonova y Serkin (1990), quienes probablemente fueron los primeros en aplicar el concepto de solitón atractor a fenómenos de óptica no lineal. Ellos usaron el concepto del modelo de Schrödinger no lineal con saturación y calcularon la dispersión inelástica de los solitones de Schrödinger y las condiciones de fusión de sus estados base que aparecen en las intensidades de los pulsos solitónicos cuando se toman en cuenta los términos superiores en la expansión en la polarizabilidad no lineal (Agrawal, 1996).

Por otra parte, durante el intercambio de energía entre ondas no lineales, una situación similar a la que aquí es considerada, siempre se realiza. De hecho, podemos considerar, por ejemplo, transferencia de energía espacialmente heterogénea durante la generación intracavitaria de pulsos de láser ultracortos y solitones de dispersión Raman estimulada bajo la conversión de frecuencia de cascada en fibras ópticas (Hasegawa, 2000; Agrawal, 1996; Dianov *et al.*, 1992; Nakazawa, 1992; Mollenauer *et al.*, 1985; Dianov *et al.*, 1985; Karpman y Maslov, 1978; Afanasiev *et al.*, 1990; Dianov *et al.*, 1986). La ganancia heterogénea cercana a la ganancia requerida hiperbólica puede ser obtenida si el pulso de bombeo por ejemplo, durante la amplificación Raman de solitones representa un pulso multi-solitónico, que está sometida a una auto-compresión tipo avalancha durante la propagación en un medio en la fase inicial de la transferencia de energía hacia el pulso de Stokes. Como se muestra en Mollenauer *et al.* (1983) y Dianov *et al.* (1986), el grado de compresión de un pulso multi-solitónico es una función hiperbólica del número de solitones en el pulso, mientras que un aumento en la potencia máxima con la longitud de la fibra se describe por una función cercana a la función hiperbólica necesaria. La ganancia requerida puede ser obtenida también mediante el uso de fibras cónicas especiales (véase la revisión Dianov y Kriukov, 2001) y los documentos (Semenov *et al.*, 1995; Bogatyriov *et al.*, 1995; Richardson *et al.*, 1995; Birks *et al.*, 2000; Liu *et al.*, 2001; Knight *et al.*, 1996).

En la auto-compresión del solitón en un medio activo en el caso de pulsos femtosegundos es necesario tener en cuenta las aproximaciones más altas de la teoría de dispersión y el cambio de frecuencia Raman (Serkin *et al.*, 2001; Yong-Xin *et al.*, 1985; Mitchke y Mollenauer, 1986; Serkin y Belyaeva, 2001a) que aparecen debido a la auto-dispersión Raman (Dianov *et al.*, 1992). Tendremos en cuenta en otras contribuciones a uno de los principales factores limitantes que es la auto-conversión Raman de la frecuencia solitónica para la amplificación óptima de solitones femtosegundos.

## PARTE II

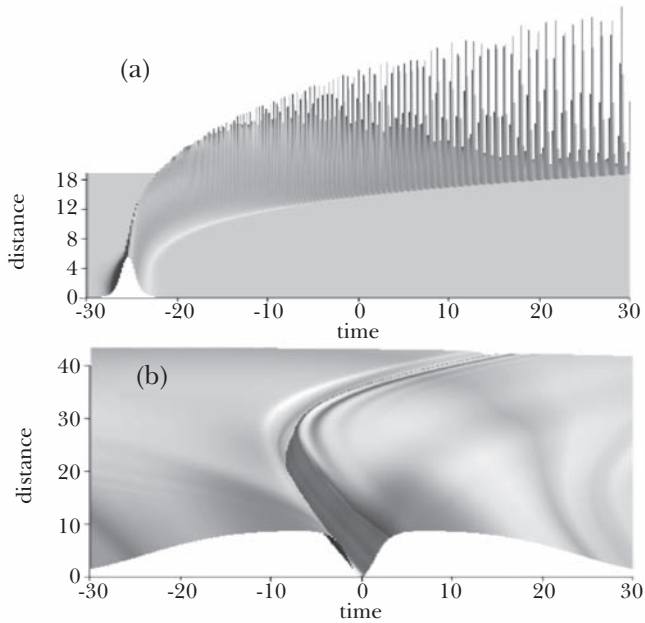
### III. SOLITONES DE COLOR

El concepto fundamental de solitones de color no autónomos en sistema no lineal y dispersivo con diferentes valores de dispersión, no linealidad y la disipación o ganancia se introdujo en los trabajos de Serkin y Belyaeva, 2001; Serkin, Hasegawa y Belyaeva, 2010; Ponomarenko y Agrawal, 2006; Wu *et al.* 2009; Porzesian *et al.*, 2007; Zhao *et al.*, 2008; Luo *et al.*, 2009; He *et al.*, 2009; Tenorio *et al.*, 2005, 2007 y Serkin *et al.*, 2010a, 2010b). Estos solitones no autónomos amplían

el concepto de solitones clásicos y generalizan la gran cantidad de solitones que interactúan elásticamente y se mueven con diferentes amplitudes, velocidades, espectros y se adaptan tanto a los potenciales externos y a las variaciones de dispersión y no linealidad.

El descubrimiento de Raman estimulando la auto-dispersión (SRSS) de solitones femtosegundos es reconocido como uno de los logros más notables de la fibra de óptica no lineal (Dianov *et al.*, 1985; Hasegawa, 1989; Mitchke y Mollenauer, 1986; Dianov *et al.*, 1992; Serkin y Belyaeva, 2001a, 2001b; Kibler *et al.*, 2007; Tsoi y De Sterke, 2007; Poletti *et al.*, 2008; Blow, 1989; Afanasiev *et al.*, 1990; Serkin *et al.*, 2003; Serkin y Hasegawa, 2000; Serkin *et al.*, 1998; Dudley *et al.*, 2006; Serkin *et al.*, 2007, 2010a, 2010b; Travers *et al.*, 2007; Liu *et al.*, 2001; Sysoliatin *et al.*, 2007, 2008). Este efecto también se llama a menudo intrapulse dispersión Raman estimulada (ISRS), o cambio de auto-frecuencia del solitón, haciendo énfasis en el régimen inusual de dispersión Raman estimulada, cuando el espectro de una alta potencia de pulso láser ultracorto resulta ser tan amplia que cubre la banda de las resonancias de Raman del medio. En este caso, el componente espectral de Stokes del campo desplazado por la frecuencia de las vibraciones moleculares está contenida dentro del pulso de la bomba en sí. La amplificación de los componentes de baja frecuencia de Stokes en el campo de los componentes anti-Stokes de alta frecuencia espectrales del mismo solitón resulta en un desplazamiento al rojo del espectro. Por ejemplo, el desplazamiento hacia el rojo de la frecuencia portadora en un sistema pasivo de fibra óptica es inversamente proporcional a la cuarta potencia de la duración del pulso y se describe por expresión (Ponomarenko y Agrawal, 2007)  $\frac{df}{dz} = 0.082/\tau_0^4$  donde  $\tau_0$  está dado en picosegundos y  $df/dz$  en  $THz/km$ . Si  $\tau_0 = 0.1 ps$ , el cambio al “rojo” cambio de frecuencia alcanza  $820 THz/km$ . A estos extraordinarios paquetes de ondas tipo solitónicos, que viajan no sólo en el espacio-tiempo, sino también en el espacio espectral, se les conocen como solitones de color. Los solitones de color juegan un papel importante en la generación solitón supercontinuo. Las típicas dinámicas de solitones de colores brillantes y oscuras se muestran en la Figura III.1.

El enigma más intrigante de solitones ópticos de color está conectado con la posibilidad de su tunelaje en el dominio espectral por medio de una barrera de potencial espectral inhomogénea de la dispersión de la velocidad de grupo (DVG) incluyendo la banda prohibida de la DVG positiva (Mitchke y Mollenauer, 1986; Serkin, 1987; Dianov *et al.*, 1992; Serkin *et al.*, 2003; Serkin y Belyaeva, 2001a, Kibler *et al.*, 2007; Tsoi y De Sterke, 2007; Poletti *et al.*, 2008).



**Figura III.1.** Dispersión Raman Típica evaluado para solitones brillantes y oscuras de la ecuación no lineal de Schrödinger.

Con el fin de observar experimentalmente el efecto SST, se requiere la presencia de una fibra con una estrecha región normal de DVG rodeado por dos regiones anómalas de DVG. Recientemente, Dudley *et al.* (2006) presentaron un estudio numérico de tunelaje espectral del solitón en una fibra de cristal fotónico con un defecto de sub-longitud de onda central en su núcleo de aire. Específicamente, para un diseño de la fibra donde dos regiones de dispersión anómala soporta una región ancha de 90 nm de dispersión normal, las simulaciones muestran un tunelaje desde 1300 hasta 1475 nm de todo el régimen de dispersión normal prohibido, con una eficiencia cerca de la unidad. Poletti, Horak y Richardson (2008) diseñaron fibras con número de índice de guiado con estructura relativamente simple que poseen propiedades dispersivas adecuadas para la observación del tunelaje espectral del solitón. Sus simulaciones sugieren que un túnel a través de una región de dispersión normal de la anchura de 150 nm cuando se bombea cercano al infrarrojo, es en principio posible utilizando sólo unos pocos metros de estas fibras.

#### IV. DINÁMICA NO LINEAL DE SOLITONES DE LA ECUACIÓN NO LINEAL DE SCHRÖDINGER (ENLS) EN TRAMPAS DEPENDIENTES DEL TIEMPO

En esta sección, se considera la influencia de la energía de enlace del solitón en la dinámica de efecto SST cuando la amplitud y la duración del tunelaje solitónico varía con el tiempo y, además, el espectro del solitón se aproxima a una barrera prohibida de DVG.

Se considera el efecto SRSS en el marco de la ENLS

$$i \frac{\partial q(z, \tau)}{\partial z} = \frac{1}{2} \frac{\partial^2 q(z, \tau)}{\partial \tau^2} + (1 - \beta) |q(z, \tau)|^2 q(z, \tau) + \beta Q(z, \tau) q(z, \tau) \quad (20)$$

para la amplitud compleja de la envolvente del pulso láser  $q(z, \tau)$  acoplado con el modelo de oscilador del efecto de la dispersión Raman estimulada (Wu *et al.*, 2009; Zhao *et al.*, 2008; Serkin *et al.*, 2001; Serkin *et al.*, 2001b; Serkin y Belyaeva, 2001a; Tenorio *et al.*, 2005).

$$\mu^2 \frac{\partial^2 Q(z, \tau)}{\partial \tau^2} + 2\gamma\mu \frac{1}{2} \frac{\partial Q(z, \tau)}{\partial \tau} + Q(z, \tau) = |q(z, \tau)|^2 \quad (21)$$

En las ecuaciones (20) y (21),  $Q(z, \tau)$  representa la amplitud real de la onda de las vibraciones moleculares, el tiempo que transcurre  $\tau = (t - z/v_0)/\tau_0$  se normaliza a la duración del impulso inicial  $\tau_0$ ;  $V_0$  es la velocidad de grupo; la distancia  $z$  se expresa en términos de la longitud de la dispersión. La Ecuación (21) describe una onda de vibraciones moleculares y contiene dos parámetros adimensionales  $\mu$  y  $\gamma$  que están relacionados con la duración inicial  $\tau_0$  del pulso, la frecuencia de vibración molecular  $\Omega_R = 2\pi/T_R$ , y la anchura de la línea Raman  $T_{2R} = 1/(\pi C \Delta v_R)$  por las siguientes expresiones:  $\mu = (\tau_0 \Omega_R)^{-1}$  y  $\gamma = (T_{2R} \Omega_R)^{-1}$ . Aquí  $T_R = 2\pi/\Omega_R$  es el periodo de vibraciones de resonancia molecular y  $T_{2R}$  es su tiempo de desfase. La contribución relativa de la respuesta molecular a la no linealidad total media (en relación con el mecanismo de Kerr puro no lineal electrónico  $n_2$ ) se describe por el parámetro  $\beta = g_R \gamma / (k n_2)$ , donde  $g_R$  es la ganancia Raman en el centro de la línea de ganancia. Cuando el parámetro  $\beta = 0$ , el sistema de ecuaciones (20) y (21) puede ser transformado a una ecuación de Schrödinger no lineal convencional para solitones canónicos. En los casos de importancia práctica, la contribución Raman a la susceptibilidad no lineal actúa como una pequeña perturbación  $\sigma = 2\beta \rho \gamma \mu = \tau_R / \tau_0$ , que es la razón de

$$Q(z, \tau) = |q(z, \tau)|^2 - 2\gamma\mu \frac{\partial}{\partial \tau} |q(z, \tau)|^2 \quad (22)$$

La Ecuación (22) conduce al modelo bien conocido del efecto SRSS

$$i \frac{\partial q(z, \tau)}{\partial z} = \frac{1}{2} \frac{\partial^2 q(z, \tau)}{\partial \tau^2} + |q(z, \tau)|^2 q(z, \tau) - \sigma q(z, \tau) \frac{\partial}{\partial \tau} |q(z, \tau)|^2 \quad (23)$$

Y describe al solitón

$$q_{sol}(z, \tau) = K_0 \operatorname{sech}[K_0(\tau + \omega_{sol}z)] \exp[i\theta(z, \tau)] \quad (24)$$

con un cambio monótonico de la frecuencia central (“color”)  $\omega_{sol}$  de un solo pulso solitónico único hasta la parte baja del espectro de frecuencia (“rojo”).

Además

$$\theta(z, \tau) = \omega_{sol}\tau - \frac{1}{2} (K_0^2 - \omega_{sol}^2)z \quad (25)$$

$$\omega_{sol}(z) = \omega_{sol}(z=0) - \frac{8}{15} K_0^4 \sigma z \quad (26)$$

y  $\sigma = 2\gamma\mu\beta = \tau_R / T_0$ , donde se introduce la característica  $\tau_R$  Raman tiempo de respuesta de Raman. Para vidrio de sílice  $\tau_R \approx 6$  fs. La intensidad espectral correspondiente

$$|q_{sol}(\omega, \tau)|^2 = \frac{1}{4} \operatorname{sech}^2 \left[ \frac{\pi}{2} \frac{\omega - \omega_{sol}(z)}{K_0} \right] \quad (27)$$

representa la principal característica de los solitones de colores que no viajan sólo en el espacio de coordenadas y tiempo (véanse las ecuaciones 24 y 25), sino que también lo hacen en el espacio espectral (véanse las ecuaciones 26 y 27).

Consideremos la Ecuación (20) como el potencial de “dispersión” del solitón  $U_{ext}(z, \tau) = \beta Q(z, \tau)$ . Cabe destacar que este potencial de dispersión no es sólo una función de coordenadas, sino que también depende del tiempo, y lo que es más, este potencial de dispersión depende de la amplitud y la duración de la dispersión de solitón en sí. Este hecho sigue directamente a partir de la solución dada por:

$$U_{ext}(Z, \tau) = \beta \int_0^\infty H(t) |q(z, \tau - t)|^2 dt, \quad (28)$$

donde el causal Green función  $H(t)$  está representado por

$$H(t) = -\frac{\mu}{\sqrt{1-\gamma^2}} \exp\left(\frac{\gamma}{\mu} t\right) \sin\left(\frac{\sqrt{1-\gamma^2}}{\mu}\right). \quad (29)$$

Tenga en cuenta que, recientemente, algunas analogías inesperadas de solitones ópticos de color se han encontrado en otras áreas de la física. En particular, se demostró que el efecto auto-dispersivo de Raman se puede describir mediante el modelo de ENLS con un potencial externo lineal. Realmente, en el supuesto de que una amplitud solitónica no varía significativamente durante su dispersión  $|q_{sol}(z, \tau)|^2 = \kappa_0^2 \operatorname{sech}^2[\kappa_0(\tau + \omega_{sol}z)]$  podemos estimar el último término de la Ecuación (23) como

$$\sigma \frac{\partial |q_{sol}(z, \tau)|^2}{\partial \tau} \approx -2\sigma \kappa_0^4 \tau \quad (30)$$

y relacionar la Ecuación (23) con el modelo ENLS siguiente:

$$i \frac{\partial q(z, \tau)}{\partial z} = \frac{1}{2} \frac{\partial^2 q(z, \tau)}{\partial \tau^2} + |q(z, \tau)|^2 q(z, \tau) + 2\sigma \kappa_0^4 \tau q(z, \tau) \quad (31)$$

El modelo ENLS Ecuación (31) con un potencial externo lineal se conoce como el modelo de Chen y Liu en un plasma linealmente homogéneo. Por eso, los solitones ópticos coloreados y solitones Alfvén en un plasma linealmente no homogéneo están estrechamente relacionados entre sí. La ecuación (31) representa uno de los llamados modelos integrables por medio del método de dispersión inversa. Este hecho explica la notable estabilidad de los solitones Raman de color en el límite  $\sigma \ll 1$ , que está garantizado por la propiedad de integrabilidad de la ENLS con potencial lineal externo (Chen y Liu, 1978).

Las propiedades de los solitones de colores han llamado nuestra atención principalmente por el papel de la energía de enlace solitónica  $\langle E_{bind} \rangle = -\kappa_0^2/3$  en su dinámica. De acuerdo con la Ecuación (27), el desplazamiento auto-frecuencia del solitón aumenta en la medida que lo hace la segunda potencia de la energía de enlace solitón

$$\omega_{sol}(z) = \omega_{sol}(0) - \frac{72}{15} \sigma \langle E_{bind} \rangle^2 z \quad (32)$$

Este hecho nos da la respuesta a la pregunta: ¿qué hay que hacer para conseguir un mejor efecto SRSS posible? Como el aumento del form-factor del solitón  $K_0$  durante su amplificación resulta en un crecimiento de la energía de enlace del solitón, entonces también el reforzamiento de la no linealidad a lo largo de distancia de propagación produce el mismo efecto.

Vamos a mostrar que la amplificación de un solitón considerando el marco de referencia del modelo siguiente

$$i \frac{\partial q(z, \tau)}{\partial z} = \frac{1}{2} \frac{\partial^2 q(z, \tau)}{\partial \tau^2} + (1 - \beta) |q(z, \tau)|^2 q(z, \tau) + \beta Q(z, \tau) q(z, \tau) + i \frac{\alpha(z)}{2} q(z, \tau), \quad (33)$$

después de la simple transformación

$$q(\xi, \eta) = \tilde{q}(\xi, \eta) \exp \frac{1}{2} \int_0^\eta \alpha(\tau) d\tau \quad (32)$$

es matemáticamente equivalente ambos a una no linealidad creciente

$$i \frac{\partial \tilde{q}(z, \tau)}{\partial z} = \frac{1}{2} \frac{\partial^2 \tilde{q}(z, \tau)}{\partial \tau^2} + (1 - \beta) \exp \left[ \int_0^z \alpha(z) dz \right] |\tilde{q}(z, \tau)|^2 \tilde{q}(z, \tau) + \beta Q(z, \tau) \tilde{q}(z, \tau). \quad (33)$$

y a una fuerza externa en aumento exponencial en el modelo de oscilador

$$\mu^2 \frac{\partial^2 Q(z, \tau)}{\partial \tau^2} + 2\gamma\mu \frac{1}{2} \frac{\partial Q(z, \tau)}{\partial \tau} + Q(z, \tau) = \exp \left[ \int_0^z \alpha(z) dz \right] |q(z, \tau)|^2 \quad (34)$$

del efecto auto-dispersivo del solitón de Raman.

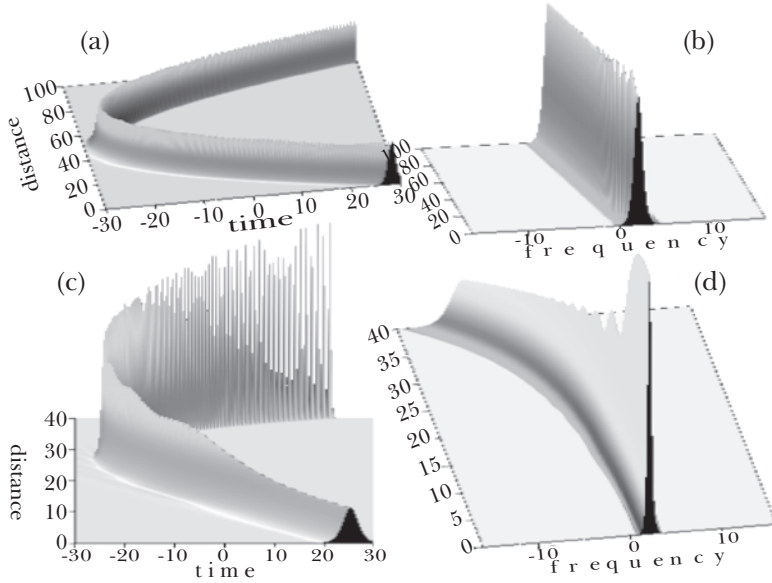
Variando no linealidad se proporciona la auto-compresión solitónica durante SRSS y debido a esto, conduce al reforzamiento del desplazamiento de auto-frecuencia del solitón.

$$\omega_{sol}(z) = \omega_{sol}(0) - \frac{2\kappa_0^4 \sigma}{15\alpha_0} \left[ \exp(4\alpha_0 z) - 1 \right] \quad (35)$$

$$\kappa(z) = \kappa_0 \exp(\alpha_0 z) \quad (36)$$

donde  $\alpha(z) = \alpha_0$  es el coeficiente de ganancia.

En la Figura IV.2 se ilustran las modificaciones más importantes del efecto de SRSS calculado en el marco del modelo dado por las ecuaciones (33, 34). Dos características de las soluciones obtenidas mediante la integración directa de las ecuaciones (33, 34) son notables: (1) En primer lugar, el incremento de la energía de enlace conlleva a la auto-compresión del solitón. Este hecho cambia significativamente las propiedades espectrales de los solitones de color, (2) La dependencia del desplazamiento de la auto-frecuencia del solitón con respecto a la distancia de propagación no es una función lineal, pero está definida como una función exponencialmente creciente.



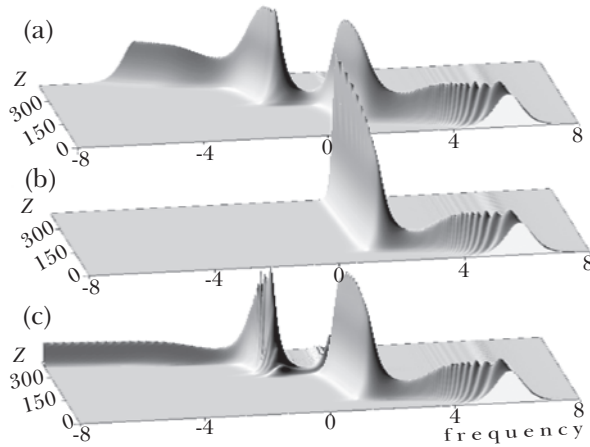
**Figura IV.2.** Efecto de auto-compresión y de Raman auto-dispersión calculado en el marco de las ecuaciones del modelo (30, 31) con aumento exponencial de la energía de enlace del solitón: (a) el comportamiento temporal de la auto-compresión de solitón coloreado y (b) auto-frecuencia desplazamiento después de elegir los siguientes parámetros:  $\mu = 0,5$ ;  $\gamma = 0,3$ ;  $\beta = 0,18$ ;  $\nu_0 = 2,25$ ;  $\kappa_0 = 0,5$ ;  $\alpha_0 = 0,05$ .

Nuestros experimentos computacionales muestran que el aumento de energía de enlace resulta en cambios significativos de las propiedades del efecto SST. Consideremos la situación cuando la frecuencia central solitónica se coloca lejos de la barrera de la DVG localizada. La dependencia de la frecuencia de la barrera DVG puede ser escrita convenientemente como

$$\frac{\partial^2 \kappa}{\partial \omega^2} = -\frac{\partial^2 \kappa}{\partial \omega^2}(\omega = \omega_{sol})[1 - f_2(\omega)] \quad (36)$$

Donde  $\frac{\partial^2 \kappa}{\partial \omega^2}(\omega = \omega_{sol})$  representa al parámetro de DVG en la frecuencia solitónica central y la  $f_2(\omega)$  describe la inhomogeneidad espectral

$$f_2(\omega) = h_{bar} \cdot \exp \left[ -\left( \frac{\omega - \Omega_{bar}}{\Delta \Omega_{bar}} \right)^n \right] \quad (37)$$



**Figura IV.3.** La influencia de la energía de enlace solitónico en el efecto túnel espectral del solitón. (a) Efecto túnel solitónico espectral a través de la barrera de DVG gaussiana centrada en la frecuencia  $h_{bar} = 0$  con la altura  $h_{bar} = 2.5$  y la  $\Delta\Omega_{bar}=3$ . El solitón se inicializa lejos de la barrera localizada y se coloca inicialmente en la frecuencia  $\omega_0 = 6$  y su comportamiento se calculó para el caso  $\alpha_0 = 0$ . (b) La supresión total del efecto SST en virtud de la altura de la barrera de DVG  $h_{bar}=5.0$  calculado para el caso  $\alpha_0=0$ . (c) Efecto túnel espectral significativamente reforzado de la auto-compresión solitónica de color calculado para el caso  $h_{bar} = 5.0$  y  $\alpha_0=0.0015$ .

donde el parámetro  $h_{bar}$  es la altura de una barrera espectral con la  $\Delta\Omega_{bar}$ , anchura espectral, que está centrado en la  $\Omega_{bar}$  frecuencia. La variable  $n = 2, 4, \dots$  hace que sea posible alterar la pendiente de la homogeneidad espectral de la DVG.

Hemos calculado diferentes escenarios de tunelaje de auto-compresión de los solitones de colores de diferentes amplitudes y velocidades, con una variedad de la pendiente  $n$  de la barrera. En lo que sigue, se define que las principales características del efecto de la SST siguen siendo los mismos. Las simulaciones numéricas muestran una diferencia dramática en todos los escenarios posibles del espectro de solitones túnel cuando la energía de enlace solitón se ha incrementado. Los resultados representativos más importantes de tales simulaciones por ordenador aparecen en la Figura IV.3.

Esta Figura puede considerarse como un resumen de todos los escenarios posibles del efecto SST no lineal. La Figura IV.3a muestra que durante el tunelaje del solitón a través de la barrera prohibida (positivo) en DVG el solitón de color sigue siendo un solitón “puro”. La Figura IV.3b muestra que el aumento de la altura de

una barrera espectral conduce a la supresión completa del efecto SST, y la Figura IV.3c ilustra el aumento del efecto túnel espectral del solitón. A partir de simulaciones de ordenadores, se infiere que las principales características del efecto túnel espectral dependen fuertemente de la energía de enlace solitónico. Se debe enfatizar que la energía de enlace solitónico tiene un impacto dramático en el efecto SST como se muestra en la Figura IV.3c. Por eso, el efecto túnel reforzado espectral de auto-compresión de solitones podrían ser de interés fundamental en el campo de la generación de solitones supercontinuos.

## CONCLUSIONES

En la primera parte del trabajo, se analizaron nuevas posibilidades de amplificación de solitones ópticos. Hemos demostrado que la condición más importante para determinar las principales posibilidades de amplificación y compresión de un solitón modulado en su frecuencia en un medio amplificado, es el mantenimiento de una modulación de frecuencia lineal durante intervalos de tiempo superior a la duración inicial del solitón. La función óptima de la ganancia solitónica se puede obtener en un amplificador de fibra bi-fotónica.

La amplificación ideal de solitones ópticos es posible porque, en el caso de la ganancia no homogénea, las condiciones se producen por el cambio del “gorjeo” solitónico de tal manera que un solitón amplificado completamente conserva sus propiedades únicas de interacción elástica con solitones similares. En cuanto a los problemas periódicos y aquellos denominados Ondas de Bloch (Chen *et al.*, 1999; Serkin y Belyaeva, 2001a; Serkin *et al.*, 2001) podemos afirmar que son completamente análogos a los problemas de generación de pulsos ultracortos durante el auto-enfoque intracavitario de la radiación (Lariontsev *et al.*, 1975; French, 1995). Nosotros intentaremos a considerar este viejo problema usando la analogía espacio temporal que nos permite generalizar fácilmente los resultados obtenidos para el caso de auto-enfocamiento intracavitario óptimo de la radiación. Como se señaló en la revisión (French, 1995), los estudios iniciados en un trabajo pionero (Lariontsev *et al.*, 1975) han anticipado la llegada de los láseres del tipo de auto-Kerr en modo bloqueado Ti: Al<sub>2</sub>O<sub>3</sub> láser.

Por tanto, el problema de óptima amplificación (en nuestra terminología, ideal) de solitones demuestra una estrecha relación con diversos problemas de la generación de pulsos láser ultracortos y del desarrollo de nuevos sistemas de comunicación.

Estos problemas están aún lejos de su completa solución. El modelo no lineal de la ecuación de Schrödinger no puede responder a la pregunta acerca de la duración límite de solitones en un medio activo. Este problema puede ser resuelto

por los métodos basados en la solución numérica directa del sistema de ecuaciones de Maxwell no lineales en un medio amplificador (Serkin *et al.*, 1996; Serkin *et al.*, 2000; Serkin *et al.*, 1998).

Aunque el problema de generación de pulsos láser ultracortos ya tiene una antigüedad de más de 40 años en su historia, últimamente el interés por estos relativamente viejos problemas se ha reavivado. Esto se demuestra por el desarrollo de líneas de comunicación de alta velocidad y la generación de un supercontinuo femtosegundo en fibras ópticas; este proceso, asimismo, se estimuló con la fabricación de nuevas fibras ópticas con características de dispersión especial óptica y fibras cónicas de cintura y cristales fotónicos.

En la segunda parte del trabajo, hemos mostrado que el aumento de la energía de enlace del solitón debido a la auto-compresión solitónica permite la prueba de un resultado teórico de grandes datos, que predice que un solitón óptico puede crear un túnel entre dos regiones de dispersión anómala a través de una región prohibida de dispersión normal (efecto túnel reforzado espectral). Nos gustaría concluir diciendo que el concepto de construcción de túneles es de primordial importancia en la naturaleza y efecto túnel solitónico porque puede ser uno de atención fundamental en la ciencia no lineal.

Resumiendo: podemos decir que el problema de la amplificación óptima de solitones es importante desde el punto de vista general de la física debido a que el solitón es una de las entidades fundamentales de la naturaleza. Esta entidad no lineal está siendo estudiada en casi todos los campos de la ciencia moderna, y es difícil no estar de acuerdo con una hipótesis planteada en el artículo (Krumhansl, 1991) de que el paradigma solitónico servirá como base de unificación para el desarrollo ulterior de la ciencia.

## REFERENCIAS

- Agrawal, G. P. (1995). *Nonlinear Fiber Optics*. 2nd Ed. Nueva York, Academic Press.
- Afanasiev, V. V. *et al.* (1990). *Pisma Zh. Eksp. Teor. Fiz.* 16, 67.
- Balakrishnan, R. (1985). *Phys Rev. A*. 32, 1344.
- Bespalov, V. G.; Krylov, V. N.; Seyfang, G.; Staselko, D. L.; Kozlov, S. A.; Shpolyansky Yu. A. y Rebane, A. (2001). *A. Proc. SPIE Int. Soc. Opt. Eng.* 4271, 159.
- Birks, T. A.; Wadsworth, W. I. y Russel, P. St. (2000). *J. Opt. Lett.* 25, 1415.
- Burtsev, S. P.; Zakharov, V. E.; Mikhailov, A. V. (1987). *Theor. Math. Phys.* 70, 227.
- Blow, K. J.; Doran, N. J. y Wood, D. (1987). *Opt. Lett.* 12, 1011.
- Bogatyryov, V. A.; Bubnov, M. M.; Dianov, E. M. y Sysolialin, A. A. (1995). *Pure Appl. Opt.* 4, 345.

- Chen H. H. y Liu C. S. (1978). *Physics of Fluid*. 21, 377.
- Chen, Y.; Kartner, F. X.; Morgner, U; Cho, S. H.; Haus, H. A.; Ippen, E. P. y Fujimoto, J. G. (1999). *Opt. Soc. Am. B*. 16, 199.
- Dianov, E. M. (2000). *Kvantovaya Elektronika*. 30, 659.
- Dianov, E. M. y Krukov, P. G. (2001). *Kvantovaya Elektronika*. 3, 877.
- Dianov, E. M.; Prokhorov, A. M. y Serkin, V. N. (1983). *Dokl. Akad. Nauk SSSR*. 273, 1112.
- Dianov, E. M.; Prokhorov, A. M. y Serkin, V. N. (1986). *Opt. Lett.* 11, 168.
- Dianov, E. M.; Karasik, A. Ya.; Prokhorov, A. M.; Serkin, V.; Dianov, E. M.; Nikonova, Z. S. y Serkin, V. N. (1986). *Kvantovaya Elektronika*. 13, 1740.
- Dianov, E. M.; Nikonova, Z. S. y Prokhorov, V. N. (1986). *Izvestia Akademiy Nauk SSSR. Ser. Fizika*. 50, 1042.
- Dianov, E. M.; Nikonova, Z. S.; Prokhorov, V. N.; Serkin V.N.; Serkin, V. N. (1985). *Doklady Akademiy Nauk SSSR*. 283, 1342.
- Dianov, E. M.; Grudinin, A. B.; Prokhorov, A. M. y Serkin, V. N. (1992). In *Optical Solitons-Theory and Experiment*. Ed. by J.R. Taylor (Cambridge: Cambridge University Press) Ch.7, p. 197-265.
- Dudley, J. M.; Genty, G. y Coen, S. (2006). *Rev. Mod. Phys.* 78, 1135-1184.
- Dudley, J. M. y Taylor, J. R. (2009). *Nat. Photon.* 3, 85-90.
- Dudley, J. M. y Taylor, J. R. (eds.) (2010). *Optical Fiber Supercontinuum Generation*. Cambridge: Cambridge University Press.
- French, P. M. W. (1995). *Reports on Progress in Physics*. 58, 169.
- Gatz S. y Herrmann, J. (1992). *IEEE J. Quantum Electron.* 28, 1732.
- Hasegawa, A. (Ed.) (2000). *Massive WDM and TDM Soliton Transmission Systems* (Kluwer: Acad. Pub.).
- He, X. G.; Zhao, D.; Li, L. y Luo, H. G., (2009). *Phys. Rev. E*. 79, 056610 .
- Katys, G. P.; Kravtsov, N. V.; Chirkov, L. E. y Konovalov, M. M. (1967). *Modulatsiya i otklonenie opticheskogo izlucheniya* (Modulation and Deflection of Optical Radiation). Moscú: Nauka.
- Karpman, V. L. y Maslov, E. M. (1978). *Sov. Phys. JETP*. 75, 504.
- Knight, J. C.; Birks, T. A.; Russel, P. St. J. y Atkin, D. M. (1996). *Opt. Lett.* 21, 1547.
- Kibler, B.; Lacourt, P. A; Courvoisier, F. y Dudley, J. M. (2007). "Soliton Spectral Tunnelling in Photonic Crystal Fibre with sub-wavelength Core Defect". *Electron. Lett.* 43, 967-968.
- Krukov, P. G. y Letokhov V. S. (1969). *Uspehi. Fizicheskikh Nauk*. 99, 169.
- Krukov, P. G. (2001). *Kvantovaya Elektronika*. 31, 95.
- Khasilev, V. Y. (1996). *Proc. SPIE. Int. Soc. Opt. Eng.* 2919, 177.
- Khasilev, V. Y.; Malomed, B. A. y Serkin, V. N. (1999). *Proc. SPIE. Int. Soc. Opt. Eng.* 3847, 224.

- Krumhansl, J. A. (1991). *Phys. Today*, 3, 33.
- Kumar, S. y Hasegawa, A. (1998). *Opt. Lett.* TL 372.
- Kumar, S. y Hasegawa, A. (1998). *Physica D*. 123, 267.
- Lariontsev, E. G. y Serkin, V. N. (1975). *Kvantovaya Elektron.* 2, 1481.
- Lenz, G. y Eggleton B. J. (1998). *Opt. Soc. Am. B*. 15, 2979.
- McKinnon, K. L. M.; Smyth, N. F.; Worthy, A. L.; Lenz, G. y Eggleton, B. J. (1999). *Opt. Soc. Am. B*. 16, 441.
- Liu, X.; Xu, C.; Knox, W. H.; Chandalia, J. K.; Eggleton, B. J.; Kosinski, S. G. y Windeler, R. S. (2001). *Opt. Lett.* 26, 358.
- Luo, H. G.; Zhao, D. y He, X. G. (2009). *Phys. Rev. A* 79, 06382.
- Mitchke, F. M. y Mollenauer, L. F. (1986). *Opt. Lett.* 11, 659.
- Mollenauer, L. F.; Stolen, R. H. e Islam, M. N. (1985). *Opt. Lett.* 10, 229.
- Mollenauer, L. F.; Stolen, R. H.; Gordon, J. P. y Tomlinson, W. L. (1983). *Opt. Lett.* 8, 289.
- Moores, J. D. (1996). *Opt. Lett.* 21, 555.
- Mustel, E. P. y Parygin, V. N. (1970). *Metody modulyatsii i skanirovaniya sveta (Methods for Light Modulation and Scanning)*. Moscú: Nauka.
- Nakazawa, Masataka (1992). *Optical Solitons-Theory and Experiment*. Ed. by J. R. Taylor (Cambridge: Cambridge University Press) Ch. 6, 52-196.
- Nikonova, Z. S. y Serkin, V. N. (1990). *Trudy LOFAN*. 23, 39.
- Poletti, F.; Horak, P. y Richardson, D. (2008). *J. IEEE Photon. Tech. Lett.* 20, 1414-1416.
- Ponomarenko, S. A. y Agrawal G. P. (2006). *Phys. Rev. Lett.* 97, 013901.
- Ponomarenko, S. A. y Agrawal, G. P. (2007). *Opt. Lett.* 32, 1659.
- Ponomarenko, S. A. y Agrawal, G. P. (2007). "Interactions of Chirped and Chirp-free Similaritons in Optical Fiber Amplifiers". *Opt. Express*. 15, 2963.
- Porsezian, K.; Hasegawa, A.; Serkin, V. N.; Belyaeva, T. L. y Ganapathy, R. (2007). *Phys. Lett. A* 361, 504.
- Richardson, D. J. C.; Jiamberlin, R. P.; Dong, L. y Pane, D. N. (1995). *Electron Lett.* 311, 1681.
- Semenov, V. A.; Belov, A. V.; Dianov, E. M.; Abramov, A. A.; Bubnov, M. M.; Semjonov, S. L.; Shchegunjaev, A. S.; Khopin, V. F.; Guryanov, A. N. y Vechkanov, N. N. (1995). *Appl. Opt.* 34, 5331.
- Serkin, V. N. y Belyaeva, T. L. (2001a). *Proc. SPIE Int. Soc. Opt. Eng.* 4271, 323.
- Serkin, V. N. y Belyaeva, T. L. (2001b). *JETP Letters*. 74, 573-577.
- Serkin, V. N.; Belyaeva, T. L.; Alexandrov, L. V. y Melo Melchor, G. (2001). *Proc. SPIE Int. Soc. Opt. Eng.* 4271, 292.
- Serkin, V. N. y Belyaeva, T. L. (2001a). *Proc. SPIE Int. Soc. Opt. Eng.* 4271, 280.
- Serkin, V. N.; Belyaeva, T. L. (2001b). *Kvanto/aya Elektron.* 31, 1016.

- Serkin, V. N. y Hasegawa, A. (2000a). *Proc. SPIE. Int. Soc. Opt. Eng.* 3927, 302.
- Serkin, V. N. y Hasegawa, A. (2000b). *Pis'ma Zh. Eksp. Teor. Fiz.* 72, 89.
- Serkin, V. N. y Hasegawa, A. (2002). *Sel. Top. Quantum Electron.* 8, 418.
- Serkin, V. N. (1986). *Pis'ma Zh. Eksp. Teor. Fiz.* 12, 756.
- Serkin, V. N. y Hasegawa, A. (2000a). *Phys Rev. Lett.* 85, 4502.
- Serkin, V. N.; Hasegawa, A. y Belyaeva, T. L. (2004). *Phys. Rev. Lett.* 92, 199401.
- Serkin, V. N.; Hasegawa, A. y Belyaeva, T. L. (2007). *Phys. Rev. Lett.* 98, 074102.
- Serkin, V. N.; Hasegawa, A. y Belyaeva, T. L. (2010a). *J. of Mod. Opt.* 57, 1-17.
- Serkin, V. N.; Hasegawa, A. y Belyaeva, T. L. (2010b). *Phys. Rev. A.* 81, 023610.
- Serkin, V. N.; Matsumoto, M. y Belyaeva, T. L. (2001). *Opt. Commun.* 196, 159.
- Serkin, V. N.; Schmidt, E. M.; Samarina, E. V. y Belyaeva, T. L. (1996). *Proc. SPIE Int. Soc. Opt. Eng.* 2800, 310.
- Serkin, V. N.; Schmidt, E. M. y Belyaeva T. L. (2000). *Proc. SPIE Int. Soc. Opt. Eng.* 3927, 323.
- Serkin, V. N.; Schmidt, E. M.; Belyaeva, T. L. y Khotyaintsev, S. N. (1998). *Dokl. Ross. Akad. Nauk.* 359, 760.
- Segev M. y Stegeman, G. (1998). *Physics Today.* 8, 42.
- Sysoliatin, A.; Belanov, A.; Konyukhov, A.; Melnikov, L. y Stasyuk, V. (2008). *IEEE J. Sel. Top. Quantum Electron.* 14, 733-738.
- Sysoliatin, A.; Senatorov, A. K.; Konyukhov, A. I.; Melnikov, L. A. y Stasyuk, V. A. (2007). *Opt. Express.* 15, 16302-16307.
- Tenorio, C. H.; Vargas, E. V.; Serkin, V. N. *et al.* (2005). *Quantum Electronics.* 35, 929-937.
- Tenorio, C.; Belyaeva, T. L. y Serkin, V. N. (2007). *Physica B- Condensed Matter,* 398, 460-463.
- Travers, J. C.; Stone, J. M.; Rulkov, A. B.; Cumberland, B. A.; George, A. K.; Popov, S. V.; Knight, J. C. y Taylor, J. R. (2007). *Opt. Express.* 15, 13203.
- Tsoi, E. N. y De Sterke, C. M. (2007). *Phys. Rev. A.* 76, 043804.
- Wu, Lei; Zhang, Jie-Fang; Finot, C. y Li, Lu (2009). *Optics Express,* 17.
- Yong-Xin, Yan; Gamble, E. B. y Nelson, K. A. (1985). *Chem. Phys.* 83, 5391.
- Zakharov, V. E. y Mikhailov, A. V. (1987). *Theor. Math. Phys.* 70, 227.
- Zakharov, V. E. y Shabat, A. B. (1971). *Zh. Eksp. Teor. Fiz.* 61, 118.
- Zakharov, V. E.; Pushkarev, A. N. y Shvets, V. F. (1988). *Pis'ma Zh. Eksp. Teor. Fiz.* 48, 79.
- Zhao, D.; Luo, H-G. y Chai, H-Y. (2008). *Phys. Lett. A.* 372, 5644.

# EL TEOREMA DE PASO DE MONTAÑA\*

*Alfredo Cano\*\**

*Fernando A. Ongay Larios\*\*\**

*Ma. de Lourdes Nájera-López\*\*\*\**

## Resumen

En la ciencia y en las aplicaciones tecnológicas son comunes los problemas de optimización. En este artículo se analizan problemas de optimización y se mencionan teoremas relacionados con la obtención de máximos o mínimos. En particular, se da la demostración del Teorema de Paso de Montaña en dimensión finita. La demostración hace referencia a ciertas sucesiones conocidas como sucesiones de Palais-Smale. Finalmente, se da una aplicación de la versión infinita del Teorema de Paso de Montaña a una ecuación diferencial parcial de tipo elíptico.

**Palabras clave:** Optimización, sucesiones de Palais-Smale, puntos críticos.

## INTRODUCCIÓN

En distintas ramas de las ciencias naturales y sociales, así como en las aplicaciones tecnológicas, aparecen problemas de optimización, como la obtención

\* Los autores agradecen a la Secretaría de Educación Pública (SEP) por el apoyo mostrado en el financiamiento al proyecto FEO1/2012 103.5/12/2126 de redes multidisciplinarias y a la SIEA de la Universidad Autónoma del Estado de México por el apoyo importante en las actividades de investigación.

\*\* Doctor en Matemáticas. Profesor de Tiempo Completo, Facultad de Ciencias, Universidad Autónoma del Estado de México, Instituto Literario 100, Toluca 50000, México. (52) (722) 2965556, <afalcaro@yahoo.com.mx>.

\*\*\* Doctor en Matemáticas. Profesor de Tiempo Completo, Facultad de Ciencias, Universidad Autónoma del Estado de México, Instituto Literario 100, Toluca 50000, México, <fernando\_ongay@yahoo.com.mx>.

\*\*\*\* Estudiante del Doctorado en Ciencias Económico-Administrativas, Universidad Autónoma del Estado de México. Investigación como estudiante adjunta a la Facultad de Ciencias. Plantel Nezahualcóyotl de la Escuela Preparatoria, Universidad Autónoma del Estado de México, Instituto Literario 100, Toluca 50000, México, <mlnajeral@uemex.mx>.

del máximo rendimiento de las tasas de interés en cuestiones de economía, o utilizar el mínimo de material para la elaboración de objetos destinados a usarse en nuestra vida diaria, tales como latas o cajas de cartón.

Estos problemas generalmente se orientan a la obtención de valores máximos y mínimos de funciones. La solución de este tipo de problemas logra tener un gran avance cuando el cálculo infinitesimal aporta el concepto de derivadas de una función. En efecto, para una función que es derivable se sabe que los valores máximos o mínimos se encuentran en aquellos puntos en donde su derivada se anula (puntos críticos). La segunda derivada proporciona un criterio para saber si se está ante un máximo o un mínimo.

El cálculo variacional es una de las ramas del análisis ocupada en estudiar las cuestiones sobre optimización. El interés específico del cálculo variacional es encontrar los puntos en donde una función particular, asociada a un problema de optimización, alcance sus valores máximos o mínimos (Gelfand y Fomin, 1963). Existe una amplia variedad de métodos para encontrar puntos críticos de una función, los cuales son los candidatos para optimizar a ésta. Con ayuda de teorías más sofisticadas del Análisis Matemático en dimensiones infinitas se desarrollan los Métodos Variacionales (Struwe, 1980; Bartsch, 1993). Una técnica de estos métodos consiste en utilizar sucesiones minimizantes, las cuales fueron usadas exitosamente en los años sesenta por R. S. Palais y S. Smale (Palais, 1963; Palais y Smale, 1964).

En este trabajo primero haremos una breve presentación de cómo aparecen y cómo se utilizan las sucesiones minimizantes en problemas de optimización, para después enfocarnos en ciertos problemas de optimización conocidos como “Teoremas de Paso de Montaña”. Estos teoremas adquieren el nombre de la siguiente analogía: supóngase que nos encontramos en el interior del cráter de un volcán y deseamos encontrar el punto “óptimo” en el relieve que nos rodea para, escalando la montaña, poder salir del cráter. Estos teoremas pueden verse, en cierto sentido, como la generalización del Teorema de Rolle. Damos también una demostración completa del caso de dimensión finita, el cual no requiere más que elementos básicos de análisis real; el caso de dimensión infinita requiere herramientas más sofisticadas de análisis funcional y análisis no lineal (referimos al lector a los textos Jabri, 2003; Rabinowitz, 1986 y Struwe, 1980).

El artículo está organizado como sigue: en la segunda sección se plantean algunos problemas clásicos de optimización, iniciando con un problema clásico y sencillo el cual se resuelve con teoría básica del Cálculo; posteriormente se analiza otro problema más sofisticado, y si bien no se trata de buscar directamente un máximo o un mínimo, se sabe que el punto en donde se tiene un valor óptimo es solución de una ecuación diferencial. En la tercera sección, con respecto al problema de optimización, se analiza la relación entre la topología del dominio

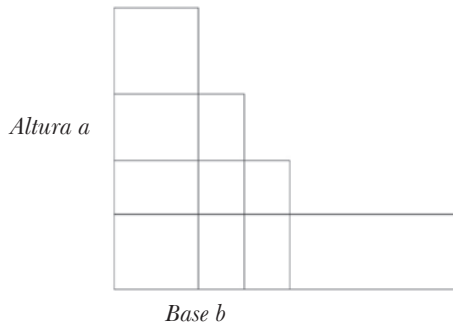
de las funciones y cuestiones como continuidad y diferenciabilidad. En la cuarta introducimos el concepto de sucesiones minimizantes y analizamos algunos detalles que se deben cuidar. En la quinta sección, que es la principal, estudiamos el Teorema de Paso de Montaña en dimensión finita. Posteriormente hablamos de las sucesiones de Palais-Smale y de la generalización del Teorema de Paso de Montaña a dimensión infinita. Y concluimos con un ejemplo de una ecuación diferencial particular en donde se visualiza el uso del Teorema de Paso de Montaña.

## I. PROBLEMAS CLÁSICOS DE OPTIMIZACIÓN

Analicemos algunos problemas típicos de optimización.

**Ejemplo 1.** Se tiene una cuerda de longitud  $P$ , y se quiere encerrar con ésta una superficie rectangular de área máxima, la pregunta entonces es ¿cómo deben ser las dimensiones del rectángulo? En este caso la solución requiere de dos observaciones: la primera es que el área  $A$  se calcula en términos de su altura  $a$  y su base  $b$  con la expresión  $A = ab$ , y la otra es que la longitud del perímetro del rectángulo es la longitud de la cuerda, esto es  $P = 2a + 2b$ .

**Figura I.1 Rectángulos de perímetro constante**



De éstas se puede plantear la siguiente función en términos de la variable  $b$ .

$$A(b) = \left(\frac{P-2b}{2}\right)b = \frac{P}{2}b - b^2.$$

Si derivamos obtenemos  $A'(b) = \frac{P}{2} - 2b = 0$ , e igualando a cero se tiene que  $b = \frac{P}{4}$ . Se puede deducir que para  $b = \frac{P}{4}$  se tiene que los lados deben ser todos iguales y para este valor se tiene un área máxima.

**Ejemplo 2.** Consideramos el problema de Dirichlet

$$\begin{cases} u''(x) = 0 \text{ para } 0 < x < 1 \\ u(0) = 0, u(1) = 0. \end{cases} \quad (*)$$

Es claro que  $u = 0$  es una solución. Pero lo que queremos describir aquí es un método para resolver este tipo de ecuaciones. Se multiplica la ecuación por una función  $v$  dos veces diferenciable con derivada continua en el intervalo  $[0, 1]$  y tal que  $v(0) = 0, v(1) = 0$  y a continuación se integra sobre el intervalo obteniendo

$$\int_0^1 u''(x)v(x)dx = 0.$$

Integrando por partes se tiene

$$\begin{aligned} \int_0^1 u''(x)v(x)dx &= u(x)v(x) \Big|_0^1 - \int_0^1 v'(x)u'(x)dx \\ &= -\int_0^1 v'(x)u'(x)dx. \end{aligned}$$

Así el problema de Dirichlet es equivalente a resolver la igualdad

$$\int_0^1 v'(x)u'(x)dx = 0 \quad (1)$$

para toda función  $v$  derivable con derivada continua. Si consideramos el funcional

$$E(u) = \int_0^1 (u')^2 dx,$$

observamos que su derivada es  $DE(u)v = \int_0^1 u'(x)v'(x)dx$ , y comparando con (1) notamos que resolver el problema de Dirichlet es equivalente a encontrar un punto (una función  $u$ ) en donde la derivada del funcional  $E$  se anule, con el problema adicional de que este punto debe ser dos veces diferenciable, el cual se resuelve utilizando resultados de regularidad (véase Brezis, 1984). Por otra parte, como  $E(u)$  es no negativa y si  $u$  es solución de (\*) observamos lo siguiente

$$\begin{aligned} E(v) &= \int_0^1 (v')^2 dx \\ &= \int_0^1 (u')^2 dx + \int_0^1 ((v-u)')^2 dx + 2 \int_0^1 u'(v-u)' dx \\ &= E(u) + E(v-u) + 2 (v-u)u' \Big|_0^1 - \int_0^1 (v-u)u'' dx \end{aligned}$$

$$\begin{aligned}
 &= E(u) + E(v-u) \\
 &\geq E(u),
 \end{aligned}$$

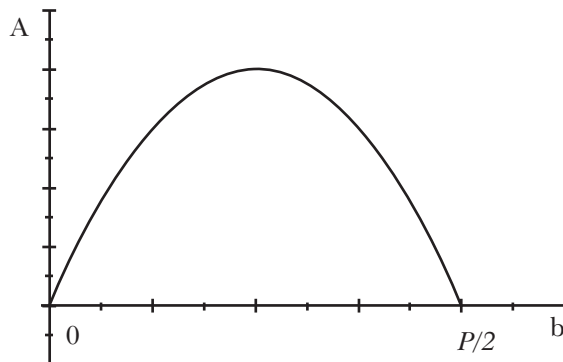
donde la evaluación en la integración por partes es igual a cero porque si buscamos funciones con las condiciones iniciales de (\*) entonces las funciones  $v$  y  $u$  deben ser iguales en la frontera del intervalo  $[0, 1]$ . De la desigualdad concluimos que la solución además debe ser en donde el funcional  $E$  tenga un mínimo.

De estos ejemplos se puede observar que si un problema real de optimización lo podemos traducir al comportamiento de una función, con ayuda de diferentes modelos matemáticos podemos saber si tiene algún máximo o mínimo, y entonces interpretar esto para resolver nuestro problema. Esto nos lleva a estudiar algunas propiedades para que nuestras funciones tengan valores extremos.

## II. FUNCIONES CONTINUAS Y COMPACIDAD

La solución de un problema de optimización puede verse beneficiada por propiedades particulares de la función que se desea optimizar y del dominio de ésta. Por ejemplo, la continuidad en un intervalo cerrado asegura que la función está acotada por abajo, es decir, que existe un número  $M \in \mathbb{R}$  tal que  $f(x) \geq M$  para toda  $x$  en  $[a, b]$ , por lo que existe el valor ínfimo de la función. Entonces podemos preguntarnos si existen puntos  $x_0$  en el dominio de  $f$  tal que  $f(x_0) = \inf_{x \in [a, b]} f(x)$ . El resultado es inmediato si recordamos que una función continua manda conjuntos compactos en conjuntos compactos, así  $f[a, b]$  es compacto, de donde se alcanza

**Gráfica II.1**  
**Función de áreas**



un valor mínimo en algún punto  $x$  del dominio de la función  $y$ , por tanto, el problema de minimizar una función continua en un intervalo cerrado tiene solución (Rudin, 1964; Protter, 1998; Kolmogorov y Fomin, 1970). Pero posiblemente el mínimo se alcance en los extremos del intervalo, como es el caso de la función área  $A(b) = \frac{P}{2} b - b^2$  del ejemplo 1 (véase Gráfica II.1).

Podemos observar que los valores ínfimos se alcanzan en  $b = 0$  y  $b = \frac{P}{2}$ . Aunque en este caso estos puntos no son de interés, ya que el área que se obtiene con estos valores es cero, en otros ejemplos podría darse esta situación.

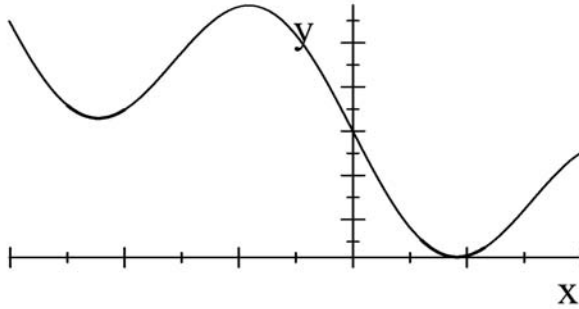
Así la compacidad del dominio y la continuidad de la función nos pueden asegurar la existencia de un mínimo. El problema ahora es cómo encontrar los puntos del dominio en donde se alcanza el valor mínimo. En el caso de intervalos o se alcanza en los extremos o se alcanza en algún punto del interior. En los cursos de cálculo hemos aprendido que los valores extremos en puntos del interior se alcanzan en puntos donde la derivada se anula cuando la función es además diferenciable. Esto se conoce como el criterio de la primera derivada. El problema con los extremos, como podemos ver en la gráfica, es que en donde se tienen valores mínimos la derivada no necesariamente se anula. En un intervalo es concebible la opción de evaluar la función en los extremos y comparar los valores obtenidos con los puntos del interior; sin embargo, en dimensiones mayores, la opción puede ya no ser viable. Por esta razón las funciones que consideramos en adelante son aquellas cuyos valores mínimos no se alcancen en la frontera del dominio.

Si se tiene un dominio en el cual no se cumple la propiedad de compacidad, aunque la función sea acotada, puede no alcanzar el valor ínfimo de la función, ver el ejemplo 4 adelante.

### III. MÍNIMOS Y PUNTOS CRÍTICOS

Supongamos que las funciones están definidas y son continuas en un dominio  $U$  dentro de un espacio métrico. Se dice que en  $z \in U$  se tiene un **mínimo absoluto** si  $f(z) \leq f(x)$  para todo  $x$  en el dominio, y se dice que es un **mínimo local** si  $f(z) \leq f(x)$  para todo  $x$  en un vecindad de  $z$ . Una condición que pediremos en adelante es que los mínimos locales no se alcancen en la frontera de  $U$  como se ilustra en la siguiente figura.

**Figura III.1**  
**Máximo y mínimos locales**



Con el objetivo de encontrar los puntos en donde se tenga un mínimo definimos lo siguiente.

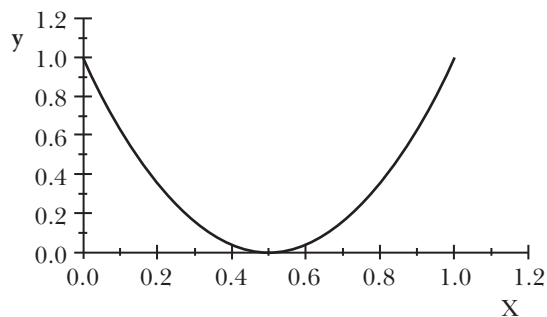
Una sucesión  $\{x_n\} \subset U$  es una **sucesión minimizante** de  $f$  si

$$f(x_n) \rightarrow c := \inf_{x \in U} f(x).$$

Lo que queremos saber es cuándo estas sucesiones minimizantes convergen a un punto en el interior de  $U$ . Analicemos un poco algunas situaciones que suelen darse con respecto a estas sucesiones.

**Ejemplo 3.** Sea  $f: [0,1] \rightarrow \mathbb{R}$  dada por  $f(x) = (2x-1)^2$ . Una sucesión minimizante es  $\left\{x_n \in (0,1) : x_n := \frac{1}{2} + \frac{1}{n} = \frac{n+2}{2n}\right\}$ , ya que  $f(x_n) = \frac{4}{n^2}$  y si hacemos que  $n$  tienda a infinito entonces  $f(x_n)$  tiende a  $\inf_{x \in U} f(x) = 0$ .

**Figura III.2**  
**Gráfica de  $f(x) = (2x-1)^2$**



Si consideramos a las sucesiones en un intervalo cerrado contenido en  $(0,1)$ , tiene una subsucesión convergente que seguimos denotando por  $\{x_n\}$

$$x_n \xrightarrow{n \rightarrow \infty} x_0.$$

Es fácil ver, en nuestro ejemplo, que  $x_0 = \frac{1}{2}$ . De la continuidad de  $f$  se sigue que

$$f(x_n) \rightarrow f(x_0) = \inf_{x \in [0, 1]} f(x).$$

De esta forma podemos decir que en  $x_0$  se encuentra un mínimo.

¿Qué pasa con la derivada de la función evaluada en la sucesión? Calculamos

$$f(x) = (2x-1)^2 \Rightarrow f'(x) = 8x - 4$$

$$f'(x_n) = 8\left(\frac{n+2}{2n}\right) - 4$$

$$= \frac{8n + 16 - 8n}{2n}$$

$$= \frac{8}{n}$$

Entonces  $f'(x_n) \rightarrow 0$ , cuando  $x_n \rightarrow \frac{1}{2}$ . Por la continuidad de la derivada (o directamente) tenemos  $f'(x_0) = f'\left(\frac{1}{2}\right) = 0$ .

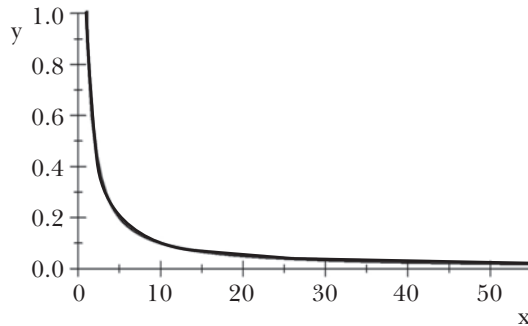
Si la función  $f$  tiene derivada continua en su dominio, recordamos que un punto  $y$  en  $U$  que cumple  $f'(y) = 0$  se dice que es un **punto crítico**. Con respecto al problema de optimización, la pregunta importante es: ¿toda sucesión minimizante  $\{x_n\}$  tal que  $f'(x_n) \rightarrow 0$  es (o contiene una subsucesión) convergente al mínimo? Consideremos el siguiente ejemplo.

**Ejemplo 4.** Considérese  $f: (0, \infty) \rightarrow \mathbb{R}$  tal que  $f(x) = \frac{1}{x}$ . Pensemos en la sucesión con  $x_n := n$ . Es claro que  $f(x_n) = \frac{1}{n} \rightarrow 0$ ,  $\inf_{x \in (0, \infty)} f(x) = 0$  y

$$f'(x_n) = -\frac{1}{n^2} \rightarrow 0 \text{ si } n \rightarrow \infty,$$

y sin embargo la sucesión  $x_n$  no converge.

**Figura III.3**  
Gráfica de  $f(x) = \frac{1}{x}$



En este ejemplo la función, de hecho, no tiene mínimos ni locales ni absolutos. Uno de los problemas aquí es que  $(0, \infty)$  no es acotado y por lo tanto no es compacto. Analicemos un ejemplo en donde el dominio sí es acotado.

**Ejemplo 5.** Ahora sea  $l_1 := \{\bar{x} = (x_1, x_2, \dots, x_n, \dots) : \sum_{i=1}^{\infty} |x_i| < \infty\}$ , con la norma dada por  $\|\bar{x}\| := \sum_{i=1}^{\infty} |x_i|$ , y definimos la función  $f: l_1 \rightarrow \mathbb{R}$  tal que

$$f(\bar{x}) = x_N^2,$$

para un  $N$  natural fijo. La derivada en este caso está dada por

$$f'(\bar{x}) \bar{y} = 2x_N \bar{y}_N.$$

Observamos que la función es positiva para  $\bar{x} \neq 0$ , de donde el valor mínimo es cero. Podemos considerar la sucesión

$$\bar{x}^{(1)} = (1, 0, 0, \dots),$$

$$\bar{x}^{(2)} = (0, 1, 0, \dots),$$

$$\vdots$$

$$\bar{x}^{(n)} = (0, \dots, 0, 1, 0, \dots)$$

$$\vdots$$

en la esfera unitaria, para la cual

$$f(\bar{x}^{(n)}) \xrightarrow{n \rightarrow \infty} 0,$$

$$f'(\bar{x}^{(n)}) \xrightarrow{n \rightarrow \infty} 0.$$

Es decir, es una sucesión minimizante tal que  $f'(\bar{x}^{(n)}) \rightarrow 0$ .

Sin embargo, la sucesión no es ni contiene una subsucesión convergente, ya que

$$\|\bar{x}^{(m)} - \bar{x}^{(n)}\| = 2$$

para cualesquiera  $m \neq n$ .

Observemos que aunque la sucesión del ejemplo anterior está en la bola cerrada unitaria, no contiene ninguna subsucesión convergente. Aquí el problema es que el espacio  $l_1$  en que se está definiendo la función es de dimensión infinita y por tanto la bola unitaria cerrada no es compacta (Kolmogorov y Fomin, 1970).

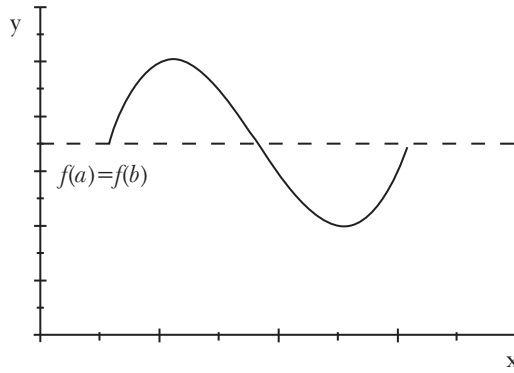
Hemos visto que para encontrar puntos en donde se tengan máximos o mínimos tendrán que estudiarse las propiedades de la función y su dominio de definición. Pero aun si se sabe de la existencia de estos extremos en muchos casos no se sabe la forma explícita de estos puntos.

Los teoremas existentes acerca de la obtención de puntos críticos de una función nos pueden ayudar a conocer otras propiedades intrínsecas de los puntos críticos, como lo veremos más adelante.

## IV. EXISTENCIA DE PUNTOS CRÍTICOS. TEOREMA DE PASO DE MONTAÑA

Uno de los resultados más importantes que estudiamos en los cursos elementales de cálculo es el siguiente.

**Gráfica IV.1**  
**Teorema de Rolle**



**Teorema 6 (de Rolle).** Sea  $f: \mathbb{R} \rightarrow \mathbb{R}$  continua, y derivable en el intervalo abierto  $(a, b)$  y tal que  $f(a) = f(b)$ . Entonces existe  $x_0 \in (a, b)$  tal que  $f'(x_0) = 0$ .

Este resultado nos asegura la existencia de un punto crítico que se encuentra dentro del intervalo  $(a, b)$ , aunque pudiera ser que esta función fuera constante, en cuyo caso no tendría interés alguno en nuestros problemas de optimización.

Debemos tener en cuenta además que el teorema nos da la existencia de un punto crítico, aunque no sea máximo o mínimo, al cual se le llama punto de inflexión. La demostración de este teorema se puede ver en Spivak (1996).

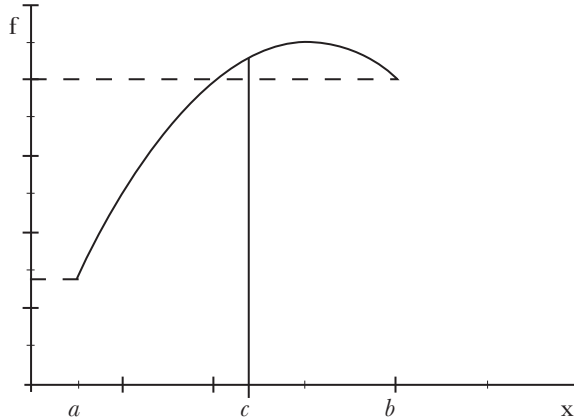
Una consecuencia del Teorema de Rolle es el siguiente resultado.

**Teorema 7.** Sea  $f: \mathbb{R} \rightarrow \mathbb{R}$  continua, y derivable en el intervalo abierto  $(a, b)$  y tal que existe  $c \in (a, b)$  con la condición

$$\max\{f(a), f(b)\} < f(c).$$

Entonces existe un punto crítico  $x_0 \in (a, b)$ .

**Gráfica IV.2**  
**Consecuencia del Teorema de Rolle**



**Demostración.** Supongamos que  $f(a) \leq f(b) < f(c)$ . De la continuidad de la función y el teorema del valor intermedio podemos encontrar  $y \in (a, c)$  tal que  $f(y) = f(b)$ . Para el intervalo  $[y, b]$  aplicamos el Teorema de Rolle y encontramos un punto crítico  $x_0$  en  $(a, b)$ .

Este teorema, al igual que el anterior, nos puede ayudar a aproximarnos a los puntos críticos. Su interpretación gráfica sería la siguiente: si se fijan los extremos  $f(a)$  y  $f(b)$  de  $f$  y se sabe que existe otro valor  $f(c)$  por arriba de éstos, entonces la gráfica debe tener una cresta por la que se debe subir para ir de  $f(a)$  a  $f(b)$  pasando por  $f(c)$ , es decir, tiene un máximo que en particular es punto crítico.

Existen generalizaciones de este último teorema a dominios mucho más grandes que  $\mathbb{R}$ . Éstos son los llamados Teoremas de Paso de Montaña. Enunciaremos y demostraremos uno de estos teoremas en el caso de dimensión finita, para lo cual necesitamos dar algunas definiciones.

**Definición 8.** Decimos que una sucesión  $\{x_n\}$  en un espacio lineal normado  $X$  es **sucesión de Cauchy** si dada  $\varepsilon > 0$ , existe  $N \in \mathbb{N}$  tal que  $\|x_n - x_m\| < \varepsilon \forall n, m \geq N$ . También decimos que  $X$  es **completo** si toda sucesión de Cauchy en  $X$  converge. Un espacio lineal normado y completo se llama **espacio de Banach**.

**Definición 9.** Una función  $\phi: W \subset X \rightarrow \mathbb{R} \cup \{\infty\}$ , con  $X$  un espacio de Banach, es **coerciva** si  $\phi(u) \rightarrow \infty$  cuando  $\|u\| \rightarrow \infty$ .

**Teorema 10 (de Paso de Montaña en dimensión finita).** Sea  $\phi: \mathbb{R}^N \rightarrow \mathbb{R}$  una función derivable con derivada continua, coerciva y tal que  $\phi$  tiene dos mínimos locales  $x_1$  y  $x_2$ . Entonces  $\phi$  tiene un tercer punto crítico,  $x_3$ , que no es mínimo local de  $\phi$  y tal que

$$\phi(x_3) = \inf_{\gamma \in \Gamma} \max_{x \in \gamma} \phi(x) =: c,$$

donde  $\Gamma := \{\gamma \subset \mathbb{R}^N: x_1, x_2 \in \gamma \text{ y } \gamma \text{ es subconjunto compacto y conexo}\}$ .

**Demostración.** Considérese una sucesión  $\gamma_m$  en  $\Gamma$  que hace que la sucesión numérica  $\max_{x \in \gamma_m} \phi(x)$  tienda a  $c$ , es decir, al posible valor más pequeño del conjunto  $\{\max_{x \in \gamma} \phi(x): \gamma \in \Gamma\}$ , así podemos decir que la sucesión  $\max_{x \in \gamma_m} \phi(x)$  es minimizante. De la coercividad de  $\phi$  tenemos que  $\gamma_m$  debe ser uniformemente acotado, ya que de otra forma existirían puntos  $x_{\gamma_m}$  en cada  $\gamma_m$  tal que  $\|x_{\gamma_m}\| \rightarrow \infty$ , y  $\phi(x_{\gamma_m}) \rightarrow \infty$  si  $m \rightarrow \infty$ , de donde nuestra sucesión minimizante tendería a infinito. Luego podemos definir

$$\tilde{\gamma} := \bigcap_{m \geq 1} \overline{\bigcup_{l \geq m} \gamma_l}.$$

Observamos que  $x_1$  y  $x_2$  están en  $\tilde{\gamma}$ , y además  $\tilde{\gamma}$  se obtiene como una intersección numerable de conjuntos compactos y conexos que se contienen, entonces  $\tilde{\gamma}$  es compacto y conexo. Con esto probamos que  $\tilde{\gamma} \in \Gamma$ .

Entonces

$$\max_{x \in \tilde{\gamma}} \phi(x) \geq \inf_{\gamma \in \Gamma} \max_{x \in \gamma} \phi(x) = c \quad . \quad (2)$$

Por otra parte, de la continuidad de  $\phi$  se tiene

$$\max_{x \in \tilde{\gamma}} \phi(x) \leq \limsup_{m \rightarrow \infty} (\max_{x \in \gamma_m} \phi(x)) = c$$

Por tanto,

$$\max_{x \in \tilde{\gamma}} \phi(x) = c, \quad (3)$$

de donde hemos encontrado un conjunto  $\tilde{\gamma}$  en  $\Gamma$  para el cual el valor  $c$  se alcanza, más aún, como  $\tilde{\gamma}$  es un conjunto compacto podemos encontrar al menos un punto  $\tilde{x} \in \tilde{\gamma}$  tal que  $\phi(\tilde{x}) = c$ .

Sin embargo, aun considerando una sucesión minimizante no podemos afirmar que el punto  $\tilde{x}$  es un punto crítico. Probemos que existe al menos uno de estos que sí es punto crítico.

Por la hipótesis sobre los puntos  $x_1$  y  $x_2$ , se tiene que

$$\max\{\phi(x_1), \phi(x_2)\} < c, \tag{4}$$

ya que tanto  $x_1$  como  $x_2$  son mínimos locales, y por tanto hay vecindades de  $x_1$  y de  $x_2$  para las cuales los valores de  $\phi$  en  $\tilde{\gamma}$  son mayores que los valores en los mínimos.

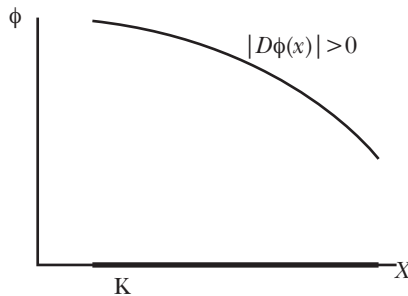
También de la continuidad de  $\phi$ , imagen inversa del valor  $c$

$$K: = \{x \in \tilde{\gamma} : \phi(x) = c\}$$

es cerrado, y es no vacío por (3), más aún, como estamos en un espacio de dimensión finita, es compacto. En este conjunto  $K$  probaremos que hay puntos críticos, lo haremos por contradicción y en varios pasos.

**Paso 1.** Ahora supongamos que los elementos en  $K$  no son puntos críticos, i.e.,  $D\phi(x) \neq 0$  para cada  $x \in K$ . Así, estamos suponiendo que el comportamiento de  $\phi$  en  $K$  es como el de la siguiente gráfica

**Gráfica IV.3**  
**Comportamiento de la gráfica de  $\phi$**



Como la derivada de la función  $\phi$  también es continua entonces existe  $\delta > 0$  tal que  $|D\phi(x)| \geq \delta$  para toda  $x \in K$ . De donde, usando la continuidad de  $D\phi$ , podemos encontrar  $\varepsilon > 0$  tal que en la vecindad abierta de  $K$ ,

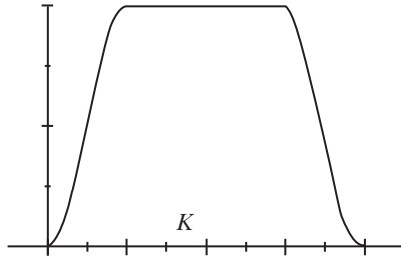
$$V := \{x \in \mathbb{R}^N : \exists y \in K, |x - y| < \varepsilon\}$$

se tiene que  $|D\phi(x)| \geq \frac{\delta}{2}$  para  $x \in V$ .

Podemos suponer que tanto  $x_1$  como  $x_2$  no están en  $V$ , pues de otro modo tomamos el valor de  $\varepsilon$  más pequeño.

**Paso 2.** Definamos a  $\eta$  como una función de corte, es decir, una función que se anule fuera de  $V$ , definida como 1 en  $K$  y que toma valores entre 1 y 0.

**Gráfica IV.4**  
**Función de corte**



Consideramos ahora el gradiente de  $\phi$

$$\nabla\phi(x) \cdot v = D\phi(x)v \quad \forall v \in \mathbb{R}^N.$$

**Paso 3.** Definimos una nueva función continua  $H: \mathbb{R}^N \times \mathbb{R} \rightarrow \mathbb{R}^N$  dada por

$$H(x, s) := x - s\eta(x)\nabla\phi(x),$$

que como se observa es diferenciable y continua con respecto a la variable  $s$  y además

$$\begin{aligned} \frac{d}{ds} \phi(H(x, s)) \Big|_{s=0} &= (\nabla\phi(H(x, s)) \cdot \frac{d}{ds} H(x, s)) \Big|_{s=0} \\ &= \nabla\phi(x) \cdot (-\eta(x)\nabla\phi(x)) \\ &= -\eta(x) |\nabla\phi(x)|^2 \leq -\eta(x) \frac{\delta^2}{4}, \end{aligned}$$

para toda  $x \in V$  donde  $\eta$  no se anula. Notemos que esto nos dice que la función  $\phi(H(x,s))$  es decreciente respecto a la variable  $s$ , que es lo que se espera, ya que la función  $H$  está definida en términos del gradiente negativo de  $\phi$ . Además por la continuidad de la derivada de  $\phi$  existe  $s_0 > 0$  tal que

$$\frac{d}{ds} \phi(H(x,s)) \leq -\frac{\eta(x)}{2} |\nabla\phi(x)|^2 \quad \forall s \in [0, s_0]. \quad (5)$$

**Paso 4.** Ahora para cada  $x \in \tilde{\gamma}$  definimos un nuevo conjunto en  $\Gamma$

$$\gamma_{s_0} := H(x, s_0).$$

Vamos a demostrar que  $\max_{x \in \gamma_{s_0}} \phi(x) < c$ . Para esto utilizamos el teorema fundamental del cálculo y (5)

$$\begin{aligned} \phi(H(x, s_0)) &= \phi(x) + \int_0^{s_0} \frac{d}{ds} \phi(H(x, s)) ds \\ &\leq \phi(x) - s_0 \frac{\eta(x)}{2} |\nabla\phi(x)|^2. \end{aligned}$$

Analizando esta última desigualdad observamos lo siguiente: si  $x \notin K$ , entonces, de (3),

$$\phi(x) - s_0 \frac{\eta(x)}{2} |\nabla\phi(x)|^2 \leq \phi(x) < c;$$

y si  $x \in K$ , entonces  $\eta(x) = 1$  y

$$\phi(x) - s_0 \frac{\eta(x)}{2} |\nabla\phi(x)|^2 \leq c - \frac{s_0}{2} \delta^2 < c.$$

De lo anterior

$$\max_{x \in \gamma_{s_0}} \phi(x) < c.$$

Como sabemos, la continuidad respeta la compacidad y la conexidad de los conjuntos, por lo que  $H$  manda compactos conexos en compactos conexos, y por la definición de  $\gamma_{s_0}$  concluimos que este conjunto también está en  $\Gamma$ .

Pero entonces hemos encontrado un conjunto en  $\Gamma$  de tal forma que su valor máximo está por debajo del ínfimo (2) de los valores máximos de los conjuntos en  $\Gamma$  lo cual es una contradicción.

Por tanto, el conjunto  $K$  sí tiene puntos críticos.

Hasta aquí sólo se ha probado que es posible encontrar sucesiones numéricas que minimizan al conjunto  $\{\max_{x \in \gamma} \phi(x) : \gamma \in \Gamma\}$ , estos valores se alcanzan en un subconjunto en  $\Gamma$ , y además que dentro de este conjunto existen puntos que pueden ser puntos críticos. Falta verificar que existe un punto crítico en  $K$  que no es mínimo local.

Supongamos que todos los puntos críticos en  $K$  son mínimos locales, para  $\phi$  en  $\mathbb{R}^N$ . Sea  $\tilde{K} := \{x \in \tilde{\gamma} : \phi(x) = c \text{ y } x \text{ mínimo local para } \phi \text{ en } \mathbb{R}^N\}$ . De la continuidad de  $\phi$  y por ser imagen inversa de  $c$  entonces  $\tilde{K}$  es cerrado. Además también es abierto por lo siguiente. Sea  $x_0 \in \tilde{K}$  que por ser mínimo local cumple

$$\phi(x_0) = c = \inf_{\gamma \in \Gamma} \max_{x \in \gamma} \phi(x) \leq \phi(y)$$

para toda  $y$  en una vecindad abierta  $V_{x_0}$  de  $x_0$ , y más aún  $\phi(x_0) = \max_{x \in \tilde{\gamma}} \phi(x) \leq \phi(y)$ ,  $\forall y \in V_{x_0} \cap \tilde{\gamma}$ , entonces

$$\max_{x \in \tilde{\gamma} \cap V_{x_0}} \phi(x) \leq \phi(x_0) \leq \phi(y) \quad \forall y \in V_{x_0} \cap \tilde{\gamma},$$

y esto se cumple sólo si  $\phi$  es constante en la vecindad local  $V_{x_0} \cap \tilde{\gamma}$ , es decir, esta vecindad de  $x_0$  es un abierto contenido en  $\tilde{K}$ , de donde podemos afirmar que  $\tilde{K}$  es abierto. Entonces como  $\tilde{K}$  es abierto y cerrado en el conexo  $\tilde{\gamma}$ , se tiene que  $\tilde{K} = \tilde{\gamma}$ . Pero esta igualdad nos lleva a una contradicción si observamos que  $x_1$  y  $x_2$  están en  $\tilde{\gamma}$ , y cumplen (4).

En este teorema la coercividad nos asegura la existencia del punto crítico. En general para  $\mathbb{R}^N$ , esta propiedad nos ayuda para asegurar la convergencia de una sucesión que minimiza un conjunto y que además converge a un punto crítico, y que aun cuando no es un mínimo local, sí lo es en un subconjunto del dominio.

## V. EL TEOREMA DE PASO DE MONTAÑA (TPM) EN DIMENSIÓN INFINITA

### Las sucesiones de Palais-Smale

Lo ilustrativo de la demostración del Teorema de Paso de Montaña (TPM) es el uso de sucesiones minimizantes. Formalicemos el tipo de sucesiones que hemos estudiado en la siguiente definición. El concepto se enuncia de una forma análoga en su versión original en Palais (1963) y Palais y Smale (1964).

**Definición 11.** Una sucesión  $\{x_n\}$  es de **Palais-Smale** para  $f$  en  $c$  si  $f(x_n) \rightarrow c$  y  $Df(x_n) \rightarrow 0$ . Y decimos que una sucesión de Palais-Smale satisface **la condición de Palais-Smale en  $c$**  si tiene una subsucesión convergente.

**Teorema 12.** Si  $\phi: \mathbb{R}^N \rightarrow \mathbb{R}$  es de clase  $C^1$  y es coerciva, entonces  $\phi$  satisface la condición de Palais-Smale.

**Demostración.** Sea  $(x_n)$  una sucesión de Palais-Smale para  $\phi$  en un valor  $c < \infty$ , es decir, se cumple

$$\begin{aligned}\phi(x_n) &\rightarrow c \\ \|D\phi(x_n)\| &\rightarrow 0\end{aligned}$$

De la coercividad deducimos que la sucesión es acotada, pues de otro modo si  $|x_n| \rightarrow \infty$ , entonces  $\phi(x_n) \rightarrow \infty$ , lo cual no es posible. Entonces cada sucesión de Palais-Smale es acotada en  $\mathbb{R}^N$ , luego por el Teorema de Bolzano-Weierstrass (Protter, 1998), esta sucesión tiene una subsucesión  $(x_{n_i})$  convergente a algún punto en  $\tilde{X} \in \mathbb{R}^N$ .

Además por la continuidad de las funciones  $\phi$  y  $D\phi$  se tiene que  $\phi(x_{n_i}) \rightarrow \phi(\tilde{x}) = c$  y  $\|D\phi(x_{n_i})\| \rightarrow \|D\phi(\tilde{x})\| = 0$ , es decir, que  $\tilde{x}$  es un punto crítico de  $\phi$ .

### El caso en dimensión infinita

Dentro de los problemas como el del Ejemplo 2, en los que la dimensión del espacio es infinita, el método de solución más interesante pero complicado de analizar. En el ejemplo se construía un funcional  $E$  definido a su vez en funciones  $u$ . Para establecer de manera correcta el método seguido se requiere una estructura más restrictiva, en cierto sentido, para el espacio de funciones  $u$  permitidas.

**Definición 13.** Un **espacio de Hilbert** es un espacio lineal con producto escalar, que es separable, completo y de dimensión infinita.

En la generalización al caso de dimensión infinita la dificultad para la obtención del punto crítico radica en que ya no se pueden aplicar resultados de compacidad. Recordemos que un conjunto que es cerrado y acotado en un espacio de dimensión infinita no necesariamente es compacto. Para este caso se tienen extensos trabajos sobre ejemplos particulares en los cuales se estudia cuándo se cumple la condición de Palais-Smale. El Teorema 10 se puede enunciar de la siguiente manera.

**Teorema 14 (de Paso de Montaña en dimensión infinita).** Sea  $X$  espacio de Hilbert,  $\phi: X \rightarrow \mathbb{R}$  una función de clase  $C^2$ ,  $r > 0$  y  $e \in X$  tal que  $\|e\| > r$  y

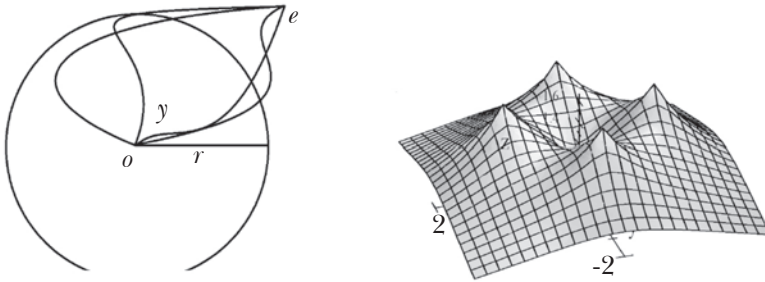
$$\inf_{\|u\|=-r} \phi(u) > \phi(0) \geq \phi(e).$$

Sea  $\Gamma := \{\gamma: [0,1] \rightarrow X: \gamma \text{ es continua y } \gamma(0)=0, \gamma(1)=e\}$  y definimos

$$c := \inf_{\gamma \in \Gamma} \max_{t \in [0,1]} \phi(\gamma(t))$$

Entonces existe una sucesión  $(u_k)$  en  $X$  tal que  $\phi(u_k) \rightarrow c$  y además  $\|\nabla \phi(u_k)\| \rightarrow 0$ .

**Figura V.1**  
**Comportamiento de paso de montaña**



La demostración requiere conocimiento del Análisis no Lineal para trabajar en espacios de dimensión infinita y puede verse por ejemplo en Willem (1996) y Jabri, (2003).

## VI. APLICACIONES DEL TPM

Daremos un ejemplo de la aplicación de este teorema. Consideramos la siguiente ecuación diferencial parcial elíptica de segundo orden, la cual es un tipo de ecuación de Schrödinger estacionaria:

$$\begin{aligned} -\Delta u - \lambda u &= u^{p-1} && \text{en } \Omega \subset \mathbb{R}^N, \\ u(x) &= 0 && \text{en } \partial\Omega, \\ u(x) &\geq 0. \end{aligned}$$

Aquí vamos a tomar  $p \in [2, 2^*)$ , donde  $2^* = \frac{2N}{N-2}$ ,  $\lambda$  un número real y  $\Omega$  un conjunto abierto acotado y conexo en  $\mathbb{R}^N$ , para  $N \geq 3$ .

Si tomamos la ecuación diferencial y la multiplicamos por una función  $v$  que sea diferenciable, que se anule en la frontera, y usando la fórmula de Green, tenemos las siguientes implicaciones

$$\begin{aligned} -\Delta uv - \lambda uv &= u^{p-1}v \\ \int_{\Omega} (\nabla u \cdot \nabla v - \lambda uv) &= \int_{\Omega} u^{p-1}v \\ \int_{\Omega} (\nabla u \cdot \nabla v - \lambda uv) - \int_{\Omega} u^{p-1}v &= 0 \end{aligned} \tag{6}$$

Se observa que el lado izquierdo de esta ecuación es la derivada de Fréchet del funcional

$$E(u) = \frac{1}{2} \int_{\Omega} (|\nabla u|^2 - \lambda u^2) - \frac{1}{p} \int_{\Omega} |u|^{p-1} u,$$

por lo que la ecuación (6) sugiere que un punto crítico del funcional  $E$  es solución de nuestra ecuación diferencial (Schwartz, 1969; Willem, 1996).

Notemos que el funcional  $E$  está definido en funciones  $u$  que deben ser diferenciables, por lo que en un principio podemos pensar que el mejor espacio en el que se puede realizar el análisis es el espacio de funciones con derivada continua definidas en  $\Omega$ , el cual se denota por  $C^1(\Omega)$ . Sin embargo este espacio no resulta ser el adecuado por lo siguiente. La ecuación (6) se puede escribir en términos de un producto escalar

$$\langle u, v \rangle := \int_{\Omega} (\nabla u \cdot \nabla v - \lambda uv),$$

y un funcional lineal

$$F(v) := \int_{\Omega} u^{p-1}v.$$

Por el Teorema de representación de Riesz-Fréchet, el cual nos dice que en un espacio completo, existe un elemento  $u$  en el espacio de tal forma que se da la relación

$$F(v) = \langle u, v \rangle. \tag{7}$$

Así nuestro problema estaría parcialmente resuelto, dado que (7) es equivalente a (6). Pero el espacio  $C^1(\Omega)$  no es completo (Kolmogorov y Fomin, 1970), por lo cual no podemos aplicar el Teorema de Riesz-Fréchet. Entonces es necesario buscar el espacio adecuado que sí sea completo.

Aquí aparecen los siguientes espacios con las propiedades importantes que se indican.

1. Los espacios  $L^p(\Omega) := \{u: \Omega \rightarrow \mathbb{R}^N: \int_{\Omega} |u|^p d\mu < \infty\}$ . Estos son espacios de Banach con norma  $\|u\|_p = (\int_{\Omega} |u|^p d\mu)^{\frac{1}{p}}$ , donde  $\mu$  es una medida de Lebesgue.
2. Los espacios de Sobolev  $H^1(\Omega) := \{u \in L^2(\Omega): \begin{matrix} \exists g_1, g_2, \dots, g_N \in L^2(\Omega) \\ \int_{\Omega} u \frac{\partial \varphi}{\partial x_i} = - \int_{\Omega} \varphi g_i \forall \varphi \in C_0^\infty(\Omega) \end{matrix} \}$  Son espacios de Hilbert con producto escalar

$$\langle u, v \rangle = \int_{\Omega} (\nabla u \cdot \nabla v + uv),$$

y con norma inducida  $\|u\|^2 = \int_{\Omega} (|\nabla u|^2 + u^2)$ . Las funciones  $g_i$  son llamadas derivadas débiles de  $u$  y suelen denotarse como las derivadas parciales  $\frac{\partial u}{\partial x_i} = g_i$ . En general, si  $u$  es diferenciable estas derivadas débiles coinciden con las derivadas parciales. El conjunto  $C_0^\infty(\Omega)$  denota al conjunto de funciones con derivadas continuas de todos los órdenes y definidas en  $\Omega$ .

3. El espacio  $H_0^1(\Omega)$  es la cerradura de  $C_0^\infty(\Omega)$  con la norma indicada en  $H^1(\Omega)$ . Este espacio es el adecuado debido a las condiciones de frontera de nuestra ecuación.

Así, buscar una solución de nuestra ecuación diferencial es buscar un punto crítico del funcional  $E$  definido en el espacio de Sobolev  $H_0^1(\Omega)$ . El parámetro  $\lambda$  de la ecuación diferencial nos deja ver otras características sobre el estudio de este tipo de ecuaciones.

**Definición 15.** Vamos a denotar por  $\lambda_1$  al **primer valor propio del operador**  $-\Delta$  en  $H_0^1(\Omega)$ , es decir, que existe una función  $e_1 \in H_0^1(\Omega)$ , distinta de cero tal que se cumple la ecuación  $-\Delta e_1 = \lambda_1 e_1$ .

Este valor propio cumple además

$$\lambda_1 = \inf_{u \in H_0^1(\Omega) - \{0\}} \frac{\|\nabla u\|_{L^2}^2}{\|u\|_{L^2}^2}$$

**Teorema 16.** Si  $2 < p < 2^* = \frac{2N}{N-2}$ , la ecuación diferencial tiene una solución no trivial si y sólo si  $\lambda < \lambda_1$ .

**Demostración.** Si  $e_1 \in H_0^1(\Omega)$  es una función propia de  $-\Delta$  y  $u$  es una solución no trivial entonces

$$-\lambda \int_{\Omega} u e_1 = \int_{\Omega} (u^{p-1} + \Delta u) e_1 > \int_{\Omega} \Delta u e_1 = \int_{\Omega} \Delta e_1 u = -\lambda_1 \int_{\Omega} e_1 u,$$

de donde  $\lambda < \lambda_1$ . Ahora vamos a demostrar la otra implicación, lo haremos en dos pasos.

**Paso 1.** Probaremos que la norma  $\|u\|_{\lambda}^2 = |\Delta u|_{L^2}^2 - \lambda|u|_{L^2}^2$  es equivalente a la norma usual en  $H_0^1(\Omega)$ . Si  $\lambda < 0$ , por la definición de  $\lambda_1$  se tiene que

$$\begin{aligned} |\nabla u|_{L^2}^2 - \lambda|u|_{L^2}^2 &= |\nabla u|_{L^2}^2 - \frac{\lambda}{\lambda_1}\lambda_1|u|_{L^2}^2 \\ &\leq |\nabla u|_{L^2}^2 - \frac{\lambda}{\lambda_1}|\nabla u|_{L^2}^2 \\ &= \left(1 - \frac{\lambda}{\lambda_1}\right)|\nabla u|_{L^2}^2, \end{aligned}$$

además  $\left(1 - \frac{\lambda}{\lambda_1}\right) > 0$ , porque  $\lambda < \lambda_1$ . Ahora si  $0 < \lambda < \lambda_1$  es claro que  $|\nabla u|_{L^2}^2 - \lambda|u|_{L^2}^2 \leq |\nabla u|_{L^2}^2$ , de donde se concluye que  $|\nabla u|_{L^2}^2 - \lambda|u|_{L^2}^2 \leq C|\nabla u|_{L^2}^2$  con  $C$  una constante positiva. De manera análoga, considerando los mismos casos para el valor de  $\lambda$ , se prueba que existe  $C_1 > 0$  tal que  $|\nabla u|_{L^2}^2 - \lambda|u|_{L^2}^2 \geq C_1|\nabla u|_{L^2}^2$ . Lo cual nos permite considerar una norma equivalente definida como  $\|u\|_{\lambda}^2 = |\nabla u|_{L^2}^2 - \lambda|u|_{L^2}^2$  en  $H_0^1(\Omega)$ .

**Paso 2.** Para finalizar, vamos a verificar las condiciones del Teorema de Paso de Montaña para el funcional  $E$ , recordando la desigualdad debida a Sobolev (Brezis, 1984; Willem, 1996; Struwe, 1990).

$$|u|_{L^p} \leq k \left\{ \int_{\Omega} |\nabla u|^2 \right\}^{\frac{1}{2}} \text{ para } p \in [2, 2^*]$$

obtenemos

$$\begin{aligned} E(u) &\geq \frac{1}{2}\|u\|_{\lambda}^2 - \frac{1}{p}|u|_{L^p}^p \\ &\geq \frac{1}{2}\|u\|_{\lambda}^2 - \frac{c}{p}\|u\|_{\lambda}^p, \end{aligned}$$

con  $c$  constante positiva, por lo que podemos encontrar  $r > 0$  tal que

$$\inf_{\|u\|_{\lambda}=r} E(u) > 0 = E(0),$$

ya que  $p > 2$  y cualquier número suficientemente pequeño elevado a la  $p$  es menor que elevado al cuadrado. Si fijamos una función  $u_0 > 0$  en  $H_0^1(\Omega)$  y consideramos todas aquellas funciones de la forma  $tu_0$  para  $t \in [0, \infty)$ , podemos modificar nuestro funcional

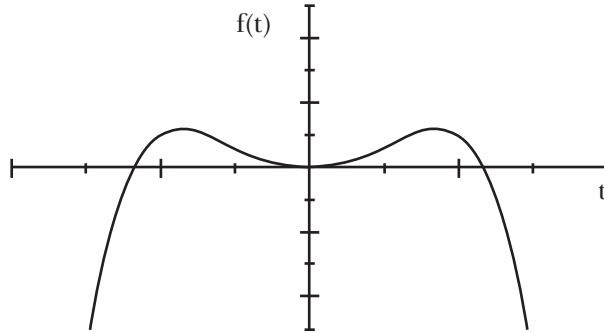
$$\begin{aligned}
 E(tu_0) &= \frac{1}{2} \int_{\Omega} (|tu_0|^2 + \lambda(tu_0)^2) - \frac{1}{p} \int_{\Omega} |tu_0|^p \\
 &= \frac{1}{2} t^2 \int_{\Omega} (|u_0|^2 + \lambda u_0^2) - \frac{1}{p} |t|^p \int_{\Omega} |u_0|^p.
 \end{aligned}$$

De esta manera, como  $u_0$  y por tanto las integrales que aparecen adquieren valores fijos, entonces estamos analizando una función real de la forma

$$f(t) = C_1 t^2 - C_2 t^p,$$

donde  $C_1$  y  $C_2$  son constantes positivas.

**Figura VI.1**  
**Gráfica de  $f(t) = C_1 t^2 - C_2 t^p$**



La gráfica de esta función nos muestra claramente que podemos encontrar  $e$  en  $H_0^1(\Omega)$  tal que  $E(e) \leq E(0)$ , por lo que se cumplen las condiciones del Teorema de Paso de Montaña.

De esta forma podemos encontrar una sucesión de Palais–Smale para el funcional  $E$  en  $d$  de tal forma que

$$d = \sup_n E(u_n) < \infty \text{ y } DE(u_n) \rightarrow 0.$$

Probaremos que esta sucesión satisface la condición de Palais–Smale y converge a un punto crítico del funcional  $E$ . Realizando algunos sencillos cálculos se tiene lo siguiente

$$\begin{aligned} d + 1 + \|u_n\|_\lambda &\geq E(u_n) - \frac{1}{p} DE(u_n)u_n \\ &= \left(\frac{1}{2} - \frac{1}{p}\right) \|u_n\|_\lambda^2, \end{aligned}$$

Podemos deducir que  $u_n$  es una sucesión acotada, pues de lo contrario estaríamos llegando a una contradicción, ya que la desigualdad indicaría que una expresión lineal tiende a infinito más rápido que una cuadrática.

Mencionamos ahora tres resultados clásicos del análisis no lineal aplicados a nuestra sucesión acotada (Royden, 1988; Brezis, 1984; Jost, 2005):

1. Si  $F$  es cualquier operador lineal definido en  $H_0^1(\Omega)$ , entonces para nuestra sucesión de Palais-Smale,  $F(u_n) \rightarrow F(u)$ .
2. Por un teorema debido a Rellich y Kondrakov (Brezis, 1984; Willem, 1996; Struwe, 1990), y como  $p < 2^*$ , se tiene una subsucesión de  $u_n$  (que denotamos de la misma forma) tal que  $u_n \rightarrow u$  en  $L^p(\Omega)$ .
3.  $u_n^{p-1} \rightarrow u^{p-1}$  en  $L^q(\Omega)$  con  $q = \frac{p}{p-1}$ .

Con lo anterior podemos observar que

$$(DE(u_n) - DE(u))(u_n - u) \rightarrow 0$$

ya que la derivada es un operador lineal. Además por la desigualdad de Hölder

$$\int_\Omega (u_n^{p-1} - u^{p-1})(u_n - u) \leq \|u_n^{p-1} - u^{p-1}\|_{L^q} \|u_n - u\|_{L^p},$$

por lo que  $\int_\Omega (u_n^{p-1} - u^{p-1})(u_n - u) \rightarrow 0$ .

Finalmente observamos que

$$\begin{aligned} (DE(u_n) - DE(u))(u_n - u) &= \int_\Omega (|\nabla(u_n - u)|^2 + \lambda(u_n - u)^2) - \int_\Omega (u_n - u)^p \\ &= \|u_n - u\|_\lambda^2 - \int_\Omega (u_n^{p-1} - u^{p-1})(u_n - u). \end{aligned}$$

Despejando obtenemos

$$= \|u_n - u\|_\lambda^2 = (DE(u_n) - DE(u))(u_n - u) - \int_\Omega (u_n^{p-1} - u^{p-1})(u_n - u),$$

y por lo tanto

$$\|u_n - u\|_\lambda^2 \rightarrow 0.$$

Así encontramos que, en efecto,  $u_n \rightarrow u$  en  $H_0^1(\Omega)$ , con las propiedades

$$\begin{aligned} E(u_n) &\rightarrow E(u) = d \\ DE(u_n) &\rightarrow DE(u) = 0, \end{aligned}$$

es decir,  $u$  es un punto crítico del funcional  $E$ .

Es importante mencionar que para pedir que una función sea punto crítico del funcional basta que tenga derivada continua, lo cual no es suficiente para que sea solución de nuestra ecuación diferencial, pues debe aplicársele el operador laplaciano que involucra segundas derivadas. Esto se resuelve con los llamados Teoremas de Regularidad (Brezis, 1984; Struwe, 1990), que se prueban con teoría muy extensa ya desarrollada, la cual indica que la suficiente diferenciabilidad de la frontera del dominio nos va a determinar la diferenciabilidad del punto crítico, y por lo tanto afirma que sea la solución de nuestra ecuación diferencial.

En este ejemplo se mostró una técnica para saber acerca de la existencia de una solución de la ecuación diferencial dada. La teoría en el sentido formal con variantes en la forma de la ecuación diferencial elíptica implica más trabajo como, por ejemplo, definir otro tipo de espacios de Sobolev, los cuales deben mantener su estructura de espacios lineales y completos para poder aplicar el Teorema de Paso de Montaña, estudiar las desigualdades que se utilizaron, como las de Sobolev, de Hölder, además de los tipos de encajes entre los espacios de Sobolev y los espacios  $L^p$ . Gran teoría se ha desarrollado para resolver ecuaciones diferenciales parciales de segundo orden y de tipo elíptico, con los Métodos Variacionales.

#### REFERENCIAS

- Bartsch, Thomas (1993). *Topological Methods for Variational Problems with Symmetries (Lecture Notes in Mathematics)*. Springer. Berlín.
- Brézis, Haim (1984). *Análisis Funcional*. Alianza Universidad Textos. Madrid.
- Gelfand, I. M. y Fomin, S. V. (1963). *Calculus of Variations*. Englewood Cliffs, Prentice-Hall Inc. N. J.
- Jabri, Youssef (2003). *The Mountain Pass Theorem: Variants, Generalizations and Some Applications*. Cambridge University Press.
- Jost, Jürgen (2005). *Postmodern Analysis*. 3a. ed. Springer. Berlín/Heidelberg.

- Kolmogorov, A. N. y Fomin, S. V. (1970). *Introductory Real Analysis*. Englewood Clift, Prentice-Hall Inc. N. J.
- Palais, R. S. (1963). "Morse Theory on Hilbert Manifold". *Topology* 2, 299-340.
- y Smale, S. (1964). "A Generalized Morse Theory". *Bull. Amer. Math. Soc.* 70, 165-171.
- Protter, M. (1998). *Basics Elements of Real Analysis*. Springer Verlag. Nueva York.
- Rabinowitz, Paul H. (1986). "Minimax Methods in Critical Point Theory with Applications to Differential Equations". Providence Rhode Island: Conference Board of the Mathematical Sciences Regional Conference Series in Mathematics 69, AMS.
- Royden, H. L. (1988). *Real Analysis*. 3a. ed. Englewood Cliffs, Prentice Hall. N. J.
- Rudin, Walter (1973). *Functional Analysis*. McGraw-Hill Book Company. EU.
- (1964). *Principles of Mathematical Analysis*. 3a. ed. McGraw-Hill Inc. Nueva York.
- Schwartz, Jacob (1969). *Nonlinear Functional Analysis*. Gordon and Breach Science Publisher. Nueva York.
- Spivak, Michael (1996). *Cálculo Infinitesimal*. 2a. ed. Reverté. México.
- Struwe, Michael (1990). *Variational Methods. Applications to Nonlinear Partial Differential Equations and Hamiltonian Systems*. Springer-Verlag. Berlín/Heidelberg.
- Willem, Michel (1996). *Minimax Theorems*. Birkhäuser. Boston/Basilea/Berlín.

# INFLACIÓN COSMOLÓGICA VISTA DESDE LA MECÁNICA CUÁNTICA SUPERSIMÉTRICA\*

*J. Socorro\*\**  
*Miguel Sabido\*\*\**  
*W. Guzmán\*\*\*\**  
*Máximo A. Agüero\*\*\*\*\**

## Resumen

En este artículo tratamos con el problema de construir el potencial de campo escalar apropiado para un escenario inflacionario. En este sentido usamos el método de Witten (1981) en la Mecánica Cuántica Supersimétrica en el contexto de variables de Grassman aplicado al modelo cosmológico de Friedmann-Robertson-Walker (FRW), incluyendo un fluido perfecto de materia ordinaria y ecuación de estado barotrópica  $p = \gamma\rho$  en la etapa inflacionaria ( $\gamma = -1$ ), así como un potencial dependiente de un campo escalar  $V(\varphi(t))$ . Durante el proceso de encontrar soluciones cuánticas empleando el formalismo canónico de Arnowitt-Deser-Misner (ADM), obtenemos una familia de potenciales que pueden satisfacer esta etapa del universo. Al considerar el proceso supersimétrico, esta familia de potenciales

\* Este trabajo fue parcialmente apoyado por Conacyt proyectos 62253, 135023 y 179881, proyecto de DAIP (2011-2012) y proyecto de Cuerpo Académico Promep. UGTO-CA-3. Red Promep de Gravitación y Física Matemática, bajo el proyecto *Aspectos Cuánticos de la Gravedad en Modelos Cosmológicos, Fenomenología y la Geometría del Espacio-Tiempo*. Este trabajo es parte de la colaboración como miembros del Instituto Avanzado de Cosmología. Muchos cálculos fueron hechos usando el programa de álgebra simbólico REDUCE 3.8.

\*\* Doctor en Física, Departamento de Física, DCEI, Universidad de Guanajuato-Campus León, A.P. E-143, C.P. 37150, León, Guanajuato, México, <josogadi@hotmail.com>. Departamento de Física, Universidad Autónoma Metropolitana, apartado Postal 55-534, C.P. 09340, México, D.F., México.

\*\*\* Doctor en Física, Departamento de Física, DCEI, Universidad de Guanajuato-Campus León, A.P. E-143, C.P. 37150, León, Guanajuato, México.

\*\*\*\* Doctor en Física, Departamento de Física, DCEI, Universidad de Guanajuato-Campus León, A.P. E-143, C.P. 37150, León, Guanajuato, México.

\*\*\*\*\* Ph. D. en Física-Matemática. Profesor de Tiempo Completo, Facultad de Ciencias, Universidad Autónoma del Estado de México, Instituto Literario 100, Toluca 50000, México, <maaguerog@uaemex.mx>.

es escaneada de tal suerte que el potencial exponencial dependiente del campo escalar  $V(\varphi) = V_0 e^{\lambda\varphi} = V_0 e^{-\nu^3\varphi}$  aparece como el más viable para garantizar este escenario, fijando el parámetro  $\lambda$ , el cual es menor al predicho mediante el proceso de obtener una solución clásica creciente tipo ley de potencias para el factor de escala de este modelo cosmológico de FRW. Por completez se presentan las contribuciones a la función de onda supersimétrica, tanto de la parte bosónica como fermiónica, así como de la densidad de probabilidad en este esquema.

**Palabras clave:** Mecánica cuántica supersimétrica, cosmología cuántica, ecuación de Wheeler-DeWitt, inflación.

## INTRODUCCIÓN

La más atractiva teoría de campo gravitacional es la teoría de la Relatividad General (RG) de Einstein. Ella describe la interacción gravitacional por la dinámica de la geometría del espacio-tiempo, pero desafortunadamente no ha sido posible construir hasta el momento la versión cuántica de la misma. En el estudio del universo temprano el proceso de inflación es el mejor candidato para explicar las observaciones a esos tiempos tempranos, por medio del campo escalar denominado inflatón, el cual genera una rápida aceleración del factor de escala en tiempos remotos, y como semilla para las perturbaciones primordiales necesarias para la formación de estructuras. El origen de estos potenciales escalares no se entiende en su totalidad, sólo se sabe que deben incluirse en los formalismos matemáticos para *pegarle* a los datos observacionales. Uno debería esperar que una teoría fundamental pudiera dar mayor luz sobre este tema; en particular, la pregunta que surge es ¿podría la teoría cuántica de la gravedad dar algunos indicios de cómo debiera ser esta construcción?

Es por ello que como aproximación en el estudio cuántico del universo uno empieza usando ciertos modelos cosmológicos en este régimen, con lo cual emerge la cosmología cuántica, con la idea de que ciertas implicaciones de este estudio se puedan extrapolar hacia la gravedad cuántica completa. En este sentido empezamos comentando que uno de los principales problemas de la cosmología inflacionaria es encontrar el mecanismo para derivar en forma natural el potencial de campo escalar más apropiado para encontrar los *e-folding* de la etapa de inflación del universo. Por “natural”, nosotros entendemos el mecanismo desde el cual alguna teoría provee el potencial de campo escalar que tenga las características convenientes de inflación que respondan a la evolución del universo. En este trabajo derivamos desde la perspectiva de la cosmología cuántica una familia de

potenciales que pueden describir el proceso inflacionario, vía sistemas integrables, y que en la literatura se han puesto a mano; que adicionando más tarde las ideas de la mecánica cuántica supersimétrica, la característica de la misma, hace que sean más restrictivas las condiciones para poder obtener un modelo inflacionario, con lo cual la familia de potenciales es escaneada, ofreciendo al final el potencial más viable.

En trabajos previos donde se usó como modelo juguete al modelo FRW se determinó una familia de potenciales que emergió al resolver la ecuación de Wheeler-DeWitt (WDW) (Guzmán *et al.*, 2007), ecuación que gobierna el escenario canónico de la cosmología cuántica y equivalente a la ecuación de Schrödinger de la cosmología estándar a energía cero. Recientemente, este tipo de potenciales escalares fueron obtenidos usando un esquema supersimétrico local en el contexto de supercampos (Escamilla-Rivera, 2009), así como en el esquema que trabajaremos en esta contribución usando el modelo cosmológico anisotrópico Bianchi I (Socorro y D'Oleire, 2010).

Hoy día es muy común hacer un estudio de la cosmología del universo introduciendo campos escalares  $\varphi$  como los agentes responsables de varios de los aspectos más intrigantes de nuestro universo (Peebles y Ratra, 2003; Lyth y Liddle, 2000; Ureña-López y Matos, 2000; Sahni y Wang, 2000; Matos y Ureña-López, 2004; Arbey *et al.*, 2001; Alam *et al.*, 2004; Sahni, 2004; Tegmark, 2005), tal como la etapa de inflación (Tsujiikawa, 2000; Copeland *et al.*, 1998), materia y energía oscura (Copeland *et al.*, 2006). La derivación usual de los potenciales escalares es una verdadera empresa que tiene el siguiente cuestionamiento: ¿qué procesos físicos nos provee el potencial de campo escalar adecuado que gobierne al universo en determinada época? Para responder esta pregunta para una etapa de inflación, desde el punto de vista teórico se usa la idea de que el universo en su evolución clásica entra en una etapa acelerada, caracterizado esto en el factor de escala del modelo correspondiente, el cual fija que el parámetro en un potencial puesto a mano tipo  $V_0 e^{\lambda\varphi}$ , el valor de  $\lambda < -\sqrt{2}$  (Chimento y Jakubi, 1996), o usando la cosmología cuántica, no se logra fijar quedando predeterminado entre los puntos críticos  $|\lambda| < \sqrt{3}$ , lo cual hace que se deban usar las ecuaciones de la cosmología cuántica supersimétrica, además de las herramientas de la cosmología cuántica supersimétrica, las cuales resultan ser más restrictivas debido a que hay más ecuaciones de restricción que se aplican a la función de onda con el fin de encontrar el potencial más viable para una época inflacionaria en esta escala de energías. En este último formalismo el parámetro queda fijado en  $|\lambda| < \sqrt{3}$  (Socorro y D'Oleire, 2010). Para hacer esto último se usa la supersimetría como una raíz cuadrada de la Relatividad General (RG) (Bene y Graham, 1994; Obregón *et al.*, 1993; Lidsey, 1995; Socorro y Medina, 2000) en la cual las variables

de Grassmann son empleadas como auxiliares y no pueden identificarse como los compañeros supersimétricos de las variables bosónicas; en este contexto más bien se habla de la contribución bosónica y fermiónica en la composición de la función de onda del universo. A partir de esta solución se encuentra que existe una tendencia para el vacío supersimétrico a permanecer cerca de sus límites semiclásicos, ya que las soluciones encontradas vienen siendo las de orden más bajo en la aproximación WKB y no corresponden a las soluciones cuánticas encontradas también en este trabajo.

Para poner en contexto el enfoque aquí empleado, brevemente describiremos las principales características de los diferentes formalismos supersimétricos aplicados a la cosmología.

1. Los definidos por medio del uso de la supersimetría, como la raíz cuadrada (Bene y Graham, 1994; Obregón *et al.*, 1993; Lidsey, 1995; Socorro y Medina, 2000). Como ya se mencionó, las variables de Grassmann no son identificadas como los compañeros supersimétricos de las correspondientes variables bosónicas cosmológicas.
2. La formulación de supercampos (Obregón *et al.*, 1996; Tkach *et al.*, 1996), la cual permite construir de manera sistemática la cosmología cuántica supersimetría local, obteniendo de manera directa los correspondientes compañeros supersimétricos fermiónicos, pudiendo incorporar materia en este contexto (Tkach *et al.*, 1997). Estos compañeros supersimétricos no son deducidos directamente del esquema del gravitino.
3. Modelos basados en supergravedad. Éstos han sido estudiados usando la formulación canónica de Arnowitt-Deser-Misner (ADM) y la descomposición espinorial de 4-componentes (Macías *et al.*, 1987; Socorro *et al.*, 1992; Obregón y Ramírez, 1998; Macías *et al.*, 1998). También han sido estudiados con las variables ADM y el formalismo espinorial de 2-componentes (D'Eath y Hughes, 1993; D'Eath *et al.*, 1993). En este esquema también la materia ha sido incluida, tanto en  $N=1$  y  $N=2$  supersimetrías (Cheng *et al.*, 1994; Cheng *et al.*, 1995; Moniz, 1996b; D'Eath, 1994). Además, las variables de Ashtekar han sido consideradas en el otro esquema canónico de la gravedad cuántica (Capovilla y Guven, 1994; Capovilla y Obregón, 1994).
4. Un cuarto formalismo que está inducido por la supersimetría se denomina análisis isoespectral (Mielnik, 1984; Nieto, 2004; Fernández, 1984; Cooper *et al.*, 1995; Cooper *et al.*, 2001; Socorro *et al.*, 2003). El análisis isoespectral consiste en encontrar una familia de potenciales que tengan los mismos eigenvalores que el potencial original (excepto el eigenvalor del estado base); con esto se obtiene en nuevo hamiltoniano para la familia de potenciales. Así,

podemos inferir que se podrán falsificar algunos potenciales y sus consabidas soluciones para un problema determinado por una ecuación diferencial en la cual se pueda usar el método isoespectral. La solución de este nuevo hamiltoniano también va a satisfacer a nuestro problema original, ya que el potencial original es parte de la familia de isopotenciales. En este contexto se trabajó el modelo Bianchi I con el potencial escalar del estilo que se maneja en este trabajo, donde la función de onda para el parámetro crítico presenta un pico pronunciado (D'Oleire, 2010).

Algunos de estos modelos y sus soluciones exactas han sido presentados en dos libros bien organizados (D'Eath, 1996; Moniz, 2010), así como en dos extensos artículos de revisión (Moniz, 1996b; Macías, 1999).

El potencial  $V_0 e^{\lambda\varphi}$  es un buen candidato, dependiendo del valor del parámetro  $\lambda$ , para estudiar inflación, materia y energía oscura o modelos taquiónicos (Sen, 2002, 2003; Gorini *et al.*, 2004; García-Compeán *et al.*, 2005). La ecuación de WDW puede ser resuelta usando un *ansatz* particular, el cual se asemeja al usado en la aproximación WKB de la mecánica cuántica, que se denomina la representación bohmiana (Bohm, 1952). Este método ha sido usado en la literatura (Obregón y Socorro, 1996) para encontrar soluciones exactas a los modelos cosmológicos Bianchi Clase A. Por otro lado, los mejores candidatos para la función de onda son aquellos que tienen un decaimiento con respecto a su factor de escala, en el sentido de que obtengamos una buena solución clásica usando la aproximación WKB en cualquier escenario en la evolución de nuestro universo (Hartle y Hawking, 1983).

Sobre la ecuación de WDW hay una cantidad enorme de artículos que tratan diferentes problemas de nuestro universo en el contexto cuántico. Por ejemplo, Gibbons y Grishchuk (1989) [50] se hacen la pregunta: ¿cuál es la función de onda típica para nuestro universo? En las referencias (Obregón *et al.*, 1996) aparece una excelente recopilación de artículos sobre cosmología cuántica donde se explica el problema de cómo el universo emerge de la singularidad conocida como big bang (gran explosión); además, un importante enfoque a este problema aparece en la propuesta de la función de onda, en el cual el universo podría ser autocontenido sin ninguna singularidad y sin ninguna edad. Sobre la estructura de los potenciales escalares, éstos aparecen en la literatura en conexión con la quintaesencia, materia y energía oscura escalar (Peebles y Ratra, 2003; Lyth y Liddle, 2000; Ureña-López y Matos, 2000; Sahni y Wang, 2000; Matos y Guzmán, 2000; Matos *et al.*, 2000; Matos y Ureña-López, 2000; Matos y Ureña-López, 2001; Lidsey *et al.*, 2002; Cabral-Rosetti *et al.*, 2002; Matos y Ureña-López, 2002; Alcubierre *et al.*, 2002).

La teoría cuántica de la gravedad (al menos en los primeros intentos de cuantización denominada canónica) está plagada de divergencias ultravioleta; en este enfoque de cuantización canónica es inicialmente aplicado a modelos cosmológicos con grandes simetrías –el cual se denomina mini-superespacio–, en los que las variables gravitacionales y de materia han sido reducidos a un número finito de grados de libertad. Por otro lado, supergravedad fue originalmente desarrollada como una teoría de campo elemental la cual debiera evitar las divergencias ultravioleta. Una forma de incorporar algunos aspectos de la supergravedad en la cosmología es trabajar con la cosmología cuántica supersimétrica, de esta manera se pueden capturar algunas características esenciales de la supergravedad y, al mismo tiempo, evitar varias de las dificultades técnicas de trabajar con la teoría completa.

Hay varios formalismos de la cosmología cuántica supersimétrica. En este trabajo presentamos el formalismo de Witten's. La idea de Witten (1981) es encontrar unos operadores denominados *supercargas*  $Q, \bar{Q}$  que produzcan un superhamiltoniano  $H_{ss}$ , y un álgebra cerrada supersimétrica (SUSY), con la condición de que la ecuación de WDW pueda ser obtenida como el sector bosónico de este superhamiltoniano en el superespacio, es decir, cuando todos los campos fermiónicos son congelados (puestos a cero). Esta idea ya se aplicó en la referencia [21] para todos los modelos cosmológicos anisotrópicos Bianchi Clase A.

Este trabajo es organizado de la siguiente manera. En la sección I presentamos la acción clásica con las correspondientes contribuciones a la densidad lagrangiana del modelo y vía transformaciones canónicas obtenemos la densidad hamiltoniana. Esta acción incluye una parte gravitacional  $S_{geométrica}$ ,  $S_{materia}$  corresponde a un fluido perfecto barotrópico con ecuación de estado  $p = \gamma\rho$ ,  $S_\varphi$  es la contribución del campo escalar, y finalmente incluimos el término cosmológico  $S_\Lambda$ . Además la densidad hamiltoniana clásica, así como la ecuación de Hamilton-Jacobi (H-J), son obtenidas; la densidad hamiltoniana juega el papel importante en la versión supersimétrica en el esquema de Witten, donde se introduce por primera vez la función superpotencial que corresponde a la fase de la función de onda en la aproximación semiclásica (Socorro y Medina, 2000), y también esencial en las soluciones supersimétricas. También la ecuación H-J es usada para construir la familia de potenciales, los cuales damos en la Tabla 1, y en particular nos enfocaremos al potencial exponencial. En la sección II damos una pequeña revisión de la propuesta de Witten para la mecánica cuántica supersimétrica. La sección III se usa para presentar de manera sistemática la solución al problema supersimétrico, dejando para la sección IV la construcción de la densidad de probabilidad no normalizada en el enfoque de Grassman; y finalmente la última sección la dedicamos a las conclusiones del trabajo.

## I. ACCIÓN CLÁSICA DEL MODELO

En cosmología uno asume ciertas simetrías, tomando como base el principio cosmológico el cual establece que el universo es homogéneo e isotrópico. Esta simetría de homogeneidad e isotropía del espacio-tiempo fue estudiada originalmente por Friedmann-Robertson-Walker (FRW). La simetría está codificada en la forma especial del siguiente elemento de línea

$$ds^2 = -N^2(t)dt^2 + e^{2\alpha(t)} \left[ \frac{dr^2}{1-\kappa r^2} + r^2 d\Omega^2 \right] \quad (1)$$

donde  $a(t)=e^{\alpha(t)}$  es el factor de escala;  $N(t)$  es la función de espaciamiento temporal, es decir, cómo podemos conectar una hipersuperficie a otra;  $\kappa$  es la constante de curvatura del espacio-tiempo que puede tomar los siguientes valores: 0, +1, -1, que corresponden a espacios-tiempo plano, cerrado y abierto, respectivamente.

La acción efectiva puede escribirse como

$$S_{tot} = S_{geométrica} + S_{materia} + S_{\Lambda} + S_{\varphi} = \int dx^4 \sqrt{-g} \left[ R + 16\pi G\rho - 2\Lambda + \frac{1}{2} g^{\mu\nu} \partial_{\mu} \varphi \partial_{\nu} \varphi + V(\varphi) \right], \quad (2)$$

donde  $\rho$  es la densidad de energía de la contribución de la materia ordinaria considerando un fluido perfecto barotrópico  $p = \gamma\rho$ ,  $\partial_{\mu} = \frac{\partial}{\partial x^{\mu}}$ , con  $x^{\mu} = (t, r, \theta, \phi)$ ,  $\varphi$ , es el campo escalar dotado en el término de potencial escalar  $V(\varphi)$ ,  $G$  la constante gravitacional de Newton, y  $\Lambda$  es el término cosmológico (constante cosmológica).

De esta manera, el lagrangiano para el modelo cosmológico FRW se lee

$$L = e^{3\alpha} \left[ 6 \frac{\dot{\alpha}^2}{N} - \frac{\dot{\varphi}^2}{2N} + N(V - 2\Lambda - 6\kappa e^{-2\alpha} + 16\pi G M_{\gamma} e^{-3(\gamma+1)\alpha}) \right] \quad (3)$$

donde se usó la forma de la densidad de energía como solución de la ecuación  $T^{\mu\nu}_{;\nu}$ , es decir

$$\rho = M_{\gamma} e^{-3(\gamma+1)\alpha} \quad (4)$$

con  $M_{\gamma}$  una constante de integración.

En este punto se considera un universo plano ( $\kappa=0$ ) y el valor particular para el parámetro  $\gamma = -1$ , el cual es conocido como el fenómeno de inflación clásica (Socorro y Medina, 2000),

$$L = e^{3\alpha} \left[ 6 \frac{\dot{\alpha}^2}{N} - \frac{\dot{\varphi}^2}{2N} + N(V - 2\Lambda + 16\pi GM_{-1}) \right] \quad (5)$$

Los momentos son dados por

$$\Pi_\alpha \equiv \frac{\partial L}{\partial \dot{\alpha}} = 12e^{3\alpha} \frac{\dot{\alpha}}{N}, \quad \dot{\alpha} = \frac{N}{12} e^{-3\alpha} \Pi_\alpha \quad (6)$$

$$\Pi_\varphi \equiv \frac{\partial L}{\partial \dot{\varphi}} = -e^{3\alpha} \frac{\dot{\varphi}}{N}, \quad \dot{\varphi} = -Ne^{-3\alpha} \Pi_\varphi \quad (7)$$

El lagrangiano canónico viene a ser

$$L_{canónico} = \Pi_\alpha \dot{\alpha} + \Pi_\varphi \dot{\varphi} - NH, \quad (8)$$

donde  $H$  es la función densidad hamiltoniana clásica, obteniendo

$$H \equiv \frac{1}{24} e^{-3\alpha} \left[ \Pi_\alpha^2 - 12\Pi_\varphi^2 - 24e^{6\alpha} (V(\varphi) - 2\Lambda + 16\pi GM_{-1}) \right]. \quad (9)$$

El procedimiento de cuantización es realizado en la forma usual, considerando a los momentos como operadores. El hamiltoniano cuántico, conocido como la ecuación de Wheeler-DeWitt, es obtenido bajo la siguiente identificación

$$H \rightarrow \hat{H}\Psi = 0, \quad \Pi_{q^\mu} \equiv -i \frac{\partial}{\partial q^\mu}, \quad (10)$$

con  $q^\mu = (\alpha, \varphi)$ ,

$$\hat{H}\Psi = e^{-3\alpha} \left[ -\frac{\partial^2}{\partial \alpha^2} + 12 \frac{\partial^2}{\partial \varphi^2} - 24e^{6\alpha} V(\varphi, \lambda_{efectiva}) \right] \Psi = 0, \quad (11)$$

donde se renombra  $V(\phi, \lambda_{efectiva}) = V(\phi) - \lambda_{efectiva}$ , y  $\lambda_{efectiva} = 2\Lambda - 16\pi GM_{-1}$  la consideramos como una constante cosmológica efectiva. Esta ecuación puede ser escrita en forma compacta, usando el ordenamiento semi-general propuesto por Hartle-Hawking (Hartle y Hawking, 1983), donde  $Q$  es una constante real que nos mide la ambigüedad en el ordenamiento de factores entre la variable  $e^{3\alpha}$  y su momento asociado a la coordenada  $\alpha$ ,

$$\square \Psi + Q \frac{\partial \Psi}{\partial \alpha} - 24e^{6\alpha} V(\varphi, \lambda_{efectiva}) \Psi = 0, \quad (12)$$

donde  $\square \equiv -\partial_\alpha^2 + \partial_\phi^2$  es el operador d'Alambertiano bidimensional y  $\phi = \frac{\varphi}{\sqrt{12}}$ .

### a. Esquema clásico

La solución cuántica a la ecuación de WDW fue obtenida en trabajos previos (Guzmán *et al.*, 2007), la cual retomaremos en la siguiente subsección, en esta parte del trabajo nuestra principal motivación es en el contexto de soluciones para ecuación de Hamilton-Jacobi. La ecuación clásica de Einstein-Hamilton-Jacobi (EHJ) se obtiene al reemplazar el momento  $\Pi_{q^\mu} \rightarrow \frac{\partial S}{\partial q^\mu}$  en la ecuación (10) y de la restricción  $H=0$  que emerge al obtener la ecuación de movimiento vía la ecuación de Lagrange para la variable auxiliar  $N$ ; la función  $S$  es conocida como la función superpotencial, además se toma  $Q=0$ . En esta forma la ecuación (12) es transformada en

$$(\nabla S)^2 - U = 0, \tag{13}$$

donde

$$(\nabla)^2 = -\left(\frac{\partial}{\partial \alpha}\right)^2 + \left(\frac{\partial}{\partial \phi}\right)^2, \quad G^{\mu\nu} = \text{diag}(-1, 1), \tag{14}$$

con  $U = 24e^{6\alpha}V(\phi, \lambda_{\text{efectiva}})$ . Se obtiene la ecuación diferencial que satisface el término de potencial  $V(\phi, \lambda_{\text{efectiva}})$ , donde algunas de las soluciones a esta ecuación son potenciales que aparecen en la literatura en conexión con la quintaesencia, materia oscura escalar, pero las estructuras de estos potenciales han sido puestos a mano (Matos y Guzmán, 2000; Matos *et al.*, 2000; Matos y Ureña-López, 2000; Matos y Ureña-López, 2001; Lidsey *et al.*, 2002; Cabral-Rosetti *et al.*, 2002; Matos y Ureña-López, 2002; Alcubierre *et al.*, 2002).

Para resolver la ecuación (13) elegimos como ansatz para la función superpotencial

$$S(q^\mu) = f(\alpha)g(\phi), \tag{15}$$

obteniendo

$$-\frac{1}{f^2} \left(\frac{\partial f}{\partial \alpha}\right)^2 + \frac{1}{g^2} \left(\frac{\partial g}{\partial \phi}\right)^2 = 24e^{6\alpha} \frac{V(\phi, \lambda_{\text{efectiva}})}{f^2 g^2}, \tag{16}$$

Por simplicidad, asumimos que  $f(\alpha) = \sqrt{24}e^{3\alpha/\mu}$ , con  $\mu$  una constante arbitraria. Así (16) viene a ser una ecuación diferencial para la función  $g(\phi)$  en términos del potencial generador como

$$\frac{dg}{d\phi} - 9g^2 = \mu^2 V(\phi, \lambda_{\text{efectiva}}). \tag{17}$$

Existen muchas soluciones a la ecuación (17), las cuales pueden ser generadas usando la idea de sistemas integrables, de la siguiente manera. Sea  $V(\phi, \lambda_{\text{efectiva}}) = g^2 F(g)$ , donde  $F(g)$  es una función arbitraria de su argumento. De este modo la ecuación (17) puede ser escrita en cuadraturas como

$$\Delta\phi = \int \frac{d \ln g}{\sqrt{9 + \mu^2 F(g)}} \tag{18}$$

La primera ecuación puede ser resuelta para  $g$  como una función de  $\phi$ , y de esto podemos encontrar el correspondiente potencial generador de este modelo. Algunas soluciones para el potencial generador son mostradas en la Tabla I.1.

**Tabla I.1**  
**Algunas soluciones exactas de la ecuación (18)\***

$F(g)$	$g(\phi)$	$V(\phi, \lambda_{\text{efectiva}}) = g^2 F(g)$
0	$e^{\pm 3\Delta\phi}$	0
$V_0 g^{-2}$	$\frac{1}{3} \sqrt{\mu^2 V_0} \sinh [v]$	$V_0$
	$v = \pm \Delta\phi + \text{Ln} \left( \frac{1}{\sqrt{\mu^2 V_0}} \right)$	
$V_0$	$\exp (\pm B\Delta\phi), \quad B = \sqrt{9 + \mu^2 V_0}$	$V_0 \exp (\pm 2B\Delta\phi)$
$V_0 g^{-n}, (n \neq 2)$	$\frac{1}{3} \sqrt{\mu^2 V_0} \sinh [p]$	$V_0 g^{2-n}$
	$p = \pm \frac{3n}{2} \Delta\phi + \text{Ln} \left( \frac{1}{\sqrt{\mu^2 V_0}} \right)$	
$V_0 \ln g$	$e^u, \quad u = \mu^2 V_0 \left( \frac{\Delta\phi}{2} \right)^2 - \frac{9}{\mu^2 V_0}$	$V_0 u e^{2u}$
$V_0 (\ln g)^2$	$e^r, \quad r = \frac{3}{\sqrt{\mu^2 V_0}} \sinh \left( \sqrt{\mu^2 V_0} \Delta\phi \right)$	$V_0 r^2 e^{2r},$

\*Aquí,  $n$  es cualquier número real. También hay una solución para (18) usando  $F(g) = (\ln g)^n$  en términos de la función hipergeométrica; sin embargo, la forma del potencial generador en este caso no se puede dar en forma cerrada.

De la tabla I.1, la solución de la función superpotencial es conocida para los potenciales generadores descritos en ella.

Es bien conocido que el escenario inflacionario se modela con un potencial con la siguiente estructura  $e^{-\lambda\varphi}$  [5, 61, 62, 63], donde  $\lambda < \sqrt{2}$ . En nuestro esquema el potencial general de la inflación corresponde a  $V=V_0e^{-2B\varphi}$ . En la sección supersimétrica usaremos el superpotencial con este potencial inflacionario.

### b. Esquema cuántico

Para resolver (12) se considera el siguiente ansatz, el cual es análogo al usado en el formalismo Bohmiano de la mecánica cuántica estándar (Bohm, 1952)

$$\Psi(\alpha,\phi) = W(\alpha,\phi)e^{-S(\alpha,\phi)}, \quad (19)$$

donde  $S(\alpha,\phi)$  es la misma función superpotencial introducida en el esquema clásico. De este modo, la ecuación (12) puede ser re-escrita por el siguiente conjunto de ecuaciones diferenciales parciales no lineales

$$(\nabla S)^2 - U = 0, \quad (20)$$

$$W \left( \square S + Q \frac{\partial S}{\partial \alpha} \right) + 2\nabla W \cdot \nabla S = 0, \quad (21)$$

$$\square W + Q \frac{\partial W}{\partial \alpha} = 0, \quad (22)$$

donde la primera ecuación es la ecuación de H-J clásica. Los diferentes términos en esta ecuación tienen la siguiente definición

$$\nabla W \cdot \nabla S \equiv - (\partial_\alpha W) (\partial_\alpha S) + (\partial_\phi W) (\partial_\phi S), \quad (\nabla)^2 \equiv - (\partial_\alpha^2) + (\partial_\phi)^2.$$

Cualquier solución exacta que satisfaga el conjunto de ecuaciones (20, 21, 22) será también una solución exacta a la ecuación de WDW original. Siguiendo la referencia (Guzmán *et al.*, 2007), primero se resuelve la (20), se introduce ésta en (21), cuyas soluciones al final deberán de satisfacer la ecuación (22), que juega el papel de ecuación de restricción para las constantes del problema considerado.

La solución de la primera ecuación se traduce exactamente a la solución del esquema clásico mencionado en la subsección anterior, por lo que nos abocaremos ahora a la solución de la función  $W$  que juega el papel de la amplitud de probabilidad de la función de onda.

Para resolver (21) asumimos que

$$W = e^{[z(\alpha)+\omega(\phi)]}, \quad (23)$$

llegando al siguiente conjunto de ecuaciones diferenciales ordinarias para las funciones  $z(\alpha)$  y  $\omega(\phi)$ .

$$2 \frac{dz}{d\alpha} - Q = k, \quad \rightarrow \quad z(\alpha) = \frac{Q+k}{2} \alpha \tag{24}$$

$$\frac{d^2g}{d\phi^2} + 2 \frac{dg}{d\phi} \frac{d\omega}{d\phi} = 3(k+3)g, \quad \rightarrow \quad \omega(\phi) = \frac{3k}{2} \int \frac{d\phi}{\partial_\phi(\ln g)} - 3\mu^2 \int \frac{d[V(\phi, \lambda_{\text{efectiva}})]}{(\partial_\phi g)^2}. \tag{25}$$

De este modo, la forma explícita para la función  $W$  viene a ser

$$W = \exp\left\{\frac{3k}{2} \left[\frac{\alpha}{3} + \int \frac{d\phi}{\partial_\phi(\ln g)}\right] + \frac{Q}{2} \alpha - 3\mu^2 \int \frac{d[V(\phi, \lambda_{\text{efectiva}})]}{(\partial_\phi g)^2}\right\}. \tag{26}$$

La ecuación de constricción (22) puede ser escrita como

$$\partial_\phi^2 \omega + (\partial_\phi \omega)^2 - \frac{k^2 - Q^2}{4} = 0, \tag{27}$$

y

$$\partial_\phi \omega = \frac{3k}{2\partial_\phi(\ln g)} - 3\mu^2 \frac{\partial_\phi[V(\phi, \lambda_{\text{efectiva}})]}{(\partial_\phi g)^2}. \tag{28}$$

Tomando en cuenta la Tabla I.1, presentamos las correspondientes funciones de onda en cada caso en la Tabla I.2.

**Tabla I.2**  
**Función de onda correspondiente a la Tabla I.1**

$g(\phi)$	Función de onda $\Psi$
$\text{Exp}[\pm 3\Delta\phi]$	$\text{Exp}\left[\left(\frac{k+Q}{2}\right) \alpha \pm \frac{k}{2} \Delta\phi\right] e^{-\frac{e^{3\alpha \pm 3\Delta\phi}}{\mu}}$
$\frac{1}{3} \sqrt{\mu^2 V_0} \sinh [v]$	$\cosh^{\frac{-k}{6}}(v) \text{Exp}\left[\left(\frac{k+Q}{2}\right) \alpha\right] e^{-\frac{e^{3\alpha+g(\phi)}}{\mu}}$
$e^{B\Delta\phi}$	$\text{Exp}\left[\left(\frac{k+Q}{2}\right) \alpha + \frac{3}{B} \left(\frac{k}{2} - 2\mu^2 V_0\right) \Delta\phi\right] e^{-\frac{e^{3\alpha+B\Delta\phi}}{\mu}}$
$\frac{1}{3} \sqrt{\mu^2 V_0} \sinh [p]$	$\text{Exp}\left[\left(\frac{k+Q}{2}\right) \alpha + \omega(\phi)\right] e^{-\frac{e^{3\alpha+g(\phi)}}{\mu}},$
	$\omega(\phi) = -\frac{2k}{3n^2} \text{Ln}\left[\text{Cosh}(p)\right]$
	$+ \frac{12(2-n)}{n^3} \left(\frac{1}{3} \sqrt{\mu^2 V_0}\right)^{2-n} {}_2F_1\left[\frac{n}{2}, \frac{n}{2}, 1 + \frac{n}{2}; \text{Sech}^2 p\right] \text{Sech}^n p$

Continúa...

$e^{u(\phi)}$	$\text{Exp} \left[ \left( \frac{k+Q}{2} \right) \alpha + \frac{3}{\mu^2 V_0} \left( k - 2 + \frac{18}{\mu^2 V_0} \right) \text{Ln} \Delta \phi - \frac{3}{2} \Delta \phi^2 \right] e^{-\frac{e^{3\alpha+u(\phi)}}{\mu}}$
$e^{v(\phi)}$	$\text{Exp} \left[ \left( \frac{k+Q}{2} \right) \alpha + \omega(\phi) \right] e^{-\frac{e^{3\alpha+v(\phi)}}{\mu}},$
	$\omega(\phi) = \frac{k}{2} \text{Arctan} \left( \text{Tanh} \left[ \frac{\sqrt{\mu^2 V_0}}{2} \Delta \phi \right] \right) - 6 \text{Cosh} \left[ \sqrt{\mu^2} \Delta \phi \right]$
	$-27 \left[ -\Delta \phi + \frac{1}{2\sqrt{\mu^2 V_0}} \text{Sinh} \left( 2\sqrt{\mu^2 V_0} \Delta \phi \right) \right]$

De esta forma, usando el formalismo cuántico en el sector inflacionario, encontramos que el potencial escalar puede tener un comportamiento exponencial, sin embargo la constante de acoplamiento es indeterminada. La cuestión que se presenta es: ¿cómo fijar esta constante de acoplamiento? La respuesta podría darse en el contexto de la cosmología cuántica supersimétrica usando operadores diferenciales como realización de las variables de Grassmann, lo cual presentaremos en la siguiente sección.

## II. FORMALISMO DE WITTEN

Existen diferentes formalismos para estudiar la cosmología cuántica supersimétrica, los cuales mencionamos brevemente en secciones anteriores; en este trabajo se usará el formalismo de Witten del cual ahora daremos una pequeña revisión. La idea de Witten presentada en la referencia (Witten, 1981) es encontrar dos operadores  $Q, \bar{Q}$  llamados operadores de supercargas que producen un superhamiltoniano  $H_{SS}$ , y que satisface una álgebra cerrada con el superhamiltoniano

$$\{Q, \bar{Q}\} = H_{SS}, \quad [H_{SS}, Q] = 0, \quad [H_{SS}, \bar{Q}] = 0. \quad (29)$$

donde el superhamiltoniano  $H_{SS}$  tiene la siguiente forma

$$H_{SS} := H_0 + \frac{\partial^2 S(x,y)}{\partial q^\nu \partial q^\mu} [\psi^\nu, \psi^\mu], \quad (30)$$

aquí  $H_0$  es el hamiltoniano bosónico y  $S$  es conocido como la función superpotencial que está relacionado con el potencial que aparece en el hamiltoniano bosónico. Nosotros estamos interesados en hamiltonianos de la siguiente forma

$$H_0\Psi = [f(x)\partial_x^2 + g(x)\partial_x + h(y)\partial_y^2 + j(y)\partial_y + V(x,y)]\Psi(x,y) = 0, \quad (31)$$

donde las funciones  $f, g, h$  y  $j$  dependen de las variables indicadas en su argumento.

Nuestro objetivo en este trabajo es generalizar y aplicar esta idea. Comenzamos dando el siguiente superhamiltoniano

$$H_{super} := H_0 + F_r(q^\mu) \frac{\partial^2 S(x,y)}{\partial q^\nu \partial q^\mu} [\bar{\Psi}^\nu, \Psi^\mu], \quad (32)$$

donde el hamiltoniano bosónico  $H_0$  es del tipo dado por la ecuación (31) y las funciones  $F_r(q^\mu)$ ,  $r = 1, 2$  están relacionadas a las funciones  $f, g, h$  y  $j$ , en completo acuerdo con el hamiltoniano  $H$ .

En principio, la función de onda del universo conduce a la probabilidad que las superficies espaciales evolucionen de un estado inicial dado. Sin embargo, aparecen ambigüedades cuando se intenta utilizar dicha interpretación debido a la naturaleza hiperbólica de la ecuación de WDW: no es posible encontrar una corriente conservada con una densidad de probabilidad positiva definida. Una solución a las dificultades mencionadas es extender la cuantización estandar del universo en un esquema supersimétrico.

### III. SOLUCIONES CUÁNTICAS SUPERSIMÉTRICAS PARA UNA ETAPA INFLACIONARIA

Siguiendo la técnica de la referencia (Socorro, 2003), damos las supercargas

$$Q = \Psi^x \left( A(x)\partial_x + \frac{\partial S(x,y)}{\partial x} \right) + \Psi^y \left( B(y)\partial_y + \frac{\partial S(x,y)}{\partial y} \right), \quad (33)$$

$$\bar{Q} = \bar{\Psi}^x \left( A(x)\partial_x - \frac{\partial S(x,y)}{\partial x} \right) + \bar{\Psi}^y \left( B(y)\partial_y - \frac{\partial S(x,y)}{\partial y} \right), \quad (34)$$

donde las funciones  $A(x), B(y)$  son funciones asociadas con el problema en consideración y  $S$ , como se mencionó, es la función superpotencial (15). La idea es obtener relaciones entre las diferentes funciones que aparecen en las ecuaciones (33, 34) y (31).

Se satisface la siguiente álgebra para las variables (Faddeev y Slavnov, 1991)  $\Psi^\mu$  y  $\bar{\Psi}^\nu$ , ( $\mu, \nu = x, y$ )

$$\left\{ \Psi^\mu, \bar{\Psi}^\nu \right\} = \eta^{\mu\nu}, \quad \left\{ \Psi^\mu, \Psi^\nu \right\} = 0, \quad \left\{ \bar{\Psi}^\mu, \bar{\Psi}^\nu \right\} = 0. \quad (35)$$

Usando la representación  $\psi^\mu = \eta^{\mu\nu} \partial / \partial \theta^\nu$  y  $\bar{\psi}^{\nu} = \theta^\nu$ , uno encuentra que el hamiltoniano en el superespacio es escrito en la forma

$$\begin{aligned}
 H_{super} \Psi &= \{Q, \bar{Q}\} \Psi = (Q\bar{Q} + \bar{Q}Q)\Psi \\
 &= (-A^2(x)\partial_x^2 - A(x)D_x A(x)\partial_x + B^2(y)\partial_y^2 + B(y)D_y B(y)\partial_y \\
 &\quad + \left(\frac{\partial S(x, y)}{\partial x}\right)^2 - \left(\frac{\partial S(x, y)}{\partial y}\right)^2 + A(x)\frac{\partial^2 S(x, y)}{\partial x^2} [\bar{\psi}^x, \psi^x] \\
 &\quad + B(y)\frac{\partial^2 S(x, y)}{\partial y^2} [\bar{\psi}^y, \psi^y] + \frac{1}{2}(A+B)\frac{\partial^2 S}{\partial x \partial y} ([\bar{\psi}^y, \psi^x] + [\bar{\psi}^x, \psi^y])\Psi, \quad (36)
 \end{aligned}$$

donde  $D_x = \frac{\partial}{\partial x}$ ,  $D_y = \frac{\partial}{\partial y}$ . Este hamiltoniano es similar en estructura al descrito por la ecuación (32), por tanto, se obtienen las siguientes relaciones entre las diversas funciones en la ecuación (31) y las supercargas (33, 34)

$$f(x) = -A(x)^2, g(x) = -A(x)D_x A(x), h(y) = B(y)^2, j(y) = B(y)D_y B(y) \quad (37)$$

$$V(x, y) = \left(\frac{\partial S(x, y)}{\partial x}\right)^2 - \left(\frac{\partial S(x, y)}{\partial y}\right)^2 \quad (38)$$

$$F_r = (A(x), B(y)), r = 1, 2. \quad (39)$$

donde (38) corresponde a la ecuación EHJ (13).

En este punto usamos el hamiltoniano de nuestro problema

$$\left(-\frac{\partial^2}{\partial \alpha^2} + \frac{\partial^2}{\partial \phi^2}\right) \Psi - 24e^{6\alpha} V(\phi, \lambda_{efectiva}) \Psi = 0, \quad (40)$$

obteniendo las siguientes relaciones entre las funciones  $x = \alpha$ ,  $y = \phi$ ,  $A = B = 1$ ,  $V(x, y) = -24e^{6\alpha} V(\phi, \lambda_{efectiva})$ .

De esta manera, el superhamiltoniano  $H_{super}$  y las supercargas para el hamiltoniano bosónico en consideración son

$$H_{super} := H + \frac{\partial^2 S(\alpha, \phi)}{\partial \alpha^2} [\bar{\psi}^\alpha, \psi^\alpha] + \frac{\partial^2 S(\alpha, \phi)}{\partial \phi^2} [\bar{\psi}^\phi, \psi^\phi] + \frac{\partial^2 S(\alpha, \phi)}{\partial \alpha \partial \phi} \left( [\bar{\psi}^\phi, \psi^\alpha] + [\bar{\psi}^\alpha, \psi^\phi] \right), \quad (41)$$

$$Q = -(\partial_\alpha + D_\alpha S) \frac{\partial}{\partial \theta^0} + (\partial_\phi + D_\phi S) \frac{\partial}{\partial \theta^1}, \quad (42)$$

$$\bar{Q} = \theta^0 (\partial_\alpha - D_\alpha S) + \theta^1 (\partial_\phi - D_\phi S), \quad (43)$$

También, en este esquema, cualquier estado físico obedece a las siguientes constricciones cuánticas

$$\bar{Q}\Psi = 0, \quad (44)$$

$$Q\Psi = 0, \quad (45)$$

esto simplifica el problema de encontrar el estado base supersimétrico debido a que la energía es conocida *a priori* y también porque la factorización de  $H_{ss} |\Psi\rangle = 0$  en las ecuaciones (44, 45), a menudo nos provee de una ecuación de primer orden para el estado base de la función de onda. La simplicidad de esta factorización está relacionada con la solubilidad de ciertos hamiltonianos bosónicos. En este trabajo encontramos que las soluciones están relacionadas con el sector vacío y lleno en el espacio fermiónico de Fock a energía cero

$$|\Psi_\pm\rangle = e^{\pm S} |\pm\rangle, \quad (46)$$

donde  $S$  es el mismo superpotencial mencionado líneas arriba, además  $\Psi_+$  y  $\Psi_-$  son las correspondientes componentes para el sector independiente y lleno de número de Grassmann en la función de onda. También observamos la tendencia del vacío supersimétrico a permanecer cerca de su límite semiclásico como se mencionó anteriormente.

En este enfoque la función de onda tiene la siguiente descomposición en la representación de variables de Grassmann

$$\Psi = A_+ + C_0\theta^0 + C_1\theta^1 + A_-\theta^0\theta^1, \quad (47)$$

donde las componentes  $A_+$  y  $A_-$  son la contribución bosónica, y  $C_0, C_1$  son la contribución fermiónica a la función de onda.

Cuando usamos la ecuación (44), obtenemos el siguiente conjunto de ecuaciones diferenciales parciales

$$\theta^0 : \partial_\alpha A_+ - D_\alpha S A_+ = 0, \quad (48)$$

$$\theta^1 : \partial_\phi A_+ - D_\phi S A_+ = 0, \quad (49)$$

$$\theta^0 \theta^1 : \partial_\alpha C_1 - D_\alpha S C_1 - \partial_\phi C_0 + D_\phi S C_0 = 0. \quad (50)$$

Introduciendo la solución para  $S$  (15), así la solución para la función  $A_+$  es (resolviendo (48-49))

$$A_+ = K_+ e^s \quad (51)$$

donde  $K_+$  es una constante de integración.

Ahora, usando la ecuación (45) tenemos otro conjunto de ecuaciones diferenciales parciales a resolver

$$\theta^0 : \partial_\phi A_- + D_\phi S A_- = 0, \quad (52)$$

$$\theta^1 : \partial_\alpha A_- + D_\alpha S A_- = 0, \quad (53)$$

$$\text{indep: } -\partial_\alpha C_0 - D_\alpha S C_0 + \partial_\phi C_1 + D_\phi S C_1 = 0, \quad (54)$$

implicando la siguiente estructura para la función  $A_-$

$$A_- = K_- e^{-s} \quad (55)$$

Para las funciones  $C_0, C_1$  se propone el siguiente ansatz

$$C_0 = \frac{\partial f_+}{\partial \alpha} e^s, \quad (56)$$

$$C_1 = \frac{\partial f_+}{\partial \phi} e^s, \quad (57)$$

donde  $f_+ = f_+(\alpha, \phi)$  es una función bosónica; con este ansatz la ecuación (50) es resuelta y la ecuación (54) puede ser escrita como

$$\square f_+ + 2 \nabla f_+ \cdot \nabla S = 0, \quad (58)$$

con la métrica  $\eta^{-\nu\nu} = \text{diag}(-1, 1)$ , podemos escribirla

$$\eta^{\mu\nu} (\partial_\mu \partial_\nu f_+ + \partial_\nu f_+ \partial_\mu S + \partial_\nu f_+ \partial_\mu S) = \eta^{\mu\nu} (\partial_\mu \partial_\nu f_+ + 2\partial_\nu f_+ \partial_\mu S) = 0, \quad (59)$$

donde una posible solución es  $f_+ = h(\alpha \pm \phi)$ , y  $h$  es cualquier función dependiente de su argumento, dando la siguiente restricción para la función superpotencial

$$\frac{\partial S}{\partial \alpha} = \pm \frac{\partial S}{\partial \phi} \quad (60)$$

y considerando la estructura del superpotencial (15) encontramos la condición de integrabilidad sobre la función  $g(\phi)$ ,

$$g(\phi) = g_0 e^{\pm 3\Delta\phi}. \quad (61)$$

Y tomando en cuenta las tablas I.1 o I.2 se encuentra la siguiente restricción para el parámetro  $B$  de nuestro modelo cuando esta última ecuación es satisfecha (60), siendo el valor extremo que puede tomar

$$B = \pm 3;$$

de esta forma, únicamente el potencial exponencial sobrevive de la tabla I.1, bajo la condición de integrabilidad, y usando la relación entre el campo escalar  $\phi = \frac{\varphi}{\sqrt{12}}$ , y de que el potencial  $V(\phi) = V_0 g^2(\phi)$  obtenemos que el potencial escalar viene dado por  $V(\phi) = V_0 e^{-\sqrt{3}\Delta\phi}$ . De esta forma, la mecánica cuántica supersimétrica fija el valor del parámetro  $B$ , y si queremos mantener que el potencial  $V_0$  sea positivo, elegimos  $\mu = i$  sin perder generalidad, lo cual haría que el superpotencial  $S = -i\Phi(\alpha, \phi) = -if(\alpha)g(\phi)$ , y todo es válido con este cambio.

La principal diferencia aparece en la densidad de probabilidad no normalizable  $|\Psi|^2$  dada en la siguiente sección, donde los términos cuadrados de las contribuciones de la función de onda son constantes.

#### IV. DENSIDAD DE PROBABILIDAD NO NORMALIZADA $|\Psi|^2$ EN LAS VARIABLES DE GRASSMANN

En esta sección seguimos la referencia (Obregón *et al.*, 1993) para obtener la densidad de probabilidad  $|\Psi|^2$  para la función de onda dada por la estructura (47) en el espacio de Grassmann. Si  $\Psi_1$  y  $\Psi_2$  son funciones que dependen de los números

de Grassmann, se usará la siguiente notación dada en la referencia (Faddeev y Slavnov, 1991) donde se define el producto escalar en este espacio

$$(\Psi_1 | \Psi_2) = \int (\Psi_1(\theta^*)) * \Psi_2(\theta^*) e^{-\sum_i \theta_i^* \theta_i} \prod_i d\theta_i^* d\theta_i, \quad (62)$$

donde la operación  $*$  es definida como  $(C\theta_1 \dots \theta_n)^* = \theta_n^* \dots \theta_1^* C^*$ , con la usual álgebra para los números de Grassmann  $\theta_i \theta_j = -\theta_j \theta_i$ , y la integral en este espacio obedece

$$\int \theta_1^* \theta_1^* \dots \theta_n^* \theta_n^* d\theta_n^* d\theta_n^* \dots d\theta_1^* d\theta_1^* = 1 \quad (63)$$

$$\int d\theta_i^* = \int d\theta_i = 0 \quad (64)$$

En nuestro caso, tenemos que  $\Psi_1 = \Psi_2 = \Psi$ . De este modo, cuando integramos sobre los números de Grassmann y usando las relaciones (63) y (64), obtenemos

$$|\Psi|^2 = A_+^* A_+ + C_0^* C_0 + C_1^* C_1 + A_-^* A_- \quad (65)$$

Usando las expresiones para las diversas contribuciones de la función de onda  $A_+$ ,  $C_0$ ,  $C_1$  y  $A_-$  dadas en (51), (56), (57) y (55) respectivamente, llegamos a la siguiente expresión para la densidad de probabilidad

$$|\Psi|^2 = K_-^2 e^{-2S} + \left[ K_+^2 + \left( \frac{\partial h}{\partial \alpha} \right)^2 + \left( \frac{\partial h}{\partial \phi} \right)^2 \right] e^{2S}. \quad (66)$$

Así, la densidad de probabilidad es escrita en términos de las componentes para la parte independiente y saturada de los números de Grassmann de la función de onda. En este sentido, observamos la tendencia del vacío supersimétrico a permanecer cercano a su límite semiclassical, tal como ocurre aquí y en otros trabajos previos (Socorro, 2003; Socorro *et al.*, 2012), donde las soluciones exactas son también las de más bajo orden en la aproximación WKB (Socorro, 2002; Kiefer *et al.*, 2005), pero no son normalizables. También es bien conocido que la existencia de soluciones normalizables del sistema (44, 45) significa que la supersimetría es rota cuánticamente. Por otro lado, en la referencia Graham y Luckock (1994) hacen mención de que el sector  $A_{\pm}$  son también distinguidas por la existencia de un mapa tipo Nicolai y están relacionadas a una interpretación estadística de la función de onda; es decir, el mapa tipo Nicolai existe en el sector de los sectores

independientes y completos en la representación de Grassmann para la función de onda, pero no en otros sectores.

Existe un enfoque similar a éste, propuesto por Lidsey y Moniz (2000), quienes obtienen soluciones cuánticas para los modelos cosmológicos Bianchi tipo I Kantowski-Sachs, en el contexto de la teoría de gravedad de Brans-Dicke, donde ellos explotan la idea de la simetría de dualidad en el factor de escala.

## CONCLUSIONES

Empleando el formalismo de la mecánica cuántica en un escenario inflacionario en el sector de la materia ordinaria, lo cual también se refleja en el comportamiento del potencial escalar, se encuentra que existe una familia de potenciales que puede gobernar este escenario; pero en particular existe uno que tiene un comportamiento exponencial como un buen candidato, el cual se refuerza cuando usamos el esquema supersimétrico. Sin embargo, en el esquema de la cosmología cuántica el parámetro de acoplamiento  $\lambda$  quedó indeterminado. En este punto, se hizo la pregunta: ¿cómo podemos fijar el valor de este parámetro? Invocamos al esquema supersimétrico para tal fin, usando operadores en el superespacio  $(q^\mu, \theta^\mu)$ , donde este parámetro es determinado bajo una condición de integrabilidad de las funciones  $g(\phi) = g_0 e^{\pm 3\Delta\phi}$ , y tomando en cuenta las tablas I.1 o II.2;  $B = \pm 3$ . De esta forma, el principal objetivo de este trabajo fue fijar el valor de este parámetro para el potencial escalar de esta etapa de evolución del universo al usar el enfoque supersimétrico, cuando el formalismo cuántico únicamente nos da la estructura general para este tipo de potencial. También a nivel cuántico encontramos que las posibles soluciones vienen a ser las contribuciones de los sectores independiente (+) y completo (-) de la función de onda en números de Grassmann del esquema supersimétrico. Es importante mencionar que la supersimetría es muy restrictiva debido a que existen más ecuaciones de restricción las cuales se aplican a la función de onda, con lo cual podemos decir que el valor obtenido para el parámetro  $\lambda$  tiene un ingrediente de fineza. Además, observamos la tendencia de que el vacío supersimétrico permanezca cerca de su límite semiclásico, esto por la estructura de las soluciones encontradas, las cuales también corresponden a soluciones en el orden más bajo de la aproximación WKB.

## REFERENCIAS

Alam, U., Sahni, V. y Starobinsky, A. A. (2004). "The Case for Dynamical Dark Energy Revisited", *J. Cosmol. Astropart. Phys.*, 008.

- Alcubierre, M., Siddhartha Guzmán, F., Matos, T., Núñez, D., Ureña-López, L. A. y Wiederhold, P. (2002). *Class. Quant. Grav.*, 5017-5024. gr-qc/0110102.
- Arbey, A., Lesgourgues, J. y Salati, P. (2001). “Quintessential Haloes around Galaxies”, *Phys. Rev. D*, 123528.
- Bene, J. y Graham, R. (1994). *Phys. Rev. D*, 799.
- Bohm, D. (1952). *Phys. Rev.*, 166.
- Cabral-Rosetti, L. G., Matos, T., Núñez, D. y Sussman, R. (2002). *Class. Quant. Grav.*, 3603-3615. gr-qc/0112044.
- Capovilla, R. y Guven, J. (1994). *Class. Quantum Grav.*, 11, 1961.
- Capovilla, R. y Obregón, O. (1994). *Phys. Rev. D*, 49, 6562.
- Cheng, A. D. Y., D’Eath, P. D. y Moniz, P. R. L. V. (1995). *Class Quantum Grav.*, 1343.
- Cheng, A. D. Y. y Moniz, P. R. L. V. (1995). *Int. J. Mod. Phys. D*, 189.
- Cheng, A. D. Y., D’Eath, P. D. y Moniz, P. R. L. V. (1994). *Phys. Rev. D*, 5246.
- Chimento, L. P. y Jakubi, A. S. (1996). *Int. J. Mod. Phys. D*, 71.
- Cooper, F., Khare, A. y Sukhatme, U. (2001). *Supersymmetry in Quantum Mechanics*, World Scientific.
- (1995). *Phys. Rep.*, 267.
- Copeland, E. J., Sami, M. y Tsujikawa, S. (2006). *Dynamics of Dark Energy*, [arXiv:hep-th/0603057].
- Copeland, E. J., Liddle, A. R. y Wands, D. (1998). *Phys. Rev. D*, 4686 (and references therein).
- D’Eath, P. D. (1996). *Supersymmetric Quantum Cosmology*. Cambridge University Press, Cambridge, Inglaterra.
- (1994). *Phys. Lett. B*, 12.
- y Hughes, D. I. (1993). *Phys. Lett. B*, 498.
- D’Eath, P. D., Hawking, S. W. y Obregón, O. (1993). *Phys. Lett. B*, 44.
- D’Oleire, M. A. (2010). “Estudio de modelos cosmológicos acoplados a campo escalar desde el punto de vista de la mecánica cuántica supersimétrica”, tesis licenciatura de Física, Universidad Autónoma del Estado de México, abril 2010. Asesores: J. Socorro y José Antonio Aguilar.
- Escamilla-Rivera, C. (2009). “Lo que la supersimetría local puede hacer por la cosmología”. Master thesis, DCEI, Universidad de Guanajuato.
- Faddeev, L. D. y Slavnov, A. A. (1991). *Gauge Fields: An Introduction to Quantum Theory* (Addison-Wesley, Reading, MA), sec. 2.5.
- Fang, L. Z. y Ruffini, R. (1987). *Quantum Cosmology*. Advances Series in Astrophysics and Cosmology, vol. 3. World Scientific.
- Fernández, D. J. (1984). *Lett. Math. Phys.*, 337.

- García-Compeán, H., García-Jiménez, G., Obregón, O. y Ramírez, C. (2007). *Phys. Rev. D*, 063517.
- Gibbons, G. W. y Grishchuk, L. P. (1989). *Nucl. Phys. B*, 736.
- Gorini, V., Kamenshchik, A. Y., Moschella, U. y Pasquier, V. (2004). *Phys. Rev. D*, 69, 123512.
- Graham, R. (1991). *Phys. Rev. Lett.*, 1381.
- y Luckock, H. (1994). *Phys. Rev. D*, 2786.
- Guzmán, W., Sabido, M., Socorro, J. y Ureña-López, L. A. (2007). *Int. J. Mod. Phys. D*, (4), 641-653. gr-qc/0506041.
- Hartle, J. y Hawking, S. W. (1983). *Phys. Rev. D*, 2960.
- Hawking, S. W. (1984). *Nucl. Phys. B*, 257.
- Kamenshchik, A. Y., Moschella, U. y Pasquier, V. (2004). *Phys. Rev. D*, 123512.
- Kiefer, C., Lück, T. y Moniz, P. V. (2005). *Phys. Rev. D*, 045006.
- Lidsey, J., Matos, T. y Ureña-López, L. A. (2002). *Phys. Rev. D*, 023514. astro-ph/0111292.
- Lidsey, J. y Moniz, P. V. (2000). *Class and Quantum Grav.*, 4823.
- Lidsey, J. (1995). *Phys. Rev. D*, R5407.
- Lyth, D. y Liddle, A. (2000). *Cosmological Inflation and Large Scale Structure*, Cambridge University Press.
- Macías, A. (1999). “The Ideas behind the Different Approaches to Quantum Cosmology”, *Gen. Rel. and Grav.*, 653-671.
- Macías, A., Mielke, E. W. y Socorro, J. (1998). *Phys. Rev. D*, 1027.
- Macías, A., Obregón, O. y Socorro, J. (1993). *Int. J. Mod. Phys. A*, 4291.
- Macías, A., Obregón, O. y Ryan, M. P. (1987). *Class Quantum Grav.*, 1477.
- Matos, T. y Ureña-López, L. A. (2004). “On the Nature of Dark Matter”, *Int. J. Mod. Phys. D*, 2287.
- (2002). *Phys. Lett. B*, 246-250. astro-ph/0010226.
- (2001). *Phys. Rev. D*, 063506. astro-ph/0006024.
- (2000). *Class. Quant. Grav.*, L75-L81. astro-ph/0004332.
- Matos, T. y Siddhartha Guzmán, F. (2000). *Class Quant. Grav.*, L9-L16. gr-qc/9810028.
- Matos, T., Siddhartha Guzmán, F. y Ureña-López, L. A. (2000). *Class Quant. Grav.*, 1707-1712. astro-ph/9908152.
- Mielnik, B. J. (1984). *Math. Phys.*, 3387.
- Moniz, P. V. (2010). “Quantum Cosmology –the Supersymmetric Perspective–”. Vol. 1 y 2. *Lecture Notes in Physics & 804*, (Springer, Berlín).
- (1996a). *Int. J. Mod. Phys. A*, 6.
- (1996b). *Int. J. Mod. Phys. A*, 4321.
- Nieto, M. M. (2004). *Phys. Lett. B*, 208.

- Obregón, O. y Ramírez, C. (1998). *Phys. Rev. D*, 1015.
- Obregón, O., Rosales, J. y Tkach, V. I. (1996). *Phys. Rev. D*, R1750.
- Obregón, O. y Rosales, J. (1996). *Int. J. Theor. Phys.*, 1381.
- Obregón, O., Socorro, J. y Benítez, J. (1993). *Phys. Rev. D*, 4471.
- Obregón, O. y Socorro, J. (1996). *Int. J. Theor. Phys.*, 35, 1381.
- Peebles, P. J. E. y Ratra, B. (2003). *Rev. Mod. Phys.*, 559.
- Ryan, M. (1972). *Hamiltonian Cosmology*. Springer, Berlín.
- Sahni, V. (2004). *Lect. Notes Phys.*, 141.
- y Wang, L. M. (2000). “A New Cosmological Model of Quintessence and Dark Matter”, *Phys. Rev. D*, 103517.
- Sen, A. (2003). *Mod. Phys. Lett. A*, 4869.
- (2002). *Mod. Phys. Lett. A*, 1797.
- Socorro, J., Martínez-Prieto, C. y Morales, E. (2012). *Classical, Quantum and SUSY Quantum Cosmological Solutions Inspired in String Theory*. En revisión.
- Socorro, J. y D’Oleire, M. (2010). “Inflation from Supersymmetric Quantum Cosmology”, *Phys. Rev. D*, 044008.
- Socorro, J. (2003). *Int. J. Theor. Phys.*, 2087.
- , Reyes, M. A. y Gelbert, F. A. (2003). *Phys. Lett. A*, 338.
- Socorro, J. (2002). *Rev. Fis.*, 112. Méx.
- y Medina, E. R. (2000). *Phys. Rev. D*, 087702-1.
- Socorro, J., Obregón, O. y Macías, A. (1992). *Phys. Rev. D*, 2026.
- Tegmark, M. (2005). “What Does Inflation Really Predict?”, *J. Cosmol. Astroport. Phys.*, 001.
- Tkach, V. I., Rosales, J. y Obregón, O. (1997). *Class Quantum Grav.*, 339.
- (1996). *Class Quantum Grav.*, 2349.
- Tsujikawa, S. (2000). *Phys. Rev. D*, 043512, [arXiv:hep-ph/0004088].
- Ureña-López, L. A. y Matos, T. (2000). “A New Cosmological Tracker Solution for Quintessence”, *Phys. Rev. D*, 081302.
- Witten, E. (1981). *Nucl. Phys. B*, 513.



# ESTUDIO DINÁMICO DEL MODELO $\Lambda$ -CDM INHOMOGÉNEO

*Roberto A. Sussman\**

*Germán Izquierdo\*\**

## Resumen

Consideramos modelos de espacio-tiempos esféricamente simétricos inhomogéneos cuya fuente es polvo con constante cosmológica positiva, dados por la métrica Lemaître-Tolman-Bondi (LTB). Estas configuraciones ofrecen una generalización sencilla pero útil del modelo  $\Lambda$ -CDM que contiene materia oscura fría (CDM) y un término  $\Lambda$ , y que parece ajustar bien los datos y observaciones actuales. La dinámica de estos modelos puede describirse completamente por ecuaciones de evolución escalares que toman la forma de un sistema dinámico asociado a un espacio de fase 4-dimensional, cuyos puntos críticos y subespacios invariantes son examinados y clasificados. La evolución en el espacio de las fases de diversas configuraciones se estudia en detalle por medio de dos subespacios bidimensionales: una proyección en el subespacio homogéneo invariante asociado con soluciones  $\Lambda$ -CDM en una métrica FLRW, y una proyección en un subespacio que representa las fluctuaciones definidas adecuadamente, que tienen en cuenta los efectos de la falta de homogeneidad. Se estudian casos que presentan expansión perpetua, casos que presentan fenómenos de retorno (un colapso inicial seguido de una expansión o una expansión seguida de colapso) y de “loitering” (la evolución se detiene en un estado estacionario inestable), así como configuraciones con patrones de cinemática “mixtos”, por ejemplo una configuración que presenta una región determinada del espacio que colapsa mientras que el resto del espacio presenta expansión. En todos los casos, las trayectorias del espacio de fases evolucionan desde un atractor pasado y convergen a un atractor futuro de una manera cualitativamente

\* Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México (ICN-UNAM), A. P. 70-543, 04510, México, D.F., México, <sussman@nucleares.unam.mx>.

\*\* Facultad de Ciencias, Universidad Autónoma del Estado de México, Toluca 50000, Instituto Literario 100, Estado de México, Mexico, <german.izquierdos@uamex.mx>.

análoga a la evolución de los modelos homogéneos FLRW. Sin embargo, podemos identificar en ambas proyecciones del espacio de fase varias características cualitativas ausentes en el límite FLRW que pueden ser útiles en la construcción de modelos sencillos de inhomogeneidades astrofísicas y cosmológicas.

**Palabras clave:** Constante cosmológica, universo inhomogéneo, formación de estructura.

## INTRODUCCIÓN

Desde un punto de vista fenomenológico y empírico, los datos observacionales cosmológicos son generalmente ajustados con el modelo  $\Lambda$ -CDM, el cual describe un espacio-tiempo de Friedman-Lemaître-Robertson-Walker (FLRW) con secciones espaciales planas, cuyas fuentes de energía son el polvo (materia oscura fría, CDM) y un campo escalar  $\Lambda$  (energía oscura) (Padmanabhan, 2003; Peebles, 2003). Aunque el universo puede ser prácticamente homogéneo a escalas de longitud mayores a la escala de homogeneidad (distancias del orden de 150-300 Mpc), quedando así justificado el uso de perturbaciones lineales en el estudio dinámico del universo, es claramente no homogéneo a escalas más pequeñas en las cuales han ocurrido formaciones de estructura en las que la CDM ha estado involucrada. Por tanto, el estudio de fuentes de polvo inhomogéneas y un “campo  $\Lambda$ ” diferente de cero es un tema muy relevante. En particular, los espacio-tiempos con este tipo de fuentes que presenten simetría esférica proporcionan generalizaciones inhomogéneas simples, pero no triviales del modelo  $\Lambda$ -CDM.

Las soluciones inhomogéneas esféricamente simétricas de las ecuaciones de Einstein de polvo con  $\Lambda = 0$ , descritas por la métrica de LTB han sido ampliamente utilizadas para construir modelos sencillos de inhomogeneidad cosmológica (Kraśiński, 1998; Kraśiński y Plebański, 2007) para una amplia discusión y revisión). Sin embargo, la literatura sobre modelos de LTB con  $\Lambda > 0$  (a los que llamaremos modelos  $\Lambda$ -LTB) es mucho menos abundante y se ha enfocado fundamentalmente en el estudio de sus singularidades y censura (Omer, 1965; Zecca, 1991).

En este capítulo presentamos un resumen y traducción del inglés de un trabajo que hemos publicado recientemente (Sussman e Izquierdo, 2011) sobre los modelos de  $\Lambda$ -LTB, en el cual utilizamos las técnicas cualitativas y numéricas genéricamente conocidas como “sistemas dinámicos”. Estas técnicas se han aplicado con éxito en una amplia variedad de espacio-tiempos (FLRW, Kantowski-Sachs, los modelos de Bianchi, espacios-tiempos auto-similares y estáticos, etcétera), en los cuales las ecuaciones de evolución para las variables de fase de “expansión normalizada” pueden reducirse a un sistema autónomo de ecuaciones diferenciales

ordinarias (ODEs) que definen un espacio de fases auto-consistente (Wainwright y Lim, 2005; Wainwright y Ellis, 1997; Coley, 2003 y 1999) para una descripción completa sobre este tema. Previamente, en la referencia (Sussman, 2008), se aplicó con éxito un análisis del sistema dinámico de las soluciones de polvo LTB con  $\Lambda = 0$ . Aquí, generalizamos este estudio para el caso  $\Lambda > 0$ .

El capítulo está organizado del siguiente modo. La clase de los espacio-tiempos  $\Lambda$ -LTB y sus objetos covariantes correspondientes son introducidos en la sección I, mientras que en la sección II mostramos que su evolución dinámica puede ser completamente determinada por variables “cuasilocales” (QL) (Sussman, 2009, 2010a, 2010b) que conducen a un conjunto de ecuaciones de evolución (Ellis y Bruni, 1989) y a una parametrización de los valores iniciales conveniente para la solución numérica de los modelos. Estas ecuaciones de evolución son transformadas en la sección III en un sistema dinámico 4-dimensional construido por “variables normalizadas de expansión” (Wainwright y Lim, 2005 y Wainwright y Ellis, 1997; Coley, 2003 y 1999). En la sección IV se enumeran los puntos críticos (atractores pasados y futuros y puntos de silla) de este sistema dinámico y se clasifican todos los subespacios invariantes del espacio de fases: el subconjunto homogéneo FLRW, el subespacio espacialmente plano, así como la solución del vacío de Kottler (Kottler, 1918) (Schwarzschild-De Sitter) y subespacios de polvo “puro” con  $\Lambda = 0$ .

Examinamos en la sección V diferentes patrones de evolución cinemática de modelos  $\Lambda$ -LTB. Las clases topológicas admisibles de las rebanadas de espacio ortogonal a la cuadrivelocidad se resumen en la sección VI. Debido a que el espacio de fases es 4-dimensional, lo descomponemos en la sección VII en dos proyecciones bidimensionales: los subsistemas homogéneo e inhomogéneo. Estas proyecciones contienen la información dinámica completa y, por tanto, son utilizadas en la sección VIII para estudiar en detalle las trayectorias del espacio de fases de las configuraciones representativas asociadas a cada uno de los patrones de cinemática y las topologías de las rebanadas espaciales enumeradas en las secciones V y VI. Para ello, resolvemos numéricamente el sistema dinámico obtenido en la sección III para cada configuración, junto con las ecuaciones de evolución QL derivadas en la sección II (lo cual es necesario cuando las variables espacio de fases divergen en configuraciones de colapso o retorno). En las conclusiones resumimos los resultados y elaboramos una discusión final.

## I. ESPACIO-TIEMPOS $\Lambda$ -LTB

Las soluciones de polvo simétricamente esféricas con constante cosmológica  $\Lambda$  diferente de cero, o (como son generalmente denominadas) modelos Lemai-

tre-Bondi-De Sitter están caracterizadas por una métrica LTB (Kraśiński, 1998; Kraśiński y Plebański, 2007).

$$ds^2 = -c^2 dt^2 + \frac{R'^2 dr^2}{1+E} + R^2[d\theta^2 + \sin^2\theta d\varphi^2] \quad (1)$$

donde  $R = R(ct, r)$ ,  $R' = \partial R/\partial r$ ,  $E = E(r)$ , y el tensor de energía-momento

$$T^{ab} = \rho u^a u^b - \Lambda g^{ab} \quad (2)$$

con  $u^a = \delta_0^a$  y  $\rho(ct, r)$  es la masa-energía en reposo. Las ecuaciones de campo  $G^{ab} = \kappa T^{ab}$  (con  $\kappa_0 = 8\pi G/c^4$ ) para (2)-(1) resultan ser:

$$\kappa_0 \rho c^2 R^2 R' = 2M' \quad (3)$$

$$\dot{R}^2 = \frac{2M}{R} + \frac{\kappa_0}{3} \Lambda R^2 + E \quad (4)$$

donde  $\dot{R} = \partial R/\partial ct$ , y la función  $M = M(r)$ , que aparece como una constante de integración, es la masa-energía en reposo efectiva dada en unidades de longitud.

El campo de cuadrivelocidades geodésico y comóvil describe una foliación de tipo espacio de hipersuperficies,  ${}^3T(t)$ , ortogonal a  $u^a$ , definido en  $t$  constante, y con una métrica inducida  $h_{ab} = u_a u_b + g_{ab} = \delta_a^i \delta_b^j g_{ij}$ , donde  $i, j = r, \theta, \varphi$ . Aparte,  $u^a$ ,  $\rho$  y  $\Lambda$  y los objetos covariantes restantes asociados con los espacio-tiempos  $\Lambda$ -LTB son dos escalares: el escalar de expansión,  $\Theta = \nabla_a u^a$  y el escalar Ricci de hipersuperficies  ${}^3T(t)$ :

$$\Theta = \frac{2\dot{R}}{R} + \frac{\dot{R}'}{R'} \quad {}^3R = -\frac{2(ER)'}{R^2 R'} \quad (5)$$

mas dos tensores de tipo espacio con traza 0, el tensor de torsión,  $\sigma_{ab} = \nabla_{(a} u_{b)}$  -  $(\Theta/3)h_{ab}$ , y el tensor eléctrico de Weyl,  $E_{ab} = u_c u_d C^{abcd}$ , que pueden ser completa y covariantemente determinados en términos de una única función escalar  $\sigma^{ab} = \Sigma \Xi^{ab}$  y  $E^{ab} = E \Xi^{ab}$ , donde  $\Xi^{ab} \equiv h^{ab} - 3\chi^a \chi^b$ , con  $\chi^a = \sqrt{h^{rr}} \delta_r^a$  siendo el vector unitario ortogonal a  $u^a$  y a las órbitas de  $SO(3)$ . Las funciones  $\Sigma$  y  $E$  son:

$$\Sigma = \frac{1}{3} \left[ \frac{\dot{R}}{R} - \frac{\dot{R}'}{R'} \right], \quad E = -\frac{\kappa_0}{6} \left[ \rho c^2 - \Lambda \right] + \frac{M}{R^3} \quad (6)$$

El conjunto  $\{\rho, \Lambda, \Theta, \Sigma, E, {}^3R\}$  brinda una representación completa de escalares covariantes locales para los modelos  $\Lambda$ -LTB.

## II. ESCALARES CUASILOCALES (QL)

Tal y como se hace en (Sussman, 2009 y 2010), definimos para los escalares  $A = \{\rho, \Theta, {}^3R, \Lambda\}$  sus escalares duales cuasilocales (QL)  $A_q = \{\rho_q, \Theta_q, {}^3R_q, \Lambda_q\}$  mediante la integral,<sup>1</sup>

$$A_q = \frac{\int_{x=0}^{x=r} AR^2 R' dx}{\int_{x=0}^{x=r} R^2 R' dx} \quad (7)$$

donde la integral es evaluada en el dominio comóvil esférico  $D = S^2_x \vartheta [r] \subset^3 T(t)$ , siendo  $S^2$  la 2-esfera unitaria parametrizada por  $(\theta, \varphi)$  y  $\vartheta[r] = \{x | 0 \leq x \leq r\}$ , donde  $x=0$  marca un centro de simetría. Es evidente que  $\Lambda_q = \Lambda$ .<sup>1</sup> A partir de (5) y (7), los escalares QL son

$$\frac{\kappa_0}{3} \rho_q = \frac{2M}{R^3}, \quad \frac{\Theta_q}{3} = \frac{\dot{R}}{R}, \quad \frac{{}^3R_q}{6} = -\frac{E}{R^2}, \quad \Lambda_q = \Lambda \quad (8)$$

Los escalares QL y locales pueden relacionarse por medio de las “desviaciones relativas”

$$\delta(A) \equiv \frac{A - A_q}{A_q} = \frac{A' / A}{3R' / R} \quad (9)$$

donde se utiliza la propiedad  $A'q = (3R'/R)(A - A_q)$ , derivada directamente de (7).

Es conveniente para el trabajo numérico usar variables adimensionales, por tanto aquí introducimos:

<sup>1</sup> Estamos utilizando una notación diferente a aquella utilizada en las referencias [15,16], donde las funciones  $A_q$  fueron denotadas por  $A_*$ . Estas mismas funciones fueron introducidas en [13], pero no fueron llamadas funciones QL. En esa referencia la notación usada fue  $l, H, S$ , respectivamente para  $L, \Theta_q/3, -\delta^{(z)}$ , y  $\Lambda^{(m)}, \Lambda^{(k)}$  para  $\delta^{(u)}, \delta^{(h)}$  mientras  $\langle \rho \rangle, \langle {}^3R \rangle$  y  $\langle \Omega \rangle$  fueron utilizadas por  $\rho_q, {}^3R_q$  y  $\hat{\Omega}^{(u)}$ . La notación  $\langle \dots \rangle$  puede ser malinterpretada debido a que (7) no es un promedio excepto cuando sea definido como un funcional (véase Sussman, 2009). Consultar además referencias (Sussman, 2009, 2010a y 2010b; Hayward, 1996) para una discusión de la naturaleza covariante de la integral (7) y su relación con la función masa-energía QL de Mister-Sharp.

<sup>2</sup> El rango radial puede ser modificado para incluir modelos que no admiten un centro de simetría, o cuando los  ${}^3T|\tau|$  están restringidos por singularidades en la curvatura, véanse referencias de Sussman (2009, 2010a y 2010b).

$$2\mu_q = \frac{\kappa_0 \rho_q}{3h_s^2}, \quad z_q \frac{\Theta q}{3h_s}, \quad \kappa_q \equiv \frac{{}^3R_q}{6h_s^2}, \quad \lambda \equiv \frac{\kappa_0 \Lambda}{3h_s^2}, \quad \tau = h_s c(t - t_i) \quad (10)$$

donde  $h_s^{-1}$  es un escalar de longitud arbitraria,  $t = t_i$  marca una hipersuperficie  ${}^3T_i$  inicial fiducial, de manera que ésta coincide con  $\tau = 0$ . Considerando las definiciones dadas en (10) y teniendo en cuenta las referencias (Sussman, 2009 y 2010a) la dinámica de los modelos LTB-dS está totalmente determinada por las siguientes ecuaciones de evolución.<sup>3</sup>

$$\frac{\partial \mu_q}{\partial \tau} = 3\mu_q z_q \quad (11)$$

$$\frac{\partial z_q}{\partial \tau} = -z_q^2 - \mu_q + \lambda \quad (12)$$

$$\frac{\partial \delta^{(\mu)}}{\partial \tau} = -3z_q \delta^{(z)} \left[ 1 + \delta^{(\mu)} \right] \quad (13)$$

$$\frac{\partial \delta^{(z)}}{\partial \tau} = -z_q \delta^{(z)} \left[ 1 + 3\delta^{(z)} \right] + \frac{2\mu_q \left[ \delta^{(z)} - \delta^{(\mu)} \right] - \lambda \delta^{(z)}}{2z_q} \quad (14)$$

junto con la condición

$$z_q^2 = 2\mu_q - \kappa_q + \lambda \quad (15)$$

Una vez que el sistema (11)-(14) se resuelve, los escalares  $\{\rho, \Lambda, \Theta, \Sigma, E, {}^3R\}$  resultan ser:

$$\mu = \mu_q [1 + \delta^{(\mu)}], \quad z = z_q [1 + \delta^{(z)}], \quad \kappa = \kappa_q [1 + \delta^{(\kappa)}], \quad (16)$$

$$\Sigma = -h_s z_q \delta^{(z)}, \quad E = -h_s m_q \delta^{(\mu)} \quad (17)$$

Como la referencia (Sussman, 2009) demuestra, las ecuaciones de evolución del tipo (11)-(14) determinan completamente la dinámica de todos los espacio-tiempos descritos por una métrica LTB en el marco comóvil (incluyendo modelos LTB-dS). En esta descripción de variables  $A_q$  (que satisfacen la dinámica FLRW)

<sup>3</sup> La ecuación (14) está incorrectamente escrita en las referencias (Sussman, 2009 y 2010a).

formalmente define un fondo FLRW, mientras que  $\delta^{(4)}$  puede ser rigurosamente caracterizado como perturbaciones covariantes, invariantes bajo transformaciones de gauge, no lineales en este fondo.

Es útil introducir una formulación de los valores iniciales para los modelos LTB-dS. Considerando una rebanada arbitraria, reescribimos las funciones métricas  $R$  y  $R'$  en (1) como factores escalares adimensionales.

$$L = \frac{R}{R_i} = \frac{R}{R_0 r}, \quad \Gamma \frac{R'/R}{R'_i/R_i} = 1 + \frac{L'/L}{R'_i/R_i} \quad (18)$$

donde el subíndice  $i$  denota de aquí en adelante la evaluación en  $t = t_i$ . Considerando como condiciones iniciales para (11)-(14) las funciones de valor inicial  $\{\mu_{qi}, \kappa_{qi}, \lambda\}$ , los parámetros libres estándar en (1)-(3) se siguen de (8) como  $Mh_s = m_{qi} R_i^3$  y  $Eh_s = \kappa_{qi} R_i^2$ , las ecuaciones (4), (8) y (16) nos conducen a unas ecuaciones de tipo Friedman:

$$z_q^2 L^2 = L_{,\tau}^2 = \frac{Q(L)}{L}, \quad Q(L) \equiv 2\mu_{qi} - \kappa_{qi} L + \lambda L^3 \quad (19)$$

y las siguientes leyes de escala:

$$\mu_q = \frac{\mu_{qi}}{L^3}, \quad \kappa_q = \frac{\kappa_{qi}}{L^2}, \quad z_q = \frac{L_{,\tau}}{L} \quad (20)$$

$$1 + \delta^{(\mu)} = \frac{1 + \delta_i^{(\mu)}}{\Gamma}, \quad \frac{2}{3} + \delta^{(\kappa)} = \frac{2/3 + \delta_i^{(\kappa)}}{\Gamma} \quad (21)$$

$$2\delta^{(z)} = \hat{\Omega}^{(\mu)} \delta^{(\mu)} + [1 - \hat{\Omega}^{(\mu)} - \hat{\Omega}^{(\lambda)}] \delta^{(\kappa)} \quad (22)$$

donde los factores ‘‘Omega’’ son:

$$\hat{\Omega}^{(\mu)} \equiv \frac{\kappa_0 \rho_q}{3z_q^2} = \frac{2\mu_q}{z_q^2} = \frac{2\mu_{qi}}{Q(L)} \quad (23)$$

$$\hat{\Omega}^{(\lambda)} \equiv \frac{\kappa_0 \Lambda}{3z_q^2} = \frac{\lambda}{z_q^2} = \frac{2\lambda L^3}{Q(L)} \quad (24)$$

La métrica (1) en esta formulación del valor inicial toma una forma del tipo FLRW

$$ds^2 = -c^2 dt^2 + R_0^2 L^2 \left[ \frac{\Gamma^2 dr^2}{1 - k_{qi} R_i^2} + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2) \right] \quad (25)$$

de manera que  $L(t, r) = 0$  y  $\Gamma(t, r) = 0$  marcan respectivamente los puntos singulares de cruce de capas (*shell crossing*) y el centro. El factor escalar  $L$  se deriva de la solución implícita de (19), el cual al contrario del caso  $\lambda = 0$ , no puede ser dado (en general) en términos de funciones elementales, sino por integrales elípticas que pueden formalmente expresarse como en las referencias (Omer, 1965; Zecca, 1991)

$$\tau = Z - Z_i, \quad Z = Z(L, \mu_{qi}, \kappa_{qi}, \lambda) = \int \frac{L^{1/2} dL}{\left[2\mu_{qi} + \lambda L^3 - \kappa_{qi} L\right]^{1/2}} \quad (26)$$

por tanto, una expresión formal para  $\Gamma$  puede obtenerse por la derivación radial implícita de (26):

$$\Gamma = 1 - \frac{2\left[\mu_{qi}(Z - Z_i)_{,\mu_{qi}} \delta_i^{(\mu)} + \kappa_{qi}(Z - Z_i)_{,\kappa_{qi}} \delta_i^{(\kappa)}\right]}{LZ_{,L}} \quad (27)$$

donde los subíndices  $_{,\mu_{qi}}$  y  $_{,\kappa_{qi}}$  denotan derivadas parciales, mientras  $Z_i$  es la integral elíptica  $Z$  evaluada en  $\tau = 0$  ( $L = 1$ ). Como consecuencia, las condiciones necesarias para evitar el cruce de capas,  $\Gamma > 0$ , no pueden ser (en general) dadas analíticamente como en el caso de  $\lambda = 0$  (véanse referencias, Matravers y Humphreys, 2001; Sussman y García, 2002).

### III. CONSTRUCCIÓN DEL SISTEMA DINÁMICO

Las ecuaciones de evolución (11)-(14) conducen de una manera natural a un sistema dinámico basado en términos de las “variables normalizadas de expansión” referencias (Wainwright y Lim, 2005; Wainwright y Ellis, 1997; Coley, 2003, 1999). Siguiendo la metodología de la referencia (Sussman, 2008) definimos las variables del sistema dinámico normalizando mediante el escalar de expansión  $Q_L$ ,  $z_q$ , y considerando los parámetros “Omega” dados por (23)-(24). Definimos, además, un parámetro de evolución conveniente,  $\xi$ , por medio de la transformación de coordenadas

$$\tau = \tau(\xi, \bar{r}), \quad r = \bar{r} \quad (28)$$

de manera que cada escalar  $A(\tau, r) = A(\tau(\xi, r), r) = A(\xi, r)$ . Seleccionando  $\xi(\tau, r) = \ln(L)$ , obtendremos  $[\partial\xi/\partial\tau]r = z_q = h_s^{-1}\dot{R}/R$ , y para trayectorias comóviles ( $r = \text{constante}$ ) el operador diferencial resultante será

$$\frac{\partial}{\partial\xi} = \frac{1}{z_q} \frac{\partial}{\partial\tau} = \frac{3}{\Theta_q} \frac{\partial}{c\partial t} \quad (29)$$

Eliminando los términos  $\mu q$  y  $\lambda$  en términos de (23)-(24) y utilizando (29), el sistema (11)-(14) se transforma en el sistema dinámico:

$$\frac{\partial \hat{\Omega}^{(\mu)}}{\partial \xi} = \hat{\Omega}^{(\mu)} [\hat{\Omega}^{(\mu)} - 2\hat{\Omega}^{(\lambda)} - 1], \quad (30)$$

$$\frac{\partial \hat{\Omega}^{(\lambda)}}{\partial \xi} = \hat{\Omega}^{(\lambda)} [2 - \hat{\Omega}^{(\lambda)} + \hat{\Omega}^{(\mu)}] \quad (31)$$

$$\frac{\partial \delta^{(\mu)}}{\partial \xi} = -3 [1 + \delta^{(\mu)}] \delta^{(z)} \quad (32)$$

$$\frac{\partial \delta^{(z)}}{\partial \xi} = -\delta^{(z)} \left[ 1 + 3\delta^{(z)} \right] + \frac{\hat{\Omega}^{(\mu)}}{2} \left[ \delta^{(z)} - \delta^{(\mu)} \right] - \hat{\Omega}^{(\lambda)} \delta^{(z)} \quad (33)$$

el cual puede integrarse especificando como condiciones iniciales en la hipersuperficie inicial conveniente denotada por la constante  $\xi$ . Nótese que las hipersuperficies de  $\xi$  constante no coinciden con  ${}^3T$  (denotado por la constante  $\tau \neq 0$ ), por tanto, los gradientes radiales de un escalar  $A$  a lo largo de las primeras son diferentes de los gradientes radiales a lo largo del segundo:

$$A' = \left[ \frac{\partial A}{\partial r} \right]_r = \left[ \frac{\partial A}{\partial r} \right]_{\xi} + \frac{\partial \xi}{\partial r} \left[ \frac{\partial A}{\partial \xi} \right]_r = \left[ \frac{\partial A}{\partial r} \right]_{\xi} + \frac{R_i(\Gamma - 1)}{R_i z_q} \left[ \frac{\partial A}{\partial \tau} \right]_{\tau} \quad (34)$$

donde utilizamos (29) y (27) para eliminar  $\xi'$  y  $\partial/\partial \xi$  en términos de  $\Gamma$  y  $\partial/\partial \tau$ . Al aplicar (34) a (23)-(24) obtenemos los gradientes de  $\hat{\Omega}^{(\mu)}$  y  $\hat{\Omega}^{(\lambda)}$  a lo largo de las superficies  $\xi = \text{constante}$

$$\left[ \frac{\partial \hat{\Omega}^{(\mu)}}{\partial r} \right]_{\xi} = \frac{3R_i}{R_i} \hat{\Omega}^{(\mu)} \Gamma \left[ \delta^{(\mu)} - 2\delta^{(z)} + \frac{\Gamma - 1}{3\Gamma} (\hat{\Omega}^{(\mu)} - 2\hat{\Omega}^{(\lambda)} - 1) \right] \quad (35)$$

$$\left[ \frac{\partial \hat{\Omega}^{(\lambda)}}{\partial r} \right]_{\xi} = -\frac{6R_i}{R_i} \hat{\Omega}^{(\lambda)} \Gamma \left[ \delta^{(z)} \frac{\Gamma - 1}{3\Gamma} \right] \left( 1 + \frac{\hat{\Omega}^{(\mu)}}{2} - \hat{\Omega}^{(\lambda)} \right) \quad (36)$$

las cuales proporcionan condiciones de consistencia para (30)-(33), pues las derivadas mixtas  $\partial^2/\partial \xi \partial r$  obtenidas a partir de estas ecuaciones coinciden con aquellas obtenidas a partir del sistema.

Las ecuaciones (30)-(33) son ecuaciones diferenciales en derivadas parciales (PDES) de evolución autónoma que contienen sólo derivadas con respecto a  $\xi$ ,  $y$ , por tanto, la dependencia radial de las funciones entra como un parámetro. Aun

así, estamos lidiando con PDEs, mientras que las técnicas de los sistemas dinámicos comunes (discutidas en las referencias, Wainwright y Lim, 2005; Wainwright y Ellis, 1997; Coley, 2003, 1999) han sido concebidas para sistemas autónomos de ODEs. Sin embargo, como se muestra rigurosamente en el Apéndice B de la referencia (Sussman, 2008) un sistema de PDEs como el (30)-(33) puede considerarse un sistema dinámico apropiado, constreñido al cumplimiento de las condiciones iniciales dadas explícitamente para  $\xi = 0$

$$\hat{\Omega}_i^{(\mu)} = \frac{2\mu_{qi}}{2\mu_{qi} + \lambda - \kappa_{qi}} \tag{37}$$

$$\hat{\Omega}_i^{(\lambda)} = \frac{\lambda}{2\mu_{qi} + \lambda - \kappa_{qi}} \tag{38}$$

$$\delta_i^{(\mu)} = \frac{R_i}{3R_i} \left( \frac{[\hat{\Omega}_i^{(\mu)}]'}{\hat{\Omega}_i^{(\mu)}} - \frac{[\hat{\Omega}_i^{(\lambda)}]'}{\hat{\Omega}_i^{(\lambda)}} \right) \tag{39}$$

$$\delta_i^{(z)} = \frac{R_i}{6R_i} \frac{[\hat{\Omega}_i^{(\lambda)}]'}{\hat{\Omega}_i^{(\lambda)}} \tag{40}$$

donde utilizamos (9) y (23)-(24) evaluadas en  $\xi = 0$  (de manea que  $L = \Gamma = 1$ ). Una vez que estas condiciones se cumplen en la hipersuperficie inicial  $\xi = 0$  (que coincide con  $\tau = 0$ ), entonces podemos aplicar a la solución numérica de (30)-(33) usando las técnicas estándar de sistemas dinámicos.

**Tabla III.1**  
**Puntos críticos de varios subconjuntos invariantes**

<i>Subespacio</i>	<i>Puntos críticos</i>
Subespacio homogéneo (FLRW)	$\Omega^{(\mu)} = 1, \Omega^{(\lambda)} = 0, \delta^{(\mu)} = 0, \delta^{(z)} = 0, h(\text{PA})$
$\delta^{(\mu)} = \delta^{(z)} = 0$	$\Omega^{(\mu)} = 0, \Omega^{(\lambda)} = 1, \delta^{(\mu)} = 0, \delta^{(z)} = 0, h(\text{FA})$
	$\Omega^{(\mu)} = 0, \Omega^{(\lambda)} = 0, \delta^{(\mu)} = 0, \delta^{(z)} = 0, h(\text{SP2})$
Subespacio espacialmente plano	$\hat{\Omega}^{(\mu)} = 1, \hat{\Omega}^{(\lambda)} = 0, \delta^{(\mu)} = -1, \delta^{(z)} = -1/2, (\text{PA})$
$\hat{\Omega}^{(\mu)} = \hat{\Omega}^{(\lambda)} = 1$	$\hat{\Omega}^{(\mu)} = 0, \hat{\Omega}^{(\lambda)} = 1, \delta^{(\mu)} \text{ arbitrario}, \delta^{(z)} = 0, (\text{FA})$
	$\hat{\Omega}^{(\mu)} = 1, \hat{\Omega}^{(\lambda)} = 0, \delta^{(\mu)} = -1, \delta^{(z)} = 1/3, (\text{SP3})$
	$\hat{\Omega}^{(\mu)} = 1, \hat{\Omega}^{(\lambda)} = 0, \delta^{(\mu)} = 0, \delta^{(z)} = 0, (\text{SP4})$
	$\hat{\Omega}^{(\mu)} = 0, \hat{\Omega}^{(\lambda)} = 1, \delta^{(\mu)} = -1, \delta^{(z)} = 2/3, (\text{SP5})$

Subespacio Schwarzschild- deSitter $\delta^{(u)} = -1$	$\hat{\Omega}^{(u)}=1, \hat{\Omega}^{(\lambda)}=0, \delta^{(u)}=-1, \delta^{(z)}=-1/2, (PA)$
	$\hat{\Omega}^{(u)}=0, \hat{\Omega}^{(\lambda)}=1, \delta^{(u)}=-1, \delta^{(z)}=0, (FA)$
	$\hat{\Omega}^{(u)}=0, \hat{\Omega}^{(\lambda)}=0, \delta^{(u)}=-1, \delta^{(z)}=-1/3, (SP1)$
	$\hat{\Omega}^{(u)}=0, \hat{\Omega}^{(\lambda)}=0, \delta^{(u)}=-1, \delta^{(z)}=0, (SP2)$
	$\hat{\Omega}^{(u)}=1, \hat{\Omega}^{(\lambda)}=0, \delta^{(u)}=-1, \delta^{(z)}=1/3, (SP3)$
	$\hat{\Omega}^{(u)}=0, \hat{\Omega}^{(\lambda)}=1, \delta^{(u)}=-1, \delta^{(z)}=-2/3, (SP5)$
Polvo LTB sin $\Gamma$  $\hat{\Omega}^{(\lambda)} = 0$	$\hat{\Omega}^{(u)}=1, \hat{\Omega}^{(\lambda)}=0, \delta^{(u)}=-1, \delta^{(z)}=-1/2, (PA)$
	$\hat{\Omega}^{(u)}=0, \hat{\Omega}^{(\lambda)}=0, \delta^{(u)} \text{ arbitrario}, \delta^{(z)}=0, (FA)$
	$\hat{\Omega}^{(u)}=0, \hat{\Omega}^{(\lambda)}=0, \delta^{(u)}=-1, \delta^{(z)}=-1/3, (SP1)$
	$\hat{\Omega}^{(u)}=1, \hat{\Omega}^{(\lambda)}=0, \delta^{(u)}=-1, \delta^{(z)}=1/3, (SP3)$
	$\hat{\Omega}^{(u)}=1, \hat{\Omega}^{(\lambda)}=0, \delta^{(u)}=0, \delta^{(z)}=0, (SP4)$

Las letras (PA) y (FA) se refieren a los atractores pasado y futuro (fuente/sumidero), mientras que (SP) es un punto de silla. Las referencias (42)-(47) tienen una descripción más detallada de estos puntos críticos (excepto para aquellos del subespacio homogéneo).

#### IV. EL ESPACIO DE FASES, SUBESPACIOS INVARIANTES Y PUNTOS CRÍTICOS

##### IV.1 Espacio de fases: caso general

Las curvas  $[\hat{\Omega}^{(u)}(\xi, r), \hat{\Omega}^{(\lambda)}(\xi, r), \delta^{(u)}(\xi, r), \delta^{(z)}(\xi, r)]$  solución del sistema dinámico (30)-(33) obtenidas por integración numérica para las condiciones iniciales (37)-(40) evolucionan en un espacio de fases 4-dimensional parametrizado como

$$P = \{\hat{\Omega}^{(u)}, \hat{\Omega}^{(\lambda)}, \delta^{(u)}, \delta^{(z)}\} \subset R^4 \tag{41}$$

y caracterizado por los puntos críticos siguientes:

- Fuente (atractor pasado): (PA)

$$\hat{\Omega}^{(u)} = 1, \quad \hat{\Omega}^{(\lambda)} = 0 \quad \delta^{(u)} = -1, \delta^{(z)} = -1/2 \tag{42}$$

- Sumidero (atractor futuro): (FA)

$$\hat{\Omega}^{(u)} = 0, \quad \hat{\Omega}^{(\lambda)} = 1, \quad \delta^{(u)} \text{ arbitrario}, \quad \delta^{(z)} = 0 \tag{43}$$

- Punto silla: (SP)

$$\hat{\Omega}^{(u)} = \hat{\Omega}^{(\lambda)} = 0, \begin{cases} \delta^{(u)} = -1, & \delta^{(z)} = -1/3 & (SP1) \\ \delta^{(u)} \text{ arbitrario} & \delta^{(z)} = 0 & (SP2) \end{cases} \quad (44)$$

$$\hat{\Omega}^{(u)} = 1 \hat{\Omega}^{(\lambda)} = 0, \begin{cases} \delta^{(u)} = -1, & \delta^{(z)} = 1/3 & (SP3) \\ \delta^{(u)} = 0 & \delta^{(z)} = 0 & (SP4) \end{cases} \quad (45)$$

$$\hat{\Omega}^{(u)} = 0, \quad \hat{\Omega}^{(\lambda)} = 1, \quad \delta^{(u)} = -1, \delta^{(z)} = -2/3, \quad (SP5) \quad (46)$$

Como se muestra en la siguiente subsección, estos puntos críticos son comunes a varios subconjuntos invariantes. Nótese que el atractor futuro es un conjunto de sumideros que definen una curva en  $P$  parametrizada por  $\delta^{(u)}$ .

#### IV.2 Subespacio homogéneo y sus centros de simetría: dinámica del caso FLRW

Teniendo en cuenta (21), (22) y (27), obtenemos:

$$\delta_i^{(u)} = \delta_i^{(z)} = 0 \quad \Rightarrow \quad \delta^{(u)} = \delta^{(z)} = 0 \quad \forall \xi \neq 0, \quad (47)$$

que define el subespacio invariante “homogéneo”:

$$H \equiv \{\hat{\Omega}^{(u)}, \hat{\Omega}^{(\lambda)}, 0, 0\} \subset P, \quad (48)$$

Caracterizado por el sistema dinámico bidimensional (30)-(33), el cual sólo involucra  $\hat{\Omega}^{(u)}$  y  $\hat{\Omega}^{(\lambda)}$ , y es formalmente idéntico al sistema dinámico asociado con los espacios-tiempos FLRW con polvo más la fuente  $\Lambda$ . Las soluciones analíticas de (30)-(33) están dadas por (23) y (24) con  $L = \exp(\xi)$ , y los puntos críticos están dados en la 1: proyecciones de (PA), (FA) y (SP2) en (48). Nótese que en general y sin estar respectivamente relacionadas con las condiciones iniciales, los siguientes límites se cumplen para todas las curvas en  $H$ :

$$\xi \rightarrow \infty \text{ (o } \sigma L \rightarrow \infty \text{)} \text{ luego } \hat{\Omega}^{(u)} \rightarrow 0, \hat{\Omega}^{(\lambda)} \rightarrow 1, \quad (49)$$

$$\xi \rightarrow \infty \text{ (o } \sigma L \rightarrow 0 \text{)} \text{ luego } \hat{\Omega}^{(u)} \rightarrow 1, \hat{\Omega}^{(\lambda)} \rightarrow 0, \quad (50)$$

Por tanto, los atractores pasado y futuro de  $H$  claramente corresponden a los límites pasado y futuro  $\xi \rightarrow \pm \infty$  de trayectorias inextensibles dadas por (23)-(24).

Si  $r = 0$  marca un centro de simetría, entonces (47) tiene que cumplirse para todo  $\xi$  en  $r = 0$ . Por tanto, independientemente de la condición inicial para

$r > 0$ , si existe un cono de luz central, éste define la única trayectoria en  $P$  cuyo desarrollo espacial de fase tiene lugar completamente en (48). Esto es también válido si hay un segundo centro de simetría marcado por  $r = r_c > 0$ .

### IV.3 El subespacio espacialmente plano

Teniendo en cuenta las formas de  $\hat{\Omega}^{(u)}$  y  $\hat{\Omega}^{(\lambda)}$  dadas por (23)-(24), se obtiene el siguiente resultado importante:

$$\hat{\Omega}^{(u)} + \hat{\Omega}^{(\lambda)} - 1 = 0 \Rightarrow \hat{\Omega}^{(u)} + \hat{\Omega}^{(\lambda)} - 1 = 0, \quad \forall \xi \neq 0. \quad (51)$$

Como  $\hat{\Omega}^{(u)} + \hat{\Omega}^{(\lambda)} = 1$  es equivalente a  $\kappa = 0$ , y  $\kappa q = 0$  implica  ${}^3R_4 = E = 0$ , lo cual (de (5)) implica  ${}^3R = 0$ , y la condición (51) define el subconjunto invariante espacialmente plano.

$$SF = \{\hat{\Omega}^{(u)}, \hat{\Omega}^{(\lambda)}, \delta^{(u)}, \delta^{(\xi)}, \delta^{(\zeta)} \mid \hat{\Omega}^{(u)} + \hat{\Omega}^{(\lambda)} - 1 = 0\} \subset P, \quad (52)$$

Nótese que la intersección de  $SF$  y  $H$  es la línea recta  $[\hat{\Omega}^{(u)}(\xi) + \hat{\Omega}^{(\lambda)}(\xi) = 1]$  en el plano  $[\hat{\Omega}^{(u)}, \hat{\Omega}^{(\lambda)}]$ . Este es el subespacio invariante que corresponde al modelo “ $\Lambda$ -CDM”: una cosmología FLRW cuya fuente es polvo (CDM) y una constante cosmológica. Para  $\hat{\Omega}^{(u)} = 0.3$  y  $\hat{\Omega}^{(\lambda)} = 0.7$ , este modelo es generalmente considerado como el que mejor ajusta las observaciones cosmológicas a gran escala y en la presente era cósmica.

El sistema tridimensional dinámico asociado con  $SF$  se obtiene desde el sistema general utilizando la restricción  $\hat{\Omega}^{(u)} = 1 - \hat{\Omega}^{(\lambda)}$ . Sus puntos críticos se pueden hallar en la Tabla 1. Los atractores pasado y futuro coinciden exactamente con aquellos del caso general. Esto indica que todos los modelos tienen un estado asintótico inicial con curvatura espacial cero, mientras que todos los modelos en los que  $\xi$  puede ser extendido infinitamente, evolucionan hacia un estado asintótico final de curvatura cero. Los puntos de silla (SP1) y (SP2) no son puntos críticos en este caso.

### IV.4 La solución Kottler: vacío Schwarzschild-De Sitter

Consideramos las condiciones iniciales con valores arbitrarios para  $\hat{\Omega}_i^{(u)}$ ,  $\hat{\Omega}_i^{(\lambda)}$ ,  $\delta_i^{(z)}$  pero  $\delta_i^{(u)} = -1$ . Entonces, si no se presenta cruce de capas ( $\Gamma > 0$ ), la ley de escala (21) implica que  $\delta^{(u)}(\xi) = -1$  se cumple para todo  $\xi$ . Insertando  $\delta^{(u)} = -1$  en (8) y considerando (3), esta condición define un subespacio tridimensional invariante.

$$V \equiv \{\hat{\Omega}^{(u)}, \hat{\Omega}^{(\lambda)}, -1, \delta^{(z)}\} \subset P \quad (53)$$

caracterizado por:  $\rho = 0$ , con  $(\kappa_0 c^2 / 3)\rho_q R^3 = 2\mu_q h R^3 = 2M_0$  y  $\mu_q = M_0 / R_3$ . Estos parámetros identifican la solución de vacío Kottler ([18]) (o vacío Schwarzschild-De Sitter) como el límite de vacío del modelo  $\Lambda$ -LTB con masa Schwarzschild  $M_0$ , y dadas en coordenadas comóviles construidas con observadores en las geodésicas radiales (nótese que (4) con  $M = M_0$  es la ecuación de estas geodésicas).

El sistema dinámico tridimensional en este caso se obtiene al sustituir (53) en (30)-(33), mientras que los puntos críticos correspondientes se hallan en la Tabla IV.1. La fuente (PA) y todos los puntos de silla excepto (SP4) coinciden exactamente con aquellos del caso general, mientras que el atractor futuro (FA) está contenido en el atractor futuro del caso general: este es el punto  $\delta^{(u)} = -1$  en la curva de sumideros parametrizada por  $\delta^{(u)}$ .

#### IV.5 Polvo LTB con $\Lambda = 0$

Si las condiciones iniciales se seleccionan con  $\lambda = 0 = \hat{\Omega}^{(\lambda)} = 0$ , entonces  $\hat{\Omega}^{(\lambda)}(\xi) = 0$ , se cumple para todo  $\xi$ . En este caso tenemos un subespacio tridimensional invariante

$$D \equiv \{\hat{\Omega}^{(u)}, 0, \delta^{(u)}, \delta^{(z)}\} \subset P \quad (54)$$

Asociado con las soluciones de polvo LTB estándares sin la constante cosmológica. Este subconjunto fue examinado en detalle en (Sussman, 2008).<sup>4</sup> El correspondiente sistema dinámico tridimensional se manifiesta cuando  $\hat{\Omega}^{(\lambda)} = 0$  en (30)-(33), con los puntos críticos que aparecen en la Tabla III.1. Nótese que la fuente coincide exactamente con la fuente (PA) del caso general, indicando como la constante cosmológica no juega ningún rol en la evolución temprana (cerca de la curvatura inicial singular  $\xi \rightarrow -\infty$ ). Los puntos de silla (SP2) y (SP5) no son puntos críticos de este subsistema. El atractor futuro (FA) es diferente al del caso general, lo cual indica que  $\lambda$  juega un rol distintivo y dominante en la evolución tardía en capas en las cuales  $\xi \rightarrow \infty$ .

<sup>4</sup> Las variables del espacio de fases  $\hat{\Omega}^{(u)}$ ,  $\delta^{(u)}$  y  $\delta^{(z)}$ , fueron denotadas en la referencia (Sussman, 2008) como  $\langle \Omega \rangle$ ,  $\Delta^{(m)}$  y  $-S$ .

## V. CINEMÁTICA DE LAS CAPAS DE POLVO

La relación entre el tipo posible de evolución cinemática y las condiciones iniciales se puede apreciar al mirar las raíces de  $L_{,\tau}$  en la ecuación (19), o equivalentemente al mirar a la raíz real positiva de la cúbica  $Q(L)$ , la cual puede reescribirse como

$$Q(L) = \beta_i L^3 - \alpha_i L + 1 = 0 \quad (55)$$

$$\beta_i = \frac{\hat{\Omega}_i^{(\lambda)}}{\hat{\Omega}_i^{(\mu)}} = \frac{\lambda}{2\mu_{qi}}, \quad \alpha_i = \frac{\hat{\Omega}_i^{(\mu)} + \hat{\Omega}_i^{(\lambda)} - 1}{\hat{\Omega}_i^{(\mu)}} = \frac{\kappa_{qi}}{2\mu_{qi}} \quad (56)$$

Aunque (56) tiene solución exacta, es más práctico examinarla cualitativamente estudiando el comportamiento del polinomio cúbico  $Q(L)$  (véase Figura V.1a). De este análisis cualitativo se pueden distinguir los siguientes casos:

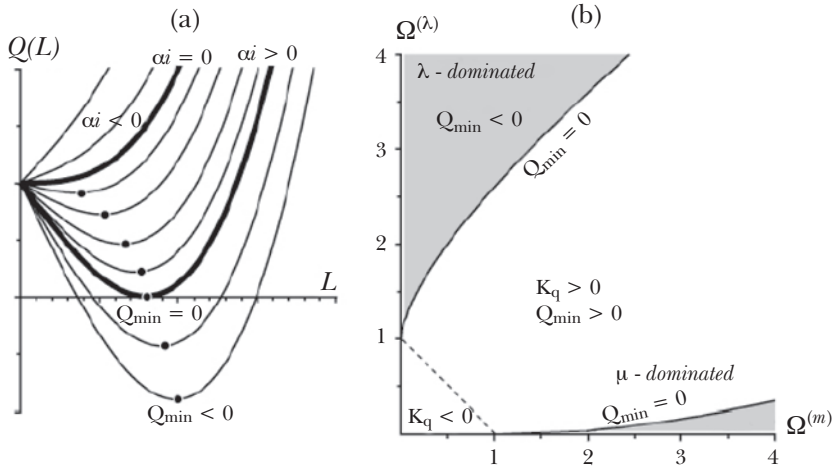
- Si  $\alpha_i \leq 0$  entonces  $Q(L)$  no tiene puntos extremales y  $Q(L) \geq 1$  se cumple para todo  $L$ . Este caso claramente corresponde a la configuración de expansión o colapso perpetuos con  $\hat{\Omega}_i^{(\mu)} + \hat{\Omega}_i^{(\lambda)} < 1$  (or  $\kappa_{qi} \leq 0$ ).
- Si  $\alpha_i > 0$  entonces  $Q(L)$  tiene un mínimo en  $L = L_{\min} = [\beta_i/(3\alpha_i)]^{1/2}$ , de manera que

$$Q_{\min} = Q(L_{\min}) = 1 - \frac{\kappa_{qi}^{3/2}}{3\sqrt{3}\lambda\mu_{qi}} = 1 - \frac{2\left[\hat{\Omega}_i^{(\mu)} + \hat{\Omega}_i^{(\lambda)} - 1\right]^{3/2}}{3\sqrt{3}\left[\hat{\Omega}_i^{(\lambda)}\right]^{1/2}\hat{\Omega}_i^{(\mu)}} \quad (58)$$

- Si  $\alpha_1 \leq 0$  entonces  $Q(L)$  no tiene puntos extremales y  $Q(L) \geq 1$  se cumple para todo  $L$ . Este caso claramente corresponde a la configuración de expansión o colapso perpetuo con  $\hat{\Omega}_i^{(\mu)} + \hat{\Omega}_i^{(\lambda)} > 1$  (o  $k_{qi} \leq 0$ ).
- Si  $\alpha_1 > 0$  entonces  $Q(L)$  tiene un mínimo en  $L = L_{\min} = [\beta_i/(3\alpha_i)]^{1/2}$ , de manera que

$$Q_{\min} = Q(L_{\min}) = 1 - \frac{\kappa_{qi}^{3/2}}{3\sqrt{3}\lambda k_{qi}} = 1 - \frac{2\left[\Omega_i^{(\mu)} + \Omega_i^{(\lambda)} - 1\right]^{3/2}}{3\sqrt{3}\left[\Omega_i^{(\lambda)}\right]^{1/2}\Omega_i^{(\mu)}} \quad (57)$$

Claramente, como  $Q(L) \geq Q_{\min}$ , entonces (55) no tiene raíces positivas para  $Q_{\min} > 0$  (véase Figura V.1a), que corresponde a la configuración de expansión/



**Figura V.1. Condiciones iniciales y cinemática.** La figura (a) muestra el polinomio cúbico  $Q(L)$  definido por (56), sus raíces determinan los ceros de  $L_\tau$ . Las configuraciones que corresponden a  $Q_{\min} > 0$  (entre las dos líneas sólidas) se expanden perpetuamente (sin ceros) pero presentan una curvatura espacial positiva. Las configuraciones que cumplen con  $Q_{\min} \leq 0$  deben rebotar (experimentan una expansión seguida de un colapso) o colapsar. La figura (b) muestra las regiones en el espacio de las condiciones iniciales  $\{\hat{\Omega}_i^{(\mu)}, \hat{\Omega}_i^{(\lambda)}\}$  para las cuales se cumple  $Q_{\min} < 0$  (región sombreada).

colapso perpetuos con  $\hat{\Omega}_i^{(\mu)} + \hat{\Omega}_i^{(\lambda)} > 1$  (o  $k_{qi} > 0$ ). Sin embargo, una condición suficiente para la existencia de raíces positivas en 56) es

$$Q_{\min} (\hat{\Omega}_i^{(\mu)} + \hat{\Omega}_i^{(\lambda)}) \leq 0, \tag{58}$$

la cual corresponde a las configuraciones que tienen valores de retorno en los que el rango de  $L$  está restringido por un punto de retorno donde  $L_\tau = 0$ .

A fin de examinar las condiciones iniciales asociadas con las posibles evoluciones cinemáticas ya descritas, examinemos la condición (59) representándola gráficamente en la Figura V.1b las curvas de nivel de  $Q_{\min}$  en el plano paramétrico  $\{\hat{\Omega}_i^{(\mu)}, \hat{\Omega}_i^{(\lambda)}\}$ . Como estos valores iniciales de las funciones dependen de  $r$ , cualquier conjunto de condiciones iniciales dadas aparecerá como una curva  $[\hat{\Omega}_i^{(\mu)}(r), \hat{\Omega}_i^{(\lambda)}(r)]$  en este plano. Las regiones donde se cumple (58) están marcadas por las áreas sombreadas en la Figura V.1b: la esquina superior izquierda corresponde a los regímenes dominados por  $\lambda$  (“ $\lambda$ -dominated”) y la esquina inferior derecha a los dominados por  $\mu$  (“ $\mu$ -dominated”). Es interesante notar que el régimen “ $\mu$ -dominated”

está más restringido por las condiciones (58), porque se necesita un mayor  $\hat{\Omega}_i^{(u)}$  aun para pequeños  $\hat{\Omega}_i^{(\lambda)} < 1$ .

Los parámetros de evolución cinemática de las capas de polvo en modelos  $\Gamma$ -LTB y su relación con las condiciones iniciales se describen a continuación:

1. *Expansión perpetua (Colapso) desde (hacia) una singularidad.* Si  $\kappa_{qi} \leq 0$ , o si  $\kappa_{qi} > 0$  pero  $\hat{\Omega}_i^{(u)}$  y  $\hat{\Omega}_i^{(\lambda)}$  son comparables para todo  $r$ , nos encontramos en la zona no sombreada de la Figura V.1b y (59) no se cumple. Véanse Figuras VIII.1.1.2, VIII.1.2.3 y VIII.1.2.4.
2. *Todas las capas iniciales se expanden y colapsan en una singularidad.* Este patrón corresponde a los extremos del régimen “ $\mu$ -dominated” en la figura 1b y al rango de restricción  $0 < L < L_1$  con el retorno (bounce)  $L_{,au} = 0$  en  $L = L_1$ . Las condiciones iniciales tienen que cumplir que  $\hat{\Omega}_i^{(u)} = 1$  y  $\hat{\Omega}_i^{(\lambda)} = 1$  para todo  $r$ . Si  $\hat{\Omega}_i^{(\lambda)} = 0$  (polvo puro) entonces  $\kappa_{qi} > 0$  es una condición suficiente para esta evolución, pero  $\kappa_{qi} > 0$  es sólo necesaria (no suficiente) si  $\hat{\Omega}_i^{(\lambda)} > 0$ . Véase Figura VIII.2.5.
3. *Todas las capas iniciales colapsan desde infinito, rebotan y se re-expanden.* Este patrón corresponde al régimen “ $\lambda$ -dominated” en la Figura V.1b y al rango de restricción  $L > L_2$  donde el punto de retorno de  $L$  ocurre en  $L = L_2$ , de manera que las capas eluden el colapso. Las condiciones iniciales tienen que cumplir con  $\kappa_{qi} > 0$ ,  $\hat{\Omega}_i^{(u)} = 1$  y  $\hat{\Omega}_i^{(\lambda)} = 1$  para todas las capas. Véase la Figura VIII.2.6.
4. *Evolución de tipo “loitering”.* Las capas iniciales que se expanden o colapsan alcanzan un punto estático inestable. Las capas pueden colapsar o expandirse más adelante (los modelos  $\Lambda$ -CDM con métrica FLRW y con esta cinemática son conocidos como modelos Eddington-Lemaitre). Las condiciones iniciales asociadas con este patrón cumplen con  $\kappa_{qi} > 0$  y  $Q_{\min} = 0$ . Véase Figura VIII.2.3.7.
5. *Patrones mixtos de evolución.* Como  $\hat{\Omega}_i^{(u)}$ ,  $\hat{\Omega}_i^{(\lambda)}$  dependen de  $r$ , las capas de polvo pueden mostrar diferentes patrones de evolución en diferentes rangos radiales (en oposición al caso FLRW homogéneo). En estos casos las condiciones iniciales tienen que ser especificadas de manera que (59) sólo se cumpla en una región determinada, es decir, la curva  $[\hat{\Omega}_i^{(u)}(r) \text{ y } \hat{\Omega}_i^{(\lambda)}(r)]$  cruza desde el área no sombreada hacia cualquiera de las áreas sombreadas de la Figura V.1b. Las dos configuraciones posibles son:
  - Las capas alrededor del centro siguen el patrón (ii) y capas “externas” evolucionan en (i). Véase Figura VIII.3.8.
  - Las capas alrededor del centro siguen el patrón (i) y capas “externa” evolucionan en (iii). Véase Figura VIII.3.1.9.

En la siguiente sección, examinamos los espacios de fases de configuraciones siguiendo los patrones de evolución (i)-(v). Recomendamos al lector consultar la referencia (Felten, 1986) para hallar gráficos detallados del factor de escala de los patrones (i)-(iv) en cosmologías FLRW.

## VI. TOPOLOGÍA DE LAS REBANADAS ESPACIALES Y ELECCIÓN DE COORDENADA RADIAL

La métrica LTB (en ambas formas (1) y (25)) admite un reescalado arbitrario de la coordenada radial. Usamos esta libertad de gauge para fijar la función  $R_i$  en (18) y (25) según la relación

$$R_i = h_s^{-1} f(r), \quad (59)$$

donde  $f$  es una función adimensional suave, mientras que  $h_s^{-1}$  (definida en (10)) representa una escala de longitud para la coordenada radial. La clase topológica de las rebanadas  ${}^3T[t]$  está dada por el número de centros de simetrías, que determina las formas permitidas de  $f$ . Si se asume que no hay cruce de capas, entonces  $\Gamma > 0$  se cumple siempre, las topologías admisibles para las rebanadas  ${}^3T(t)$  son:

- *Modelos abiertos.* Las rebanadas  ${}^3T[t]$  son homeomórficas a  $R^3$  y admiten sólo un centro de simetría en  $r = 0$ . Como  $R' > 0$  y  $R'_i > 0$  tiene que cumplirse para todo  $r$ , entonces  $f$  puede ser cualquier función monótonamente creciente que cumpla con  $f(0) = 0$ . La opción más simple es  $f(r) = r$ , pero usaremos  $f(r) = \tan r$ , de manera que  $R_i \rightarrow \infty$  corresponda al límite finito  $r \rightarrow \pi/2$ .
- *Modelos cerrados.* Las rebanadas  ${}^3T[t]$  son homeomórficas a  $S^3$  y admiten dos centros de simetría en  $r = 0$  y  $r = r_c$ . Como  $R(t, 0) = R(t, r_c) = 0$  para todo  $t$ , entonces tiene que existir (al menos) un valor de retorno  $r = r_{tw}$  de manera que  $R'(ct, r_{tw}) = 0$  donde  $0 < r_{tw} < r_c$ . La función  $f$  debe cumplir  $f(0) = f(r_c) = 0$  y  $f'(r_{tw}) = 0$ . Una opción conveniente es  $f(r) = \sin r$ .
- *Agujeros de gusano.* Las rebanadas  ${}^3T[t]$  pueden ser homeomórficas a  $S^2 \times R$  o a  $S^2 \times S^1$ . Mientras que las rebanadas  ${}^3T[t]$  no admiten centros de simetría, debe de haber (al menos) un valor de retorno  $R'(ct, r_{tw}) = 0$ . La función  $f$  no puede tener ceros, pero tiene que cumplir con  $f'(r_{tw}) = 0$ . Usaremos  $f(r) = \sec r$  en los ejemplos de la sección VIII.

A fin de examinar la relación entre la clase topológica de los  ${}^3T[t]$ , el signo de la curvatura espacial y la evolución cinemática de las capas de polvo, haremos

énfasis en que, aun cuando  $\Gamma > 0$  se cumple,  $R_i$  tiene que cumplir con la siguiente condición asociada con la regularidad de los coeficientes métricos  $g_{rr}$  en (1) y (25) (Sussman, 2010a, 2010b; Matravers y Humphreys, 2001; Sussman y García, 2002).

$$\frac{R'_i}{\left|1 - \kappa_{qi} R_i^2 h_s^2\right|^{1/2}} = \frac{h_s^{-1} f'}{\left|1 - \kappa_{qi} f^2\right|^{1/2}} \neq 0 \quad (60)$$

Por tanto, si hubiera un valor de retorno  $r = r_{tv}$  tal como  $R'_i(r_{tv}) = 0$ , los denominadores en la expresión anterior tienen que tener el mismo orden cero en  $r = r_{tv}$ . Esto conlleva a la siguiente restricción para  $f$  y  $\kappa_{qi}$

$$f'(r_{tv}) = 0 \Leftrightarrow \kappa_{qi}(r_{tv}) = \frac{1}{f^2(r_{tv})} > 0 \quad (61)$$

Si (60)-(61) no se cumplen, tendremos una singularidad en la capa de superficie en  $r = r_{tv}$  (Sussman, 2010a, 2010b; Matravers y Humphreys, 2001; Sussman y García, 2002). Por tanto, (62) implica que todos los modelos regulares con topologías cerradas y los agujeros de gusano siempre tendrán una curvatura espacial positiva:  $\kappa_{qi} > 0$  o  $\hat{\Omega}_i^{(u)} + \hat{\Omega}_i^{(\lambda)} - 1 > 0$ . En el caso de los modelos LTB de polvo puro ( $\lambda = 0$ ), la implicación de (62) significa que las capas de polvo para todos los modelos tienen retorno y colapso (dado que si  $\kappa_{qi} > 0$ ,  $L_\tau \tau$  tiene necesariamente un cero). Sin embargo, la situación es diferente si  $\lambda > 0$ , pues las capas de polvo se pueden expandir para toda  $\tau$  cuando  $\kappa_{qi} > 0$  siempre que los valores iniciales de las funciones no cumplan (58) (véase Figura VIII.1). Por tanto, los modelos  $\Lambda$ -LTB con topologías cerrada o de agujero de gusano totalmente regulares y que se expandan perpetuamente son posibles.

## VII. LA DINÁMICA DEL ESPACIO DE FASES EN TÉRMINOS DE PROYECCIONES HOMOGÉNEAS E INHOMOGÉNEAS

Como los subespacios homogéneos invariantes  $H$  definidos en (48) corresponden a cosmologías FLRW de polvo- $\Lambda$ , las variables dinámicas de  $\hat{\Omega}_i^{(u)}$  y  $\hat{\Omega}_i^{(\lambda)}$  se corresponden con los parámetros Omega de FLRW y el sistema dinámico para  $H$  (el cual es (30) y (31)) es equivalente al sistema dinámico de estas cosmologías FLRW (en las variables normalizadas de expansión (Wainwright y Lim, 2005; Wainwright y Ellis, 1997; Coley, 2003, 1999). Definimos para el último el espacio de fases bidimensional.

$$H_{(0)} = \{\hat{\Omega}^{(u)}, \hat{\Omega}^{(\lambda)}\} \subset R^2, \quad \text{tal que} \quad H = i(H_{(0)}) \quad (62)$$

donde  $i$  es el mapa de inclusión  $i: R^2 \rightarrow R^4$ ,  $i(x^1, x^2) = [x^1, x^2, 0, 0]$ .

Como  $H_{(0)}$  está dotado con un significado geométrico y físico claro (el espacio fase del límite homogéneo del espacio tiempo FLRW), mientras que  $\{\delta^{(u)}, \delta^{(z)}\}$  representa la desviación dinámica con respecto del perfil homogéneo (es decir, representan las perturbaciones no lineales covariantes e invariantes gauge (Sussman, 2009)). Asimismo, es útil expresar el espacio de fases general  $P$  como el producto directo

$$P = H_{(0)} \times I, \quad I \equiv \{\delta^{(u)}, \delta^{(z)}\} \subset R^2, \quad (63)$$

de modo que las trayectorias y los puntos críticos de  $P$  pueden describirse y analizarse por medio de los siguientes mapas de proyecciones para cada  $p = [\hat{\Omega}^{(u)}, \hat{\Omega}^{(\lambda)}, \delta^{(u)}, \delta^{(z)}] \in P$ :

Proyección homogénea.

$$\Pi_H: P \rightarrow H_{(0)}, \quad \Pi_H(p) = [\hat{\Omega}^{(u)}, \hat{\Omega}^{(\lambda)}] \in H_{(0)} \quad (64)$$

Proyección inhomogénea.

$$\Pi_I: P \rightarrow I, \quad \Pi_I(p) = [\delta^{(u)}, \delta^{(z)}] \in I \quad (65)$$

Mientras el uso de (64) y (65) resuelve el problema práctico de lidiar con el espacio de fases 4-dimensional  $P$ , estas proyecciones resaltan características dinámicas interesantes de  $P$ , que se discuten a continuación.

Como las ecuaciones dinámicas (30) y (31) asociadas a  $H_{(0)}$  son completamente independientes de  $\delta^{(u)}, \delta^{(z)}$  (las coordenadas de  $I$ ), es posible mapear uno a uno cada trayectoria de espacio de fases FLRW en  $H_{(0)}$  en una trayectoria de espacio fase de  $P$  proyectada en  $H_{(0)}$  por  $\Pi_H$ . Sin embargo, las condiciones iniciales para cada trayectoria en  $H_{(0)}$  son dos constantes  $\hat{\Omega}_i^{(u)}, \hat{\Omega}_i^{(\lambda)}$  y, por tanto, cada trayectoria de espacio de fases corresponde a una sola configuración espacio tiempo. Por otra parte, debido a las inhomogeneidades de los espacios tiempos  $\Lambda$ -LTB, las condiciones iniciales están dadas (en general) en funciones de  $r$ . Por tanto, estas trayectorias forman una familia de curvas de un parámetro, de manera que cada curva (cuando se proyecta en  $H_{(0)}$  a través de  $\Pi_H$ ) es equivalente a una trayectoria de un espacio tiempo FLRW distintivo.

Los puntos críticos de  $H_0$  son simplemente los puntos críticos de  $H$  (véase Tabla III.1) proyectados bajo  $\Pi_H$ , y corresponden a casos particulares de espacios tiempos:

$$\hat{\Omega}_i^{(u)} = 1, \hat{\Omega}_i^{(\lambda)} = 0 \quad \text{Atractor pasado: FLRW espacialmente plano,} \quad (66)$$

$$\hat{\Omega}_i^{(\mu)} = 0, \hat{\Omega}_i^{(\lambda)} = 1 \quad \text{Atractor futuro: De Sitter,} \quad (67)$$

$$\hat{\Omega}_i^{(\mu)} = 0, \hat{\Omega}_i^{(\lambda)} = 0 \quad \text{Punto de silla: Minkowski} \quad (68)$$

Sin embargo, cada uno de estos puntos críticos de  $P$  listados en la sección VI es también proyectado por  $\Pi_H$  en uno de los puntos críticos de  $H_0$  listados a continuación:

- (PA), (SP3) y (SP4) son proyectados en el atractor pasado (66)
- (FA) y (SP5) son proyectados en el atractor futuro (67)
- (SP1) y (SP2) proyectados en el punto de silla (68)

Estas proyecciones claramente sugieren asociar cada punto crítico de  $P$  con uno relacionado con un espacio tiempo limitante conocido, tal como FLRW espacialmente plano, De Sitter o Minkowski. La degeneración al proyectar los puntos críticos de  $P$  en alguno de los puntos críticos (66)-(68) ocurre cuando sólo miramos las proyecciones de las trayectorias de  $P$  en  $H_{(0)}$  por  $\Pi_H$ , pero ésta desaparece cuando examinamos la proyección de estas curvas en  $I$  utilizando (66), cuyas coordenadas  $\{\delta^{(\mu)}, \delta^{(z)}\}$  representan los efectos de inhomogeneidad. Por tanto, se obtiene una representación global de la evolución de las trayectorias en  $P$  y su relación con los puntos críticos cuando combinamos la descripción en las dos proyecciones (64) y (65).

Las proyecciones (64) y (65) serán utilizadas en la siguiente sección para analizar las trayectorias de espacio de fases de varias configuraciones  $\pi$ -LTB representativas asociadas con los patrones de cinemática y topologías de las rebanadas de espacio discutidas en las secciones V y VI.

## VIII. EJEMPLOS NUMÉRICOS Y GRÁFICOS

Examinamos, por medio de las soluciones numéricas de (30)-(33) bajo las condiciones iniciales (37)-(40), la evolución en el espacio de fases de varias configuraciones  $\Lambda$ -LTB características. Al representar gráficamente la curva  $[\hat{\Omega}_i^{(\mu)}(r), \hat{\Omega}_i^{(\lambda)}(r)]$  en el espacio de parámetros de la Figura 1b (sección VIII), podemos inferir para cualquier conjunto de funciones iniciales uno de los patrones de evolución cinemático (i)-(iv) descritos en la sección V. La clase topológica de las rebanadas de espacio dependen de la selección de  $f(r)$  en (59). Para evitar lidiar con singularidades de cruce de capas, verificamos numéricamente que se cumpla la condición  $\Gamma > 0$  para cada configuración (nótese que  $\Gamma_{,\tau} = 3\Gamma_{z_q} \delta^{(z)}$  se deriva de (6), (17) y (27)).

Como el espacio de fases  $P$  es 4-dimensional, representando gráficamente las trayectorias (curvas parametrizadas por  $\xi$  con  $r$  constante) para cualquier conjunto de condiciones iniciales en ocho figuras compuestas de dos paneles, y cada panel contiene una gráfica bidimensional correspondiente a:

- Panel “(a)” (lado izquierdo): la congruencia de las curvas  $[\hat{\Omega}^{(u)}(\xi, r), \hat{\Omega}^{(\lambda)}(\xi, r)]$  dada por la proyección homogénea de  $P$  definida en (65).
- Panel “(b)” (lado derecho): la congruencia de las curvas  $[\delta^{(u)}(\xi, r), \delta^{(z)}(\xi, r)]$  dada por la proyección inhomogénea de  $P$  definida por (66).

Estos gráficos muestran las congruencias de las trayectorias, junto con los puntos críticos señalados como puntos gruesos (atractores pasado,  $PA$ , y futuro,  $FA$ , y los puntos de silla  $SP1$ ,  $SP2$ , etc.). En las configuraciones que contienen capas de polvo que colapsan o rebotan (subsecciones VIII.2 y VIII.3, Figuras 6-9) las variables del espacio de fases  $\hat{\Omega}^{(u)}$ ,  $\hat{\Omega}^{(\lambda)}$  y  $\delta^{(z)}$  divergen cuando  $z_q \rightarrow 0$ . Para estos casos representamos la arctan de las variables de manera que  $\pm$  infinito es igual a  $\pm \pi/2$ . Sin embargo, más allá de estos puntos la integración numérica (trayectoria) de (30)-(33) no se puede resolver. En estos casos debemos utilizar el sistema (11)-(14) con la variable  $s = -z_q \delta^{(z)}$  (que esta acotada cuando  $z_q \rightarrow 0$ ) para examinar el comportamiento de  $\hat{\Omega}^{(u)}$ ,  $\hat{\Omega}^{(\lambda)}$  y  $\delta^{(z)}$  en el rango completo de evolución.

### VIII.1 Configuración con expansión perpetua: patrón cinemático (i)

Si las condiciones iniciales se seleccionan de tal manera que la curva  $[\hat{\Omega}^{(u)}(r), \hat{\Omega}^{(\lambda)}(r)]$  está contenida completamente en la región no sombreada de la figura VI.1b y  $Z_{qi} > 0$  para todo  $r$ , las configuraciones resultantes se expanden o colapsan monótonamente para todo  $\xi$ . Como una evolución de colapso es equivalente a una expansión bajo cambio de signo de  $\xi$ , solamente se considerará la expansión en los ejemplos a continuación.

#### VIII.1.1 Curvatura espacial negativa

Considerando  $f(r) = \tan r$  en (60) en el rango  $0 \leq r < \pi/2$ , de manera que  $R_i \rightarrow \infty$  cuando  $r \rightarrow \pi/2$ , se obtienen las primeras funciones iniciales como:

$$\mu_{qi} = m_{10} + \frac{m_{11} - m_{10}}{1 + f^3(r)}, \quad (m_{10} = 0, m_{11} = 15.3), \quad (69)$$

$$\kappa_{qi} = k_{10} + \frac{k_{11} - k_{10}}{1 + f^4(r)}, \quad (k_{10} = -1.2, k_{11} = -0.1), \quad (70)$$

$$\lambda = 0.1 \quad (71)$$

Las trayectorias del espacio de fases se muestran en la Figura VIII.1.1.2, destacando el atractor pasado PA de donde emergen todas las trayectorias (y puede asociarse con la singularidad inicial). Las curvas tienden al punto de silla mSP2 y terminan en un punto de la línea de atractores futuros FA (línea gruesa gris del panel 2). Debido a la degeneración en la coordenada  $\delta^{(u)}$  en SP2 y FA, las curvas del panel (b) cortan el eje de las  $Dz$  en diferentes valores de  $\delta^{(u)}$ . Como  $\mu_{qi} \rightarrow 0$  cuando  $R_i \rightarrow \infty$ , esta configuración es radialmente asintótica a la configuración de vacío de Schwarzschild-De Sitter. Esto es consistente con el hecho que en el límite  $r \rightarrow \pi/2$  (que implica  $R_i \rightarrow \infty$ ) se tienen condiciones del vacío cercano  $\hat{\Omega}^{(a)} \approx 0$  y  $\hat{\Omega}^{(u)} \approx 0$  (Sussman, 2010b) y de esta manera las trayectorias en el panel (a) tienden al punto de silla de la esquina derecha inferior en el estado intermedio de la evolución entre PA y FA. Además, las curvas en el panel (b) en el límite  $r \rightarrow \pi/2$  alcanzan FA con  $\delta^{(u)} \rightarrow -1$ , lo que caracteriza el subespacio invariante de un vacío de Schwarzschild-De Sitter (véase sección IV.4).

Si consideramos las condiciones iniciales (69)-(71), pero con  $m_{10} = 0.7$  de manera que  $\mu_{qi} \rightarrow 0.7$  cuando  $R_i \rightarrow \infty$ , obtenemos una configuración con la misma evolución temporal pero radialmente asintótica a un modelo de  $\pi$ -polvo FLRW con curvatura negativa (Sussman, 2010b). Las trayectorias de espacio de fases (mostradas en la Figura VIII.1.2.3) son cualitativamente análogas a aquellas de la Figura VIII.1.2.2, pero las diferentes condiciones asintóticas radiales implican que las trayectorias cuando  $r \rightarrow \pi/2$  ( $R_i \rightarrow \infty$ ) no llegan a alcanzar la esquina inferior izquierda del panel (a) y no llegan hasta  $\delta^{(u)} \rightarrow -1$  en el panel (b) (como se esperaría si el modelo es radialmente asintótico a un modelo FLRW que contiene  $\Lambda$ -polvo).

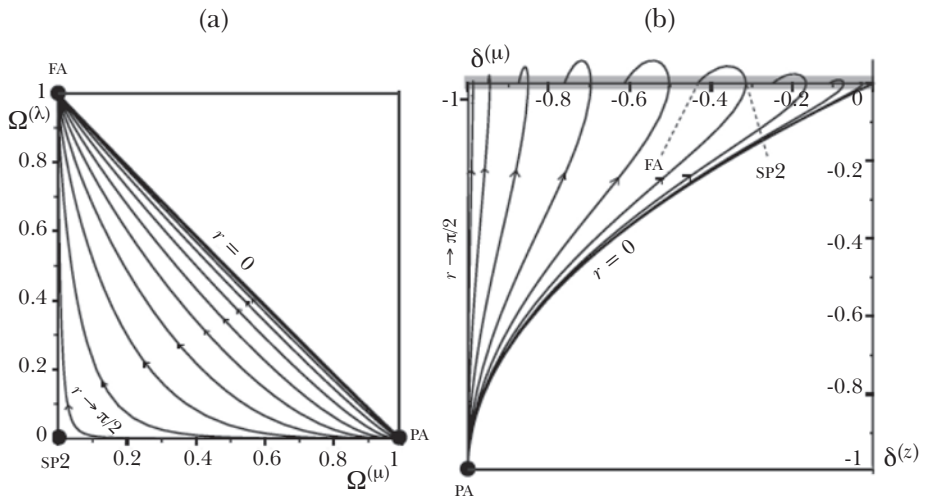
### VIII.1.2 Curvatura espacial positiva con topología abierta

Seleccionamos  $f(r) = \tan r$  en el rango  $0 \leq r < \pi/2$  para las siguientes funciones iniciales

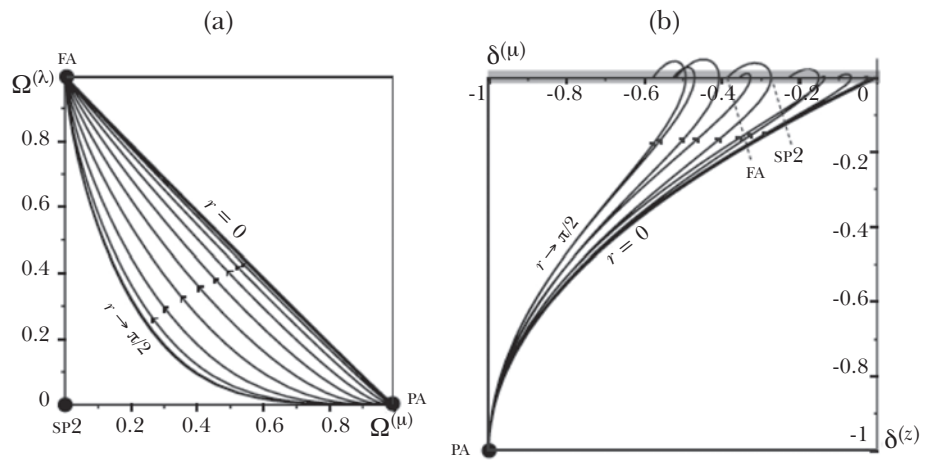
$$\mu_{qi} = m_{10} + \frac{m_{11} - m_{10}}{1 + f^2(r)}, \quad (m_{10} = 1.1, m_{11} = 50.0), \quad (72)$$

$$\kappa_{qi} = k_{10} + \frac{k_{11} - k_{10}}{1 + f^4(r)}, \quad (k_{10} = 30.0, k_{11} = 0.5), \quad (73)$$

$$\lambda = 0.10 \quad (74)$$

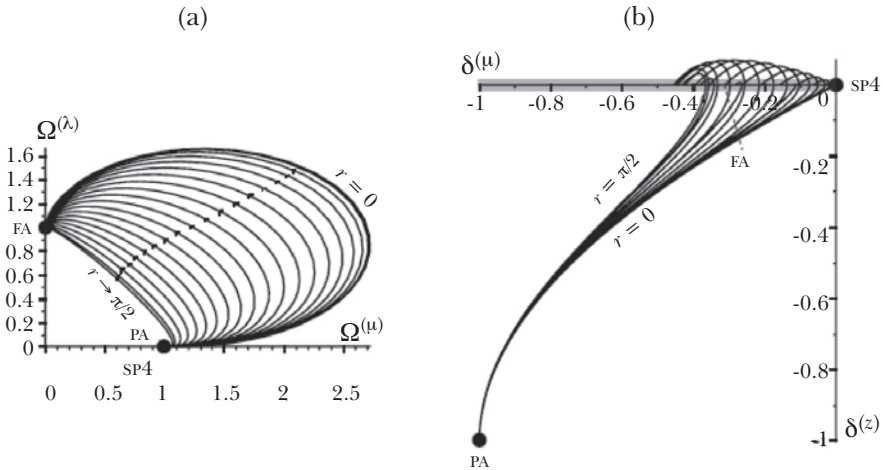


**Figura VIII.1.1.2.** Configuración de expansión con curvatura negativa y asintótica a Schwarzschild-De Sitter. Estas trayectorias del espacio de fases cumplen con las condiciones iniciales (69)–(71).



**Figura VIII.1.2.3.** Configuración de expansión con curvatura negativa y asintótica a FLRW. Las condiciones iniciales y las convenciones de la figura son las mismas utilizadas en la Figura 2 pero con  $\mu_{qi}$  que tiende asintóticamente a una constante.

Estas configuraciones no tienen un equivalente en los modelos de polvo puro  $\hat{\Omega}^{(\lambda)} = \lambda = 0$ , porque en este caso la curvatura positiva necesariamente implica inversión de expansión y colapso (véase sección VI). Las trayectorias de espacio de fases se muestran en la Figura VIII.1.2.3. Como en los casos anteriores, las curvas emergen desde un atractor pasado PA (*big bang*), se expanden monótonamente y terminan en un atractor futuro FA, pero ahora las trayectorias en el panel (a) están localizadas por encima de la línea  $\hat{\Omega}^{(u)} + \hat{\Omega}^{(\lambda)} = 1$  característica de la curvatura espacial cero. Las curvas ahora tienden al punto de silla SP4, cuya proyección coincide con la proyección de PA en el panel  $\{\hat{\Omega}^{(u)}, \hat{\Omega}^{(\lambda)}\}$  (panel (a)).



**Figura VIII.1.2.4.** Modelo de expansión abierto con curvatura positiva y topología abierta. Las condiciones iniciales están dadas por (72)-(74). Consultar el texto para explicación más amplia.

### VIII.1.3 Curvatura espacial positiva con topología cerrada o de agujero de gusano

Como en los casos anteriores, estas configuraciones no tienen equivalentes en el caso de polvo puro:

- *Topología cerrada:* Tomando  $f(r) = \sin r$ , de manera que el rango radial esté restringido por  $0 \leq r \leq \pi$ , con dos centros de simetría en  $r = 0, \pi$  y  $r = \pi / 2$  marcando el “ecuador” de la 3-esfera  $S^3$  donde  $R$  toma su máximo valor y  $R'(\pi / 2) = 0$  (véase sección VI).

- *Agujero de gusano (Wormhole)*: Utilizando  $f(r) = \sec r$ , de manera que  $-\pi/2 < r < \pi/2$  y  $R$  no tiene ceros (no hay centros de simetría), mientras  $R \rightarrow \infty$  cuando  $r \rightarrow \pm \pi/2$ . Esta configuración conduce a rebanadas de espacio homeomórficas a  $S^2 \times R$ , donde el locus  $r = 0$  puede identificarse con la “garganta” del agujero de gusano ( $R$  toma su valor mínimo y  $R'(0) = 0$ ).

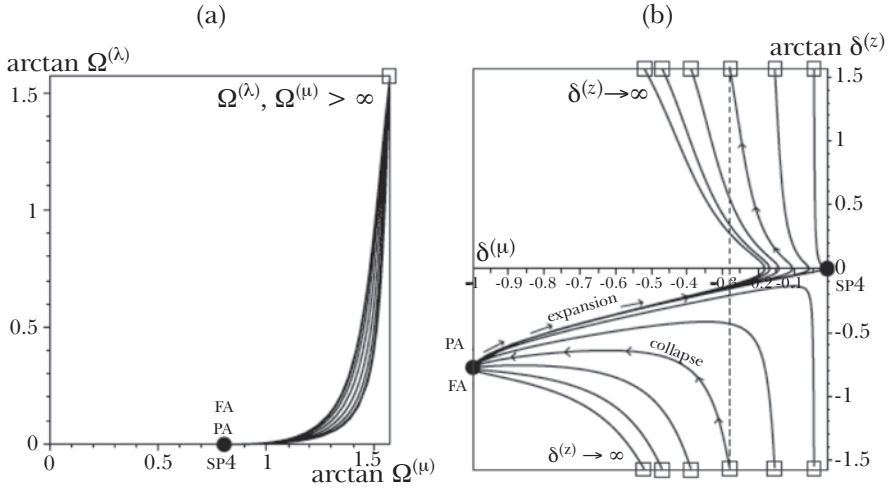
No mostramos las trayectorias de espacio de fases para estos casos porque para cualquier conjunto de condiciones iniciales, éstas son similares a las de los modelos abiertos de la Figura VIII.1.2.3, aunque es necesario recalcar algunas características distintivas. En el caso de la topología cerrada las trayectorias de los dos centros en la proyección homogénea (panel (a)) son idénticas entre sí y equivalentes a la curva  $r = 0$  en el panel (a) de la Figura VIII.1.2.3. Las trayectorias en el rango  $\pi/2 < r < \pi$  se corresponden únicamente a aquellas en el rango  $0 < r < \pi/2$ . La curva para  $r = \pi/2$  (el “ecuador” de  $S^3$ ) es análoga a la curva cercana a la línea  $\hat{\Omega}^{(\mu)} + \hat{\Omega}^{(\lambda)} = 1$  correspondiente al radio infinito en los modelos abiertos ( $r \rightarrow \pi/2$  en ambos paneles de la Figura VIII.1.2.3). En el caso del agujero de gusano las curvas para  $r < 0$  y  $r > 0$  (“izquierda” y “derecha” de la garganta  $r = 0$ ) son idénticas y pueden ser mapeadas entre sí, mientras la curva del espacio de fases de la garganta es análogo a aquel del centro  $r = 0$  en ambos paneles de la Figura VIII.1.2.3.

## VIII.2 Configuración en la cual la expansión es revertida o interrumpida para todas las capas de polvo

Estas configuraciones se obtienen al escoger las condiciones iniciales que se cumplen en (58), de manera tal que la curva resultante  $[\hat{\Omega}^{(\mu)}(r), \hat{\Omega}^{(\lambda)}(r)]$  está totalmente contenida en el régimen “dominado por  $\mu$ ” o el “dominado por  $\lambda$ ” (áreas sombreadas en la Figura V.1b).

### VIII.2.1 Modelos que colapsan: patrón cinemático (ii)

Si se selecciona  $\mu_{qi} \approx \kappa_{qi} > 1$  con  $\lambda \mu_{qi}, \kappa_{qi}$  para todo  $r$ , se obtienen configuraciones  $\Lambda$ -LTB fuertemente dominadas por  $\mu$  que son similares al caso del polvo puro  $\Lambda = 0$  con curvatura positiva en las cuales las capas se expanden desde una curvatura inicial singular (*big bang*), rebotan y colapsan a una segunda singularidad (*big crunch*). Nótese que, en general, ninguna singularidad coincide con una hipersuperficie simultánea. Las rebanadas de espacio pueden tener topologías cerradas, abiertas o de agujero de gusano. Un ejemplo de condiciones iniciales que conlle-



**Figura VIII.2.5.** Todas las capas colapsan. Esta configuración tiene rebanadas de espacio con topología cerrada y se contienen a partir de las condiciones iniciales (75)–(77). Consultar el texto para explicación más ampliada de la evolución en la evolución del espacio de fases.

van a este tipo de evolución se obtiene al seleccionar  $f = \sin r$  (topología cerrada) junto con

$$\mu_{qi} = m_{10} + \frac{m_{11}}{1 + f^2(r)}, \quad (m_{10} = 0.3, m_{11} = 2.0), \quad (75)$$

$$\kappa_{qi} = k_{10} + \frac{k_{11}}{1 + f^2(r)}, \quad (k_{10} = 0.01, k_{11} = 1.98), \quad (76)$$

$$\lambda = 0.01 \quad (77)$$

donde  $\kappa_{qi}$  cumple con (60) y (61). Como las capas de polvo pueden sólo evolucionar en el rango  $0 < L < L_1$  (véase sección V), la expansión inicial se invierte a un estado colapsante en un “punto de retorno”  $L = L_1(r)$  que es en general diferente para cada  $r$ . En este punto de retorno tenemos  $z_q = 0$  y, por tanto,  $\hat{\Omega}^{(\mu)}$ ,  $\hat{\Omega}^{(\lambda)}$  y  $\delta^{(z)}$  divergen. Como consecuencia, no es posible encontrar el estado en que las trayectorias colapsan (que está más allá de este punto) con las soluciones numéricas del sistema dinámico (30)–(33). Las trayectorias completas en este caso se encuentran a partir de las soluciones numéricas del sistema (11)–(14).

Las trayectorias en el espacio de las fases se muestran en la Figura VIII.2.5, donde se resalta marcada con flechas en el panel (b) una trayectoria representativa (para facilitar la explicación). La proyección homogénea de la evolución del espacio fase (panel (a)) es muy simple: las curvas de expansión emergen desde el atractor pasado PA, alcanzan el punto de retorno donde  $z_q \rightarrow 0$  y  $\hat{\Omega}^{(u)}, \hat{\Omega}^{(\lambda)}$  divergen (marcado por un cuadrado), la expansión se invierte a colapso y las curvas regresan siguiendo las mismas trayectorias (pero en dirección inversa) hacia PA, que es ahora un atractor futuro en el estado de colapso. En la proyección no homogénea (panel (b), véase curva representativa), las curvas tienen caminos diferentes en los estados de expansión y colapso. Todas las curvas inicialmente se expanden y emergen de un PA, se acercan al punto de silla SP4 y se dirigen hacia los “puntos de retorno” marcados por los cuadrados superiores donde  $z_q \rightarrow 0$  y  $\delta^{(z)} \rightarrow -\infty$  para la variable  $\delta^{(u)}$ . En el estado de colapso (véase curva representativa) las curvas comienzan en los puntos de retorno (cuadrados inferiores) con  $\delta^{(z)} \rightarrow -\infty$  para la variable  $\delta^{(u)}$  y evolucionan hacia PA, el cual es ahora el atractor futuro. El hecho de que las trayectorias de expansión y colapso sean las mismas (con  $\xi \rightarrow -\xi$ ) en la proyección homogénea (panel (a)), pero totalmente diferente en el caso de la proyección inhomogénea (panel (b)) es una característica importante, inherente a la inhomogeneidad de los modelos.

Se pueden obtener el mismo tipo de configuraciones de colapso con rebanadas que tengan topologías abiertas y de agujero de gusano utilizando similares condiciones iniciales pero con  $f = \tan r, \sec r$ .

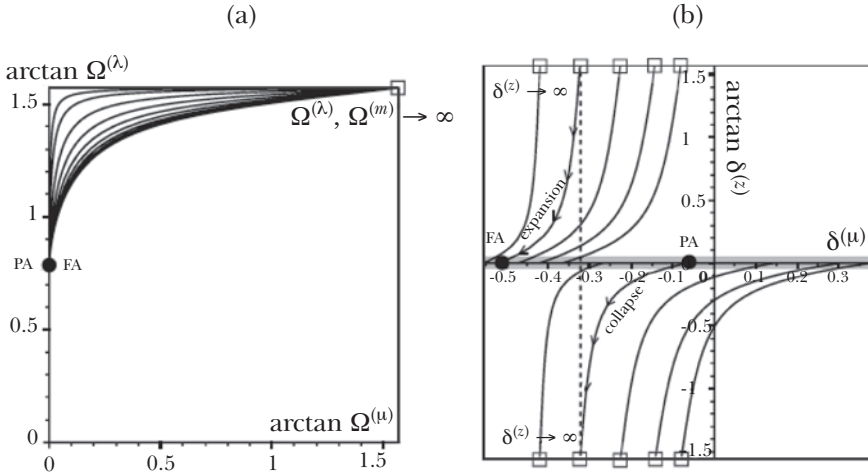
### VIII.2.2 Modelos con retorno sin colapso: patrón cinemático (iii)

Configuraciones fuertemente dominadas por  $\lambda$  se obtienen cuando  $\mu_{qi} \gg \kappa_{qi}$  y  $\mu_{qi} \gg \lambda$  para todo  $r$ . Todas las capas comienzan con un colapso desde el infinito  $L$  pero rebotan y evitan el colapso a una singularidad, estas configuraciones no tienen equivalentes en el caso del polvo puro ( $\lambda = \hat{\Omega}^{(\lambda)} = 0$ ). Las rebanadas de espacio pueden tener topologías cerradas, abiertas o de agujero de gusano. Un ejemplo de las condiciones iniciales que conllevan a este tipo de evolución se obtiene al seleccionar  $f = \tan r$  (topología abierta) y

$$\mu_{qi} = m_{10} + \frac{m_{11}}{1 + f^2(r)}, \quad (m_{10} = 0.01, m_{11} = 1.5), \quad (78)$$

$$\kappa_{qi} = k_{10} + \frac{k_{11}}{1 + f^2(r)}, \quad (k_{10} = 10.0, k_{11} = 30.0), \quad (79)$$

$$\lambda = 31.0 \quad (80)$$



**Figura VIII.2.2.6.** Todas las capas colapsan. Esta configuración se obtiene a partir de las condiciones iniciales (79)-(81).

Las trayectorias de espacio fase se muestran en la Figura VIII.2.2.6. Como las capas de polvo sólo pueden evolucionar en el rango  $L > L_2$  (véase sección V), el colapso inicial desde el infinito retorna a una expansión cuando  $L = L_2(r)$ , que es generalmente diferente para cada  $r$ . Como en el caso de colapso, en este punto de rebote tenemos  $z_q = 0$ , por tanto,  $\hat{\Omega}^{(\mu)}, \hat{\Omega}^{(\lambda)} = \delta^{(z)}$  divergen; por tanto, el gráfico de la Figura VIII.2.5 muestra la arctan de estas variables. Como en el caso anterior, se utilizó el sistema (11)-(14) para obtener el estado de colapso de la evolución.

La curva de los atractores pasados PA se obtiene de (43), y es la misma que la curva de los atractores futuros para la configuración de expansión perpetua de las Figuras VIII.1.1.2, VIII.1.2.3 y VIII.1.2.4.

Cada trayectoria de espacio fase comienza de un punto (marcado como PA) en esta curva de atractores pasados y termina en un punto diferente de la misma curva (de atractores futuros, marcados como FA). Estos atractores están proyectados en el mismo punto de la proyección homogénea (panel (a)) y en dos puntos distintos en la línea gris gruesa de la proyección homogénea (panel (b)). En el panel (a) las trayectorias de colapso inicialmente comienzan en PA, evolucionan hacia el punto de retorno cuando  $z_q \rightarrow 0$  y  $\hat{\Omega}^{(\mu)}, \hat{\Omega}^{(\lambda)} \rightarrow \infty$  (cuadrados) y regresan al mismo punto (ahora un atractor futuro FA) siguiendo la misma trayectoria. En el panel (b) (véase trayectoria representativa marcada con flechas), las trayectorias comienzan en un punto PA de la curva de atractores (línea gris gruesa) y evolucionan hacia abajo hacia los puntos de retorno donde  $z_q \rightarrow 0$  y  $\delta^{(z)} \rightarrow \infty$  (cuadrados

inferiores) y variable  $\delta^{(u)}$ . Las trayectorias de fase de la expansión comienzan en un punto de rebote  $z_q \rightarrow 0$  y  $\delta^{(z)} \rightarrow \infty$  (cuadrados superiores) y van descendiendo hacia el punto FA de la curva de atractores. Como en el caso anterior (capas colapsantes), también se tiene un tiempo de reflexión simétrica  $\xi \rightarrow -\xi$  de la evolución de colapso/expansión de las variables homogéneas, pero no en las variables inhomogéneas.

Las configuraciones de rebote que evitan el colapso pueden construirse con rebanadas de espacio que tengan topologías cerradas y de agujero de gusano utilizando condiciones iniciales similares, pero con  $f = \sin r$ ,  $\sec r$  y  $\kappa_{qi}$  cumple con las condiciones (60) y (61). Sus trayectorias de espacio de fases son cualitativamente similares a aquellas de la Figura V.1.

### VIII.2.3 Modelos de “loitering”: patrón cinemático (iv)

Las condiciones iniciales para estos modelos cumplen con  $\kappa_{qi} > 0$  y los valores de las funciones iniciales están restringidos de manera que se cumple  $Q_{\min} = 0$ . Considerando la  $Q_{\min}$  dada en (57), examinemos el siguiente caso específico:

$$\kappa_{qi} = k_{10} + \frac{k_{11}}{1 + f^4(r)}, \quad (k_{10} = 2.1, k_{11} = 10.5), \quad (81)$$

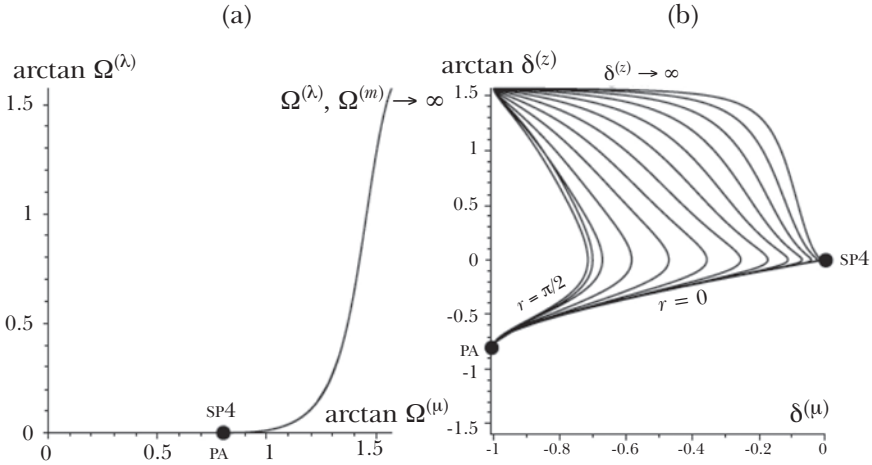
$$\lambda = 0.1 \quad (82)$$

$$\mu_{qi} = \frac{\kappa_{qi}^{3/2}}{3\sqrt{3\lambda}} \quad (83)$$

donde  $f = \tan r$ . La Figura VIII.2.3.7 muestra los diagramas de espacio fase. Todas las trayectorias de espacio de fases en ambos paneles (a) y (b) comienzan en el atractor pasado PA (*big bang*) y evolucionan hacia el estado estático caracterizado por  $z_q \rightarrow 0$ , que implica  $\hat{\Omega}^{(u)}$ ,  $\hat{\Omega}^{(\lambda)}$ ,  $\delta^{(z)}$ . Es evidente que todas las curvas se acercan al punto de silla SP4 en la proyección  $\{\delta^{(u)}, \delta^{(z)}\}$  (panel (b)).

La inestabilidad del estado estático en estas configuraciones viene dado por la sensibilidad extrema de las condiciones iniciales exactas que cumplen con  $Q_{\min} = 0$  (véase sección V y Figura V.1) y, por tanto, los gráficos son muy sensibles a errores numéricos en la integración de (30)-(33). Como consecuencia, las capas pueden colapsar o expandirse (dependiendo de las condiciones iniciales) si la integración numérica se realiza en un rango suficientemente largo de valores de

$\tau$  o  $i$ . Si las capas recolapsan o se expanden, sus trayectorias de espacio de fases se determinan por medio del sistema (11)-(14), resultando en trayectorias de espa-

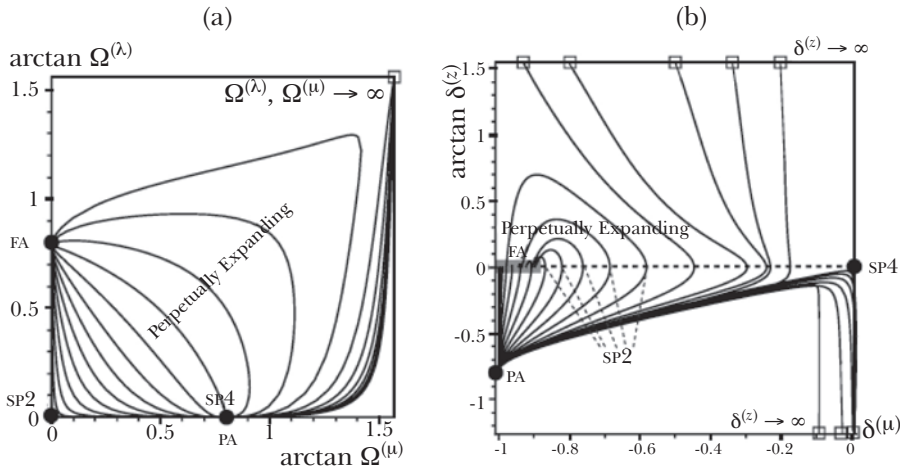


**Figura VIII.2.3.7.** Capas que presentan “Loitering”. Las condiciones iniciales para esta configuración son (82)–(84) y cumplen  $Q_{\min} = 0$ .

cio de fases similares a las curvas que presentan colapso en la Figura VIII.2.5 o a las que se expanden en la Figura VIII.2.2.6.

### VIII.3 Configuraciones con patrones cinemáticos mixtos: patrones cinemáticos (v)

Como los valores de las funciones iniciales dependen de  $r$ , se pueden dar unas condiciones iniciales de manera que diferentes patrones cinemáticos ocurran en la misma configuración. En términos prácticos tales condiciones iniciales deben cumplir las condiciones de (59) y de ellas se obtienen curvas  $[\hat{\Omega}_i^{(\mu)}(r), \hat{\Omega}_i^{(\lambda)}(r)]$  que atraviesan las áreas sombreadas y no sombreadas de la Figura V.1b. Nótese que estas configuraciones de patrones “mixtos” pueden ocurrir fácilmente en un único modelo inhomogéneo, pero no podrían existir en un único modelo homogéneo.



**Figura VIII.3.8.** Escenario de formación de estructura. Las condiciones iniciales para esta configuración son (85)–(87).

### VIII.3.1 Formación de estructura

Configuraciones interesantes que pueden asociarse con modelos sencillos de “formación de estructura” se obtienen a partir de condiciones iniciales en donde (59) se cumple (región sombreada dominada por  $\mu$  en la Figura V.1b) en un rango radial que contiene un centro de simetría ( $0 \leq r \leq r_b$ ) y deja de cumplirse (región no sombreada en la Figura V.1b) para todo  $r > r_b$ . Ésta corresponde a una región esférica interna en colapso local rodeada por un “fondo” en expansión. Mientras este escenario de formación de estructura sólo ocurre con soluciones de polvo puro, en este caso la frontera  $r > r_b$  entre la región en colapso y en fondo en expansión corresponde al cono de luz donde la curvatura espacial pasa de positiva a negativa o a cero. Sin embargo, si  $\lambda > 0$  es posible que este escenario también ocurra cuando todas las capas tengan curvatura espacial positiva, porque un fondo en expansión perpetua con curvatura positiva es perfectamente posible.

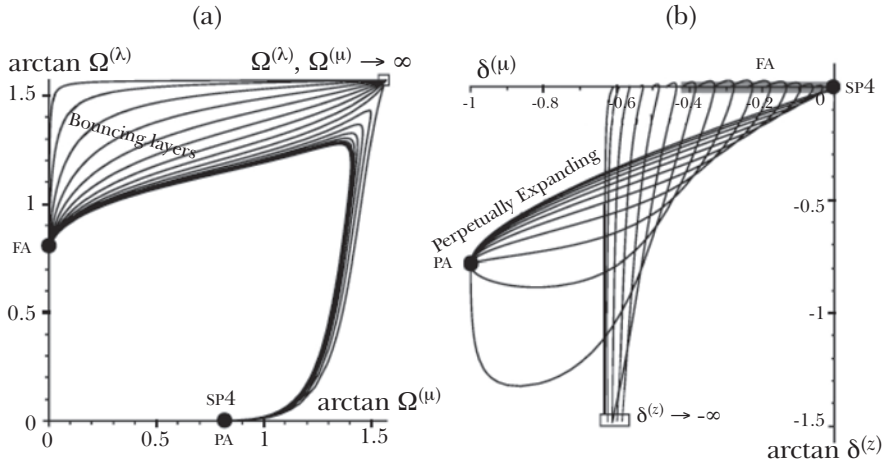
Un ejemplo de condiciones iniciales que llevan a un modelo sencillo de formación de estructura se obtiene al seleccionar  $f(r) = \tan r$  (topología abierta) con

$$\mu_{qi} = m_{10} + \frac{m_{11} - m_{10}}{1 + f^3(r)}, \quad (m_{10} = 0.0, m_{11} = 20.1), \quad (84)$$

$$\kappa_{qi} = k_{10} + \frac{k_{11} - k_{10}}{1 + f^4(r)}, \quad (k_{10} = -4.1, k_{11} = 35.5), \quad (85)$$

$$\lambda = 0.75 \tag{86}$$

Las trayectorias espacio fase se muestran en la Figura VIII.3.8. Todas las curvas comienzan su evolución en el atractor pasado PA marcado por la singularidad inicial (*big bang*). Sin embargo, las capas “interiores” alrededor del centro ( $r = 0$ ) y las capas “exteriores”  $r > r_b$  tienen una evolución en el espacio de fases diferente.



**Figura VIII.3.1.9.** Expansión mixta y retorno. Las condiciones iniciales para esta configuración son (88)-(90).

Como las capas de polvo con un radio  $r > r_b$  se expanden perpetuamente, sus evoluciones son análogas a las de las configuraciones de expansión perpetua mostradas en las Figuras VIII.1.1.2, VIII.1.2.3 o VIII.1.2.4 (dependiendo de la curvatura espacial y de las asíntotas de  $\mu_{qi}$  y  $\kappa_{qi}$ ). En el ejemplo particular se considera que estas trayectorias (véase Figura 8) tienen formas análogas a las representadas en estas tres figuras. Desde un atractor pasado PA se acercan al punto silla SP2 y terminan en la línea del atractor de futuro FA (cuya proyección inhomogénea se representa como una línea gris gruesa en el panel (b)).

Las capas interiores, por otro lado, evolucionan de forma similar a las trayectorias de configuraciones de colapso representadas en la Figura VIII.2.5, inicialmente se expanden alejándose del atractor pasado PA, acercándose al punto de silla SP4 y continuando hacia los puntos de retorno de manera que  $z_q \rightarrow 0$ , con  $\hat{\Omega}^{(\mu)}$ ,  $\hat{\Omega}^{(\lambda)}$ , y  $\delta^{(z)}$  divergen (los puntos de retorno están representados como peque-

ños cuadrados en ambos paneles). La diferencia principal con el caso de colapso “puro” de la Figura VIII.2.5 es que las trayectorias de capas internas más cercanas a las capas externas, que se expanden perpetuamente, alcanzan sus puntos de retorno (véase panel (b)) cuando  $\delta^{(z)} \rightarrow \infty$ , mientras que las internas cercanas al centro lo hacen cuando  $\delta^{(z)} \rightarrow -\infty$ . El periodo de colapso de estas capas no se muestra en la Figura 8, pero se puede obtener a partir del sistema (11)-(14) y es análogo al mostrado en V. Las trayectorias que presentan colapso en el panel (a) son simplemente las que se expanden considerando una inversión del tiempo, mientras que en el panel (b), comienzan en un punto de retorno hacia un atractor futuro que exactamente coincide con PA (la singularidad del colapso) pero siguiendo un camino diferente al de las curvas en expansión.

### VIII.3.2 Modelos de retorno/expansión

Otra configuración cinemática mixta es posible con las siguientes características: las capas de polvo alrededor del centro se expanden desde una singularidad inicial, mientras que las capas exteriores colapsan desde el infinito hasta un punto de retorno, a partir del cual se expanden. En este caso las condiciones iniciales necesitan cumplir  $\hat{\Omega}_i^{(u)} \approx \hat{\Omega}_i^{(\lambda)}$  cerca del centro y  $\hat{\Omega}_i^{(\lambda)} \neq \hat{\Omega}_i^{(u)}$  cuando  $r$  se incrementa, de manera que la curva  $[\hat{\Omega}_i^{(u)}(r), \hat{\Omega}_i^{(\lambda)}(r)]$  atraviesa desde el área no sombreada de la Figura V.1b hasta la región sombreada dominada por  $\lambda$ . Las condiciones iniciales que se requieren para que se presente este patrón de evolución son:

$$\mu_{qi} = m_{10} + \frac{m_{11} - m_{10}}{1 + f^2(r)}, \quad (m_{10} = 0.01, m_{11} = 15.0), \quad (87)$$

$$\kappa_{qi} = k_{10} + \frac{k_{11} - k_{10}}{1 + f^2(r)}, \quad (k_{10} = 10.0, k_{11} = 30.0), \quad (88)$$

$$\lambda = 10.0 \quad (89)$$

con  $f = \tan r$ . Las trayectorias obtenidas se representan en los diagramas del espacio de fases de la Figura VIII.3.1.9. Las capas que se expanden perpetuamente desde una singularidad inicial se comportan como las curvas de la Figura VIII.1.2.4: comienzan en un atractor pasado PA, se acercan a un punto silla SP4 y terminan en el atractor futuro FA. Las curvas que presentan retorno son análogas a las representadas en la Figura VIII.2.2.6: tienen su origen en un atractor pasado PA y evolucionan hacia el punto de retorno  $z_q \rightarrow 0$  (o  $\hat{\Omega}_i^{(u)}, \hat{\Omega}_i^{(\lambda)} \rightarrow \infty, \delta^{(z)} \rightarrow -\infty$  con la variable  $\delta^{(u)}$ ) y se expanden de regreso a PA (que ahora es atractor futuro). En

la Figura VIII.3.1.9 sólo se muestra el estado de colapso de esta evolución, pues las curvas del estado de expansión son análogas a las curvas de expansión de la Figura VIII.2.2.6.

## CONCLUSIONES

Hemos realizado un estudio exhaustivo y detallado de los sistemas dinámicos de los espacio-tiempos  $\Lambda$ -LTB mediante las variables covariantes cuasi-locales (QL), generalizando estudios previos de los sistemas dinámicos de modelos LTB con  $\Lambda = 0$  (Sussman, 2008). Como las trayectorias en el espacio de las fases del sistema (30)-(33) son 4-dimensionales, hemos examinado y representado gráficamente las trayectorias del espacio de las fases en dos subconjuntos bidimensionales: el subespacio “homogéneo” y el subespacio “inhomogéneo”,  $H_0$  e  $I$ , definidos por las proyecciones de (64) y (65) y representados en los paneles (a) y (b) de las Figuras VIII.1.1.2 a VIII.3.1.9, respectivamente. La descomposición de las trayectorias en  $H_0$  e  $I$  nos proporciona una conexión teórica entre el estudio de sistemas dinámicos  $\Lambda$ -LTB y el formalismo de perturbaciones presentado en Sussman (2009, 2010a). En ese formalismo, las fluctuaciones  $\{\delta^{(u)}, \delta^{(z)}\}$  (que parametrizan  $H_0$ ) son caracterizadas rigurosamente como perturbaciones no lineales covariantes e invariantes gauge de un fondo FLRW definido por escalares cuasi-locales, que satisfacen la dinámica FLRW y, por tanto, se corresponden con las proyecciones homogéneas del subespacio  $H_0$  parametrizadas por  $\{\hat{\Omega}^{(u)}, \hat{\Omega}^{(\lambda)}\}$ .

En los modelos LTB con  $\Lambda = 0$  todas las configuraciones, que muestran rebanadas espaciales  ${}^3T[t]$  presentan topología cerrada ( $S^3$ ) o de tipo agujero de gusano ( $S^3 \times R$  o  $S^3 \times S^1$ ). Las condiciones de regularidad (60)-(61) implican que la curvatura espacial debe de ser estrictamente positiva y, por tanto, no se puede dar expansión perpetua (Sussman, 2010a; Matravers y Humphreys, 2001; Hayward, 1998). Sin embargo, los modelos con  $\Lambda > 0$  no presentan esta restricción. Aunque (60)-(61) aun implican necesariamente curvatura espacial positiva en estas topologías, la curvatura positiva no evita que exista expansión perpetua (consúltese la sección VI). Entonces, modelos  $\Lambda$ -LTB regulares con expansión perpetua con topología cerrada o de tipo agujero de gusano son perfectamente posibles.

Las proyecciones homogéneas en los paneles (a) de las Figuras VIII.1.1.2 a VIII.3.1.9 muestran que las trayectorias de los espacios de fase son cualitativamente análogas a aquellas de los modelos FLRW polvo- $\Lambda$ . Sin embargo, cada configuración de espacio tiempo FLRW se corresponde con una (y sólo una) trayectoria de espacio de fases en un diagrama  $\{\hat{\Omega}^{(u)}, \hat{\Omega}^{(\lambda)}\}$  (porque cada configuración espacio tiempo está determinada por las condiciones iniciales dadas por las constantes  $\hat{\Omega}_i^{(u)}, \hat{\Omega}_i^{(\lambda)}$ ). En cambio, como las condiciones iniciales para una sola configuración

$\Lambda$ -LTB dependen de  $r$ , todas las curvas mostradas en los diagramas de los paneles (a) de cada una de estas figuras corresponde a una sola configuración espacio tiempo. En otras palabras, en términos de la proyección (64), cada configuración  $\Lambda$ -LTB formalmente equivale a una familia superpuesta de un parámetro de modelos FLRW de polvo- $\Lambda$ .

En todas las configuraciones  $\Lambda$ -LTB (excepto en aquellas como la representada en la Figura VIII.2.2.6) las trayectorias de espacio de fases inician en una misma fuente o atractor pasado (PA), definida en (PA) y están asociadas a una singularidad inicial. Este punto crítico puede proyectarse en el atractor pasado del subespacio homogéneo e identificarse con una cosmología de polvo FLRW espacialmente plana  $\hat{\Omega}^{(u)} = 1, \hat{\Omega}^{(\lambda)} = 0$ . Dado que  $\delta^{(u)} = \delta^{(z)} = -1$ , este punto crítico coincide exactamente con el atractor pasado de las soluciones de polvo LTB con  $\Lambda = 0$  y modelos espacialmente planos con  $\Lambda > 0$ , indicando así que los efectos de  $\Lambda > 0$  y de la curvatura espacial son despreciables en los estados tempranos de la evolución cercanos al *big bang*. El atractor pasado (PA) puede también identificarse con condiciones auto-similares (Sussman, 2008) que es consistente con el hecho de que ambos espacios tiempos de polvo LTB y FLRW espacialmente planos con  $\Lambda = 0$  son compatibles con un vector Killing homotético. Este atractor pasado (42) se convierte en el atractor futuro o sumidero para trayectorias de espacio de fases de capas que colapsan a una singularidad de la curvatura (*big crunch*). Esto se aplica entonces tanto para un colapso que coincide con la inversión trivial del tiempo de una expansión perpetua (Figuras VIII.1.1.2-VIII.1.2.4), como para configuraciones donde todas las capas se expanden y colapsan (Figura VIII.2.5), o que presentan dinámica mixta en las cuales algunas capas se expanden perpetuamente y algunas colapsan (Figura VIII.3.8).

Las trayectorias de espacio fase de capas que se expanden perpetuamente ( $z_q > 0$ ) terminan en el mismo atractor (sumidero) (FA), definido en (43) y se representan en las Figuras VIII.1.1.2-VIII.1.2.4 y en las trayectorias que se expanden perpetuamente de las Figuras VIII.2.2.6, VIII.3.8 y VIII.3.1.9. Este punto crítico se puede proyectar en el atractor futuro del subespacio homogéneo  $\hat{\Omega}^{(u)} = \hat{\Omega}^{(\lambda)} = 1$  y también es el atractor futuro para configuraciones espacialmente planas. Por tanto, puede identificarse con el espacio tiempo De Sitter y condiciones espacialmente planas, indicando que los efectos de  $\Lambda$  son dominantes en el futuro de estados tardíos de la evolución de configuraciones  $\Lambda$ -LTB en las cuales los conos de luz de las capas de polvo en expansión son inextensibles. Para capas de polvo que colapsan desde el infinito ( $L \rightarrow \infty$ ), las trayectorias del espacio de fases emergen desde un atractor pasado que exactamente coincide con el atractor futuro (FA).

Observando la proyección inhomogénea (65) (paneles (b)), la evolución hacia (FA) o las trayectorias del espacio de fases de las capas que se expanden per-

petuamente, se obtiene información adicional no disponible en la proyección (64). Las trayectorias de capas con curvatura espacial negativa (Figuras VIII.1.1.2-VIII.1.2.3) se acercan al punto silla (SP2) (Minkowski), mientras que las capas con curvatura positiva se acercan al punto silla (SP4) (Figuras VIII.1.2.4, VIII.2.5, VIII.3.8 y VIII.3.1.9) asociadas a condiciones espacialmente planas. La evolución hacia (SP2) claramente indica un estado de densidad baja intermedio, y también ocurre en trayectorias del espacio de fases FLRW (nótese que (SP2) es proyectado por (64) en un punto de silla único de este espacio de fases). Sin embargo, la evolución hacia (SP4) no tiene equivalente en las trayectorias de las capas de espacios de fase FLRW que evolucionan a un atractor de futuro.

En configuraciones que contienen capas de polvo que retornan y colapsan ( $z_q$  cambia el signo: Figuras VIII.2.5-VIII.3.1.9) el sistema (11)-(14) se utiliza para obtener las trayectorias completas de espacio de fases. En todos los casos, las trayectorias de espacio de fases en la proyección homogénea comienzan en el atractor pasado (PA), alcanzan el infinito a valores  $\hat{\Omega}^{(u)}, \hat{\Omega}^{(\lambda)} \rightarrow \infty$  cuando  $z_q = 0$  y retornan por la misma trayectoria al mismo atractor (que ahora es un atractor futuro). Sin embargo, en la proyección inhomogénea (paneles (b) de las Figuras VIII.2.2.5 y VIII.2.2.6) las curvas no retornan al atractor por la misma trayectoria: si el patrón es de expansión/colapso (Figura VIII.2.5) alcanzan  $\delta^{(z)} \rightarrow \infty$  cuando  $z_q = 0$  y retornan al atractor desde  $\delta^{(z)} \rightarrow -\infty$ . Si el patrón es de colapso/retorno (Figura VIII.2.2.6), las trayectorias alcanzan  $\delta^{(z)} \rightarrow -\infty$  y retornan al atractor desde  $\delta^{(z)} \rightarrow \infty$ . El mismo efecto ocurriría en las configuraciones de patrones mixtos mostradas en las Figuras VIII.3.8 y VIII.3.1.9, pero no se incluyeron en los gráficos, pues la representación de las curvas completas hubiera generado un patrón confuso. Esta diferencia en el comportamiento de las curvas que emergen/retornan desde/hacia un atractor es un efecto inherente en la inhomogeneidad de las soluciones LTB.

Los modelos de retorno y “*loitering*” merecen mención aparte pues no tienen equivalencia en el caso  $\Lambda = 0$ . En los modelos de retorno (Figura VIII.2.1.6), las capas inicialmente colapsan desde el infinito, por tanto sus atractores pasados coinciden con el atractor futuro de las configuraciones que se expanden perpetuamente (Figuras VIII.1.1.2-VIII.1.2.4). Como se mencionó anteriormente, las curvas retornan al atractor siguiendo una trayectoria diferente. En la configuración “*loitering*” que se examinó (Figura VIII.2.3.7), las trayectorias del espacio de fases emergen desde el mismo atractor pasado (FA) como todas las curvas para capas que se expanden desde una singularidad, pero como las capas se vuelven asintóticamente estáticas  $z_q \rightarrow 0$  las trayectorias evolucionan hacia  $\hat{\Omega}^{(u)}, \hat{\Omega}^{(\lambda)}, \delta^{(z)} \rightarrow \infty$ . En la proyección homogénea (panel (a) de la Figura VIII.2.3.7) las trayectorias son indistinguibles de aquellas que alcanzan estos valores divergentes en confi-

guraciones de expansión/colapso (Figura VIII.2.5.), pero en la proyección inhomogénea (panel (b) de la Figura VIII.2.3.7) las curvas tienen un comportamiento diferente: alcanzan  $\delta^{(z)} \rightarrow \infty$  por diferentes caminos y convergen a  $\infty$ . Este efecto es una característica inherente de inhomogeneidad y no ocurre en diagramas de espacio fase FLRW de modelos “loitering”.

La dinámica de modelos  $\Lambda$ -LTB no es muy diferente (hablando cualitativamente) de la de los modelos FLRW homogéneos que se generalizan. Esencialmente, el atractor pasado es un estado cercano a FLRW espacialmente plano y de autosimilaridad, mientras que el atractor futuro es estado de De Sitter dominado por  $\Lambda$ . Los dos efectos más significativos debidos a la inhomogeneidad de los modelos  $\Lambda$ -LTB son:

- a) La posibilidad de acomodar patrones cinemáticos totalmente diferentes en la misma configuración espacio tiempo (véanse figuras VIII.3.8 y VIII.3.1.9). Sin embargo, aun cuando los patrones cinemáticos de todas las capas sean cualitativamente similares, cada espacio tiempo  $\Lambda$ -LTB único puede entenderse como un modelo inhomogéneo compuesto por algún tipo de superposición de modelos FLRW (pues la trayectoria espacio fase de cada cono de luz comóvil puede ser proyectada por (65) en la trayectoria espacio fase de un modelo FLRW único).
- b) La asimetría en las evoluciones desde un atractor pasado y un punto de retorno y desde ese punto hacia el mismo atractor, que se convierte en un atractor futuro (véanse paneles (b) de las Figuras VIII.2.5 y VIII.2.2.6).

Sin embargo, a pesar de las simplificaciones en que se incurre al asumir una simetría esférica y movimiento geodésico, el estudio del presente sistema dinámico claramente muestra que aun niveles idealizados de inhomogeneidad proveen grados de libertad adicionales que pueden ser muy útiles para construir modelos sencillos de inhomogeneidades astrofísicas y cosmológicas.

## REFERENCIAS

- Bruni, M., Dunsbury, P. K. S. y Ellis, G. F. R. (1992). *Astroph. J.* 34-53.  
Carroll, S. M., Press, W. H. y Turner, E. L. (1992). *A R A A.* 499.  
Coley, A. A. (2003). *Dynamical Systems and Cosmology*. Kluwer Academic Publishers.  
——— (1999). *Dynamical Systems in Cosmology in Proceedings of ERE-99*. Preprint arXiv:gr-qc/9910074v1.

- Copeland, E. J. *et al.* (2003). *Preprint* hep-th/0603057.
- Deshingkar, S. S., Chamorro, A., Jhingan, S. y Joshi, P. S. (2001). *Phys Rev. D* 124005.
- Ellis, G. F. R. y Bruni, M. (1989). *Phys. Rev. D* 1804.
- y van Elst, H. (1998). *Cosmological Models* (Cargèse Lectures 1998). *Preprint* arXiv gr-qc/9812046 v4.
- Felten, J. E. y Isaacman, R. (1986). *Rev Mod Phys.* 689.
- Goncalves, S. M. C. V. (2001). *Phys Rev. D* 064017.
- (2001). *Phys Lett. A* 53.
- Hayward, S. A. (1996). *Phys. Rev. D* 531938 (*Preprint* ArXiv gr-qc/9408002).
- (1998). *Class. and Quant. Gravity.* 3147-3162. *Preprint* ArXiv gr-qc/9710089v2.
- Humphreys, N. P., Maartens, R. y Matravers, D. R. (1998). *Regular Spherical Dust Spacetimes. Preprint* gr-qc/9804023v1.
- Ilha, A., Kleber, A. y Lemos, J. P. S. (1999). *J. Math Phys.* 3509.
- Ilha, A. y Lemos, J. P. S. (1997). *Phys Rev. D* 1788.
- Kraśniński, A. (1998). *Inhomogeneous Cosmological Models*. Cambridge University Press.
- y Plebański, J. (2007). *An Introduction to General Relativity and Cosmology*. Cambridge University Press.
- Kottler, F. (1918). *Ann Phys.* 410. Germany.
- Lake, K. (2000). *Phys Rev. D* 027301.
- Matravers, D. R. y Humphreys, N. P. (2001). *Gen. Rel. Grav.* 531-552.
- Nakao, K. (1992). *Gen Rel Gravit.*
- Omer, G. C. (1965). *Proc. Nat. Ac. Sci.* 1.
- Padmanabhan, T. (2003). *Phys. Rep.* 235.
- Peebles, P. J. E. y Ratra, B. (2003). *Rev. Mod. Phys.* 559.
- Sahni, V. (2004). *Lect. Notes Phys.* 141-180 (*Preprint* arXiv:astro-ph/0403324v3).
- Sussman, R. A. (2008). *Class. and Quant. Gravity.* 015012.
- (2009). *Quasi-Local Variables and Scalar Averaging in LTB Dust Models*, in Proceedings of the “Invisible Universe International Conference”, J. M. Alimi (ed).
- (2010a). *A New Approach for Doing Theoretical and Numeric Work with Lemaitre-Tolman-Bondi Dust Models*. *Preprint* arXiv:1001.0904v2 [gr-qc].
- (2010b). *Radial Asymptotics of Lemaitre-Tolman-Bondi Dust Models* (to appear in *Gen. Rel. Grav.*) *Preprint* arXiv:1002.0173 [gr-qc].
- y García-Trujillo, L. (2002). *Class. and Quant. Gravity.* 2897-2925.
- e Izquierdo, G. (2011). *Class. and Quant. Gravity.* 045006. *Preprint* arXiv:1004.0773.

Wainwright, J. y Lim, W. C. (2005). *J Hyperbol Diff Equat.* 437-470. Preprint arXiv:gr-qc/0409082.

——— y Ellis, G. F. R. (1997). *Dynamical Systems in Cosmology*. Cambridge: Cambridge University Press.

Zecca, A. (1991). *Nuovo Cimento*. B 413.

# NEURONAS DE INTEGRACIÓN Y DISPARO\*

*Fernando A. Ongay Larios\*\**, *Humberto Carrillo Calvet\*\*\**,  
*José A. Aguilar Sánchez\*\*\*\**, *Máximo A. Agüero\*\*\*\*\**  
*y Miguel Ángel Mendoza Reyes\*\*\*\*\**

## Resumen

Dentro del universo de las oscilaciones no lineales son de especial interés las de *integración* y *disparo*, características de varios sistemas naturales como los que se conocen como sistemas excitables: son aquellos que tienen un umbral de excitación y que responden a un estímulo siempre y cuando éste rebase el umbral de excitación, tal respuesta se conoce como *disparo*. Cuando el estímulo es sostenido por encima del umbral de excitación, el sistema produce trenes de disparos. Un ejemplo paradigmático de sistemas excitables es la membrana celular, particularmente la de las neuronas, razón por la cual a algunos modelos abstractos que exhiben oscilaciones de este tipo se les conoce como *neuronas de integración y disparo*. En este trabajo se expone la teoría fundamental y parte de los resultados obtenidos por nuestro grupo de

\* Los autores agradecen a la Secretaría de Educación Pública (SEP) por el financiamiento al proyecto FEO1/2012 103.5/12/2126 de redes multidisciplinarias y a la SIEA de la UAEM, al Conacyt por el apoyo importante en las actividades de investigación.

\*\* Doctor en Matemáticas. Profesor de Tiempo Completo, Facultad de Ciencias, Universidad Autónoma del Estado de México, Instituto Literario 100, Toluca 50000, México, (52) (722) 2965556, <fernando.ongay@yahoo.com.mx>.

\*\*\* Doctor en Matemáticas. Facultad de Ciencias, Departamento de Matemáticas, Universidad Nacional Autónoma de México, Circuito exterior, Ciudad Universitaria, Col. Copilco el Bajo, Del. Coyoacán, C.P. 04510, México, DF., <carr@servidor.unam.mx>.

\*\*\*\* Doctor en Física-Matemática. Profesor de Tiempo Completo, Facultad de Ciencias, Universidad Autónoma del Estado de México, <polaresquis@yahoo.com.mx>.

\*\*\*\*\* Doctor en Física-Matemática. Profesor de Tiempo Completo, Facultad de Ciencias, Universidad Autónoma del Estado de México, Instituto Literario 100, Toluca 50000, México, <maaguerog@uaemex.mx>.

\*\*\*\*\* Candidato a Doctor en Matemáticas. Profesor de Tiempo Completo, Universidad de la Ciudad de México, <correos001@yahoo.com.mx>.

investigación interdisciplinario e interinstitucional de la UAEM y de la UNAM, de modelos de neuronas de integración y disparo gobernados por una ecuación diferencial de la forma

$$\frac{dx}{dt} = F(t, x)$$

## OSCILACIONES NO LINEALES

El funcionamiento orgánico de un sistema complejo es posible gracias a la sincronización de sus diferentes componentes o subsistemas. Es así que el estudio de sistemas interactuantes constituye un tema central en cualquier disciplina científica. Sin embargo, el análisis matemático de estos sistemas es un problema difícil; de hecho, el estado actual del conocimiento –científico y matemático– aún dista mucho de poder comprender todos los posibles comportamientos de los sistemas complejos. El problema radica, por un lado, en la gran cantidad de componentes, variables y parámetros que pueden estar interactuando, lo que obliga a analizar simplificaciones del problema; y por otro, a que las interacciones generalmente son no lineales. Un paso fundamental en el estudio de estas cuestiones es analizar el forzamiento de osciladores no lineales, asunto sobre el cual se ha vertido una gran cantidad de trabajo.

*Grosso modo*, podemos afirmar que en la naturaleza solamente hay dos tipos de dinámicas: las que presentan oscilaciones o las más o menos impredecibles, como las que se observan en los llamados sistemas caóticos. De hecho, a ciertas escalas el universo pareciera presentar algún orden regido por leyes precisas (razón por la cual se le llama *Cosmos*). El entendimiento humano ha permitido ir descifrando algunas de estas leyes y (parcialmente) comprender este Cosmos, lo que, a su vez, ha redituado en alcanzar altos niveles en materia de tecnología y salud.

Esto ha sido posible gracias a la capacidad de abstraer y modelar sistemas dinámicos. La modelación de las dinámicas oscilatorias ha sido un reto y una inquietud presente en la mente de los investigadores desde los inicios de la ciencia moderna; ha sido también, en buena medida, el motor del desarrollo del cálculo diferencial y el análisis matemático, entre otras áreas de la matemática.

Ya Kepler, en 1609 y 1619, dio uno de los primeros modelos matemáticos, las famosas *Leyes de Kepler*, que explican de manera satisfactoria el movimiento de los planetas. En el siglo XVII, Newton y Leibniz crearon el cálculo diferencial, lo que hizo posible dar un salto cualitativo en la modelación matemática de las leyes físicas que gobiernan la dinámica del movimiento planetario y otros sistemas físicos. El siglo XIX fue toda una explosión de resultados teóricos y experimentales en física y matemáticas; el afortunado encuentro que tuvieron el análisis y la geometría dio lugar a herramientas muy poderosas, como el análisis de Fourier, las ecuaciones diferenciales ordinarias y parciales, la topología, los grupos de Lie y

muchas otras. Esto permitió el desarrollo de modelos más sofisticados, como las leyes de Maxwell del electromagnetismo, sistemas hamiltonianos y la teoría de la relatividad, por mencionar solamente algunos de los más conspicuos, los cuales han permitido una mejor comprensión del mundo físico.

En la modelación de las oscilaciones esencialmente se presentan dos problemas: por un lado, modelar sistemas que presentan oscilaciones autosostenidas (esto es, autónomas de cualquier estímulo externo; a estos sistemas se les llama *osciladores*), y por otro, determinar lo que sucede cuando estos sistemas son sometidos a perturbaciones o estimulaciones (forzamientos) por parte de otros sistemas, y el acoplamiento de varios osciladores. Un problema típico de esto último es: si un oscilador recibe un estímulo periódico, determinar condiciones para que la respuesta del sistema esté sincronizada con el estímulo.

Sin embargo, la gama de comportamientos oscilatorios que se observan en los sistemas de la vida real es muy grande y algunos son muy complejos. Tenemos, por ejemplo, el tipo de oscilaciones llamadas por Van der Pol de *relajación*, estudiadas por éste y muchos otros autores (1926; Cartwright, 1952) con resultados, a decir verdad, todavía muy modestos. La complejidad de estos sistemas es la razón por la cual, para muchos sistemas importantes, todavía no es posible tener modelos científicamente satisfactorios o que sean realmente útiles, ya que mucha de la modelación hecha ocurre en escenarios muy simplificados –modelos lineales, por ejemplo–, o en los que su aplicabilidad está restringida a casos muy especiales o, por el contrario, por su excesiva complejidad o un alto costo en su uso, los hace prácticamente inmanejables.

Desde el punto de vista del análisis matemático, la complejidad de los modelos de osciladores se debe a que éstos suelen ser no lineales y de dimensión mayor que uno. Comúnmente, los osciladores se modelan en términos de un sistema de ecuaciones diferenciales autónomo de la forma

$$\frac{dx}{dt} = F(x)$$

con  $F: R^n \rightarrow R^n$  como los sistemas unidimensionales no son capaces de presentar oscilaciones autónomas, ya que todas sus soluciones son monótonas, oscilaciones autosostenidas ocurren en sistemas con al menos dos variables de estado interactuando ( $N \geq 2$ ). En la estructura del modelo, el forzamiento se traduce en una dependencia temporal del sistema de ecuaciones diferenciales, de donde un modelo minimal de oscilador forzado tendrá la forma

$$\frac{dx}{dt} = f(t, x, y)$$

$$\frac{dy}{dt} = g(t, x, y)$$

No obstante, existen otras alternativas para modelar osciladores, como discutiremos más adelante.

## I. NEURONAS DE INTEGRACIÓN Y DISPARO

La membrana celular es un sistema excitable que dispara un impulso eléctrico (*potencial de acción*) cuando un pulso breve supraumbral de corriente eléctrica le es aplicado. Si en lugar de un pulso, una corriente eléctrica continua  $I$  es aplicada, la célula dispara trenes periódicos de potenciales de acción; es decir, periódicamente el voltaje a través de la membrana celular se incrementa, alcanza el umbral y vuelve al reposo otra vez. Desde este punto de vista, la neurona actúa como un transductor convirtiendo la intensidad  $I$  de la corriente continua aplicada a una señal periódica con frecuencia característica  $w(I)$ .

El estudio de la respuesta de las células nerviosas a una corriente eléctrica aplicada es un problema familiar en fisiología. Es bien conocido entre los fisiólogos que la frecuencia del tren de oscilaciones aumenta con la intensidad de la corriente  $I$  (en ciertos intervalos de intensidad). Este hecho puede ser claramente entendido con la ayuda de modelos incluso tan sencillos como el de la neurona mecánica, discutido más adelante.

Algunas neuronas (marcapasos neuronales) tienen una actividad periódica espontánea. Éstas exhiben patrones de distribución muy variados, incluyendo ráfagas de oscilaciones, en lugar de trenes de oscilaciones sencillos (Baer *et al.*, 1995).

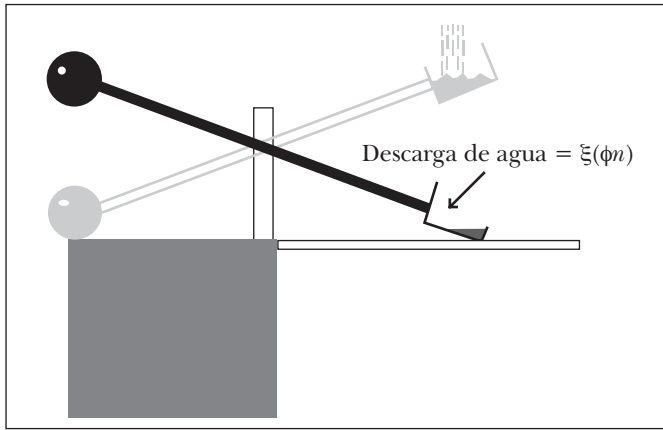
Como la forma (curso temporal) y el tamaño de los potenciales de acción son característicos de cada célula, no es necesario monitorear el voltaje de membrana durante todo el tiempo que dura el evento, sino solamente la sucesión de tiempos  $\{T_n\}$  de los disparos del potencial de acción. Esto es equivalente a imaginar los disparos como procesos instantáneos. Esta idealización de comprimir el tiempo, y la correspondiente reducción del análisis al estudio de sucesiones temporales, es útil para estudiar procesos en los cuales fenómenos estereotípicos ocurren recurrentemente. El concepto ideal de *tiempo de disparo*  $T_n$ , y *sucesión de disparo*  $\{T_n\}$ , emergen de esta idea de compresión temporal. A un sistema excitable junto con la compresión temporal ideal es a lo que llamamos *neurona u oscilador de integración y disparo*.

### a) Neurona mecánica

Como un paradigma de una neurona de integración y disparo consideraremos un sistema mecánico simple, al que llamaremos la neurona mecánica (Mendoza

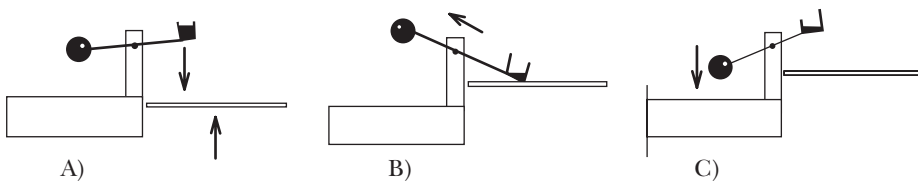
*et al.*, 1998): considérese un subibaja como se muestra en la Figura I.1: un lado de la balanza tiene un contenedor que recibe agua de una llave a una razón constante; en el lado opuesto, la balanza tiene un contrapeso  $h$ . Suponiendo una adecuada relación entre sus constituyentes, esta *neurona mecánica* transluce la intensidad del flujo continuo de agua en la señal periódica de las descargas del contenedor de agua.

**Figura I.1**  
**Neurona mecánica**



Llamando  $v(t)$  el peso del agua en el contenedor al tiempo  $t$  entonces, mientras  $v(t) \leq h$ , tenemos  $v(t) = v_0 + It$ , donde  $v_0 = v(0)$ . Después, una descarga vacía el contenedor a un nivel de reposo ( $v_R$ ); desde ese peso, el contenedor reanuda su incremento de peso hasta alcanzar el umbral una vez más. Si hacemos una modificación más al sistema agregando un elevador, como se muestra en la Figura I.2, obtenemos un sencillo sistema mecánico forzado que simula la dinámica de una célula nerviosa periódicamente forzada (Mendoza *et al.*, 1998).

**Figura I.2**  
**Neurona mecánica forzada**



## II. FORZAMIENTO Y SINCRONIZACIÓN

Es un hecho que en la naturaleza interactúan innumerables sistemas y subsistemas de las dinámicas más variadas. Con frecuencia, un sistema que perturba a un segundo, se ve a sí mismo perturbado por ese otro en una retroalimentación en ocasiones incesante. En otros, debido a estímulos externos, el comportamiento de un sistema se ve afectado sin que éste, aparentemente, influya en aquéllos.

En algunos sistemas pueden identificarse varios componentes oscilatorios que interactúan. Esta interacción puede ser muy complicada, pero generalmente la sincronización juega un papel muy importante. El tipo más general de sincronización es en el que los diferentes componentes del sistema tienen frecuencias conmensurables; es llamado *sincronización racional* (Carrillo, 1984).

El fenómeno de sincronización racional es abundante en el campo de la música. Es también conocido que, cuando escuchamos música, el latido de nuestro corazón tiende a sincronizarse racionalmente al ritmo de ella. El latido del corazón mismo es causado por la acción sincrónica de un gran número de osciladores individuales a nivel celular. Aparentemente, algunas enfermedades del corazón como extracístole (saltarse un latido) o fibrilación (el corazón deja de latir) son causados por la pérdida de sincronización entre los diferentes centros oscilatorios.

En un oscilador forzado periódicamente, la sincronización racional se da cuando el periodo de la respuesta del sistema es conmensurable con el periodo,  $T$ , del forzamiento.

## III. MODELACIÓN DE NEURONAS DE INTEGRACIÓN Y DISPARO

El modelo más común para estudiar estimulaciones de neuronas de integración y disparo es construido usando una ecuación diferencial de la forma (1)

$$\frac{dx}{dt} = F(t, x)$$

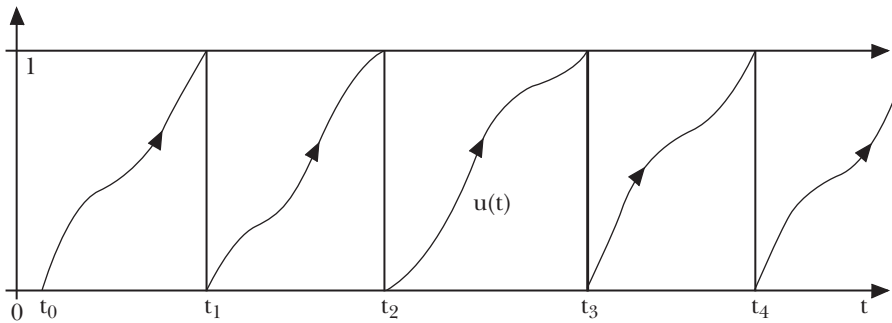
Esta ecuación es usada para modelar el proceso de integración (o acumulación) de una variable de estado,  $v$ , hasta que el valor umbral  $v_T$  es alcanzado y el sistema dispara una respuesta que relaja el valor de  $v$  a su valor de reposo,  $v_R$ , por medio de la *condición de disparo o salto* (2)

$$\lim_{t \rightarrow \tau^+} v(\tau) = v_R \text{ cada tiempo, } \tau, \text{ para el cual } v(\tau) = v_T$$

Es decir,  $v(t)$  es una función discontinua: mientras  $v_R \leq v(t) \leq v_T$ ,  $v$  es igual a la solución de la ecuación diferencial (1) que satisface la condición inicial  $x(\tau) = 0$  sus discontinuidades son de salto y van del valor umbral  $v_T$  al valor de reposo  $v_R$  como lo indica la condición de disparo (2).

Denotemos por  $x(t, \tau, \eta)$ , a la solución de (1) que satisface la condición inicial  $x(\tau) = \eta$  El sistema dispara desde el tiempo inicial  $\tau$ , si  $\tau$  es tal que la solución  $x(t, \tau, 0)$  alcanza el valor umbral  $v_T$  para alguna  $t$  mayor que  $\tau$ . Obsérvese que pueden existir tiempos  $\tau$  para los cuales el sistema nunca dispara:  $x(t, \tau, 0) \neq v_T$ , para toda  $t > \tau$ . La repetición sucesiva del proceso de acumulación y relajación mediante la condición de disparo genera la secuencia de tiempos de disparo  $\{t_n\}$  (véase Figura III.3).

**Figura III.3**  
**Sucesión de disparos de una neurona de integración y disparo modelada por un modelo diferencial**



#### a) El modelo KHR

J. P. Keener, F. C. Hoppensteadt y J. Rinzel (Keener *et al.*, 1981) analizaron un modelo de integración y disparo de la respuesta de una membrana nerviosa a un forzamiento periódico, previamente estudiado por otros autores (Knight, 1972; Rescigno *et al.*, 1970). El modelo describe la evolución del potencial de membrana  $v(t)$  en su proceso de acumulación hasta alcanzar un valor umbral  $v_T$  y emitir un disparo (potencial de acción). Después del disparo, el potencial de membrana regresa a un valor de reposo y el proceso se repite. El modelo matemático que ellos formulan es

$$\frac{dv}{dt} = -\gamma v + S_0 + S_m \cos(\omega t + \varphi)$$

y condición de disparo (3)

$$v(t^+) = 0 \text{ si } v(t) = v_T,$$

donde  $S_0$  es el promedio del estímulo aplicado y  $S_m$ ,  $\omega$  y  $\varphi$  son la amplitud, frecuencia y fase del estímulo, respectivamente.

Después de un cambio de variable y el reescalamiento de los parámetros, el análisis se reduce al estudio de las posibles dinámicas generadas por la función de disparo del modelo diferencial

$$x = -\sigma x + S + H \cos(2\pi t)$$

con umbral normalizado a uno:  $v_T = 1$ . La interpretación del modelo puede ser la de un oscilador autónomo gobernado por la ecuación  $x = -\sigma x + S$  y la condición de disparo (3), sujeto al forzamiento armónico  $H \cos(2\pi t)$ .

#### IV. FUNCIONES DE DISPARO

Se tiene una función  $a$ , la función de disparo, que asocia a cualquier tiempo inicial  $\tau$  (del cual el sistema dispara), el valor  $a(\tau)$  del tiempo de disparo, que es el mínimo tiempo mayor a  $\tau$  en el cual la variable de estado del sistema alcanza su valor umbral. Para simplificar la exposición supondremos que el valor de reposo es cero:  $v_R = 0$ , y que el valor umbral,  $v_T$ , es igual a uno.

Definimos la *función de disparo* del modelo diferencial dado por (1) y (2) como

$$a(\tau) = \min \{t > \tau : x(t; \tau, 0) = 1\}$$

Esta función genera, mediante iteraciones, la familia de sucesiones de disparo del modelo; si para todo tiempo inicial  $\tau$ , el sistema dispara (*i.e.* el dominio es  $D_a = N$ ), la secuencia de disparos  $\{t_n\}$  está bien definida para todo  $n \in N$  y las secuencias de disparo son las órbitas, en la recta real, del semi-sistema dinámico discreto en  $R$  determinado por la función de disparo  $a$ . Puede ocurrir que para algunos tiempos iniciales  $\tau \in R$ , el sistema no dispare ( $D_a$  es un subconjunto propio de  $R$ ). En el caso en el que  $D_a$  esté contenida en la imagen de la función de disparo  $a$ , la sucesión de disparo  $\{t_n\}$  está bien definida para toda  $n \in N$  y son las órbitas del

semi-sistema dinámico discreto generado por la función  $a$ . La función de disparo puede definirse sin mayor problema en el dominio donde el sistema sí dispara, pero no siempre las sucesiones de disparo definen un semi-sistema dinámico y se pueden tener sucesiones finitas de disparos, fenómeno conocido como *muerte del oscilador*.

## V. SINCRONIZACIÓN Y FUNCIONES EN LA CIRCUNFERENCIA

La mayor parte de la investigación acerca de los sistemas forzados ha centrado su atención en el caso de estimulación o forzamiento periódico. Para las neuronas de integración y disparo esto significa que la sucesión de disparo,  $\{t_n\}$ , es periódica ( $t_{n+q} = t_n + P$ , para algún entero positivo  $q$ , algún real positivo  $P$  y para toda  $n$ ) y su periodo  $P$  es un múltiplo entero del periodo del forzamiento:  $P = pT$ . En otras palabras, por cada  $p$  ciclos del forzamiento, el sistema produce  $q$  disparos, y el proceso se repite, de modo que los siguientes disparos se dan siempre en las mismas fases, esto es  $t_{n+q} \equiv t_n \pmod{T}$ . Llamamos a  $q$  el *periodo* y a  $p$  la *envolvencia* de la sincronización.

Para una neurona de integración y disparo periódicamente forzada, la *sucesión de fases de disparo*,  $\{x_n\}$ , es obtenida tomando la clase residual, módulo  $T$ , de la sucesión de disparo:  $x_n = t_n \pmod{T}$  ( $0 \leq x_n < T$ ), donde  $T$  es el periodo del forzamiento. La *función de fases de disparo* es la función definida por

$$\alpha(x) = a(x) \pmod{T}$$

La sucesión  $x_n = t_n \pmod{T}$  está contenida en el intervalo  $[0, T)$  y está formada por las fases en las que se dan todos los disparos del sistema respecto al ciclo del forzamiento. En el caso de sincronización racional, la sucesión de fases de disparo es finita ( $x_n = x_{n+q}$  para toda  $n$ ). Este tipo de sincronización es llamada también *phase locking* por otros autores (Pérez y Glass, 1982).

En el modelo, forzamiento periódico significa que  $F$  es periódica en la variable  $t$ :  $F(t + T, v) = F(t, v)$ . Como  $F$  es periódica en  $t$  es bien conocido que si  $v(t)$  es una solución del modelo diferencial (1), entonces  $v(t + T)$  es también solución (Coddington y Levinson; 1955 Hale, 1969). A partir de este hecho se pueden probar las siguientes afirmaciones (Mendoza *et al.*, 1998)

$$(i) D_a = D_a + T$$

$$(ii) a(t+T) = a(t) + T$$

Las consecuencias de estos hechos son importantes: cuando  $D_a = R$ , podemos pensar a la función de disparo,  $a$ , como un levantamiento de una función de grado uno en la circunferencia, tal función viene siendo precisamente la función de fases de disparo. Tenemos así el siguiente resultado (Mendoza *et al.*, 1998): *Las secuencias de fases de disparo  $\{x_n\}$  de una neurona de integración y disparo periódicamente forzada son las órbitas de un semi-sistema dinámico en la circunferencia, generado por la función de fases de disparo, si  $D_a = R$ .*

Visto de este modo, las propiedades sincrónicas de la respuesta del oscilador al forzamiento periódico están codificadas en la función de fases de disparo. La sincronización se revela como la existencia de órbitas periódicas atractoras. El problema de estudiar las propiedades de sincronización del sistema al forzamiento periódico se vuelve ahora un problema técnico de funciones en la circunferencia: si la función de fases de disparo es un homeomorfismo, la teoría del número de rotación de Poincaré es una herramienta útil para detectar órbitas periódicas atractoras. Si la función de fases de disparo no es un homeomorfismo, pueden ser utilizadas generalizaciones de la noción de número de rotación. En la siguiente sección haremos un breve resumen de los aspectos más relevantes de estas teorías.

## VI. TEORÍAS DE ROTACIÓN

Si la función de disparo es un homeomorfismo de la circunferencia, el *número de rotación* de Poincaré (4)

$$\rho(a) = \lim_{n \rightarrow \infty} \frac{a^n(s)}{n} \bmod 1$$

es útil para determinar la existencia de órbitas periódicas (Nitecki, 1971): si este número es un racional  $p/q$  la función tiene órbitas de periodo  $q$ ; más aún, si la función de fases de disparo no es una rotación, entonces tiene órbitas periódicas atractoras, lo cual, como hemos dicho, interpretamos como sincronización. Cuando el número de rotación no es racional, entonces todas sus órbitas son aperiódicas y densas en todo el círculo o en un subconjunto de Cantor. En este caso no tenemos sincronización del oscilador con el forzamiento.

Si la función de fases de disparo no es un homeomorfismo, se requiere generalizar el concepto de número de rotación. Cuando la función es continua pero no monótona, el límite (4) depende del punto donde es calculado. Esto nos da un intervalo cerrado de números de rotación (Newhouse *et al.*, 1983; Ito, 1981). Por cada número racional  $p/q$  en este intervalo, la función de fases de disparo tiene al menos una órbita periódica de periodo  $q$  (Misiurewicz, 1982), pero la mayoría de ellas son inestables. Algunos resultados, en su mayoría numéricos, muestran que es posible

la coexistencia de órbitas periódicas atractoras de diferente periodo (Pérez y Glass, 1982; Boyland, 1986), fenómeno conocido como *multiestabilidad*; también es posible la existencia de atractores caóticos e, inclusive, la coexistencia de atractores periódicos y caóticos (Martínez-Meckler *et al.*, 1986). Es interesante notar cómo, en este caso, la teoría nos muestra la posibilidad de tener diferentes sincronizaciones, dependiendo de la condición inicial del oscilador forzado. Es más, cabe la posibilidad de tener osciladores forzados para los cuales algunas condiciones iniciales darán lugar a sincronización y otras a comportamientos caóticos.

Si la función de fases de disparo es monótona pero no continua, el número de rotación queda también bien definido por un límite análogo a (4). En este caso se tiene una vez más que, si  $\rho \in Q$ , entonces existen órbitas periódicas atractoras (salvo en el caso de rotaciones), y si  $\rho \notin Q$  las órbitas son densas en un conjunto de Cantor (Keener, 1980). Aquí, como en el caso de homeomorfismos, número de rotación racional implica sincronización. Hasta donde nosotros sabemos, no existe en la literatura una teoría desarrollada para estudiar el caso en el que la función de la circunferencia no es ni monótona ni continua.

De acuerdo con los resultados de estas teorías, para estudiar las propiedades de sincronización de un sistema de integración y disparo periódicamente forzado, necesitamos elementos que nos permitan determinar la regularidad (continuidad e inyectividad) de la función de fases de disparo. Al trabajar sistemas con parámetros, se tiene una familia parametrizada de funciones de fases de disparo  $\alpha_\lambda$ . En el espacio de parámetros podemos considerar las siguientes cuatro regiones, de acuerdo con la regularidad de la función de fases de disparo:

- I. Donde  $a(t)$  es un homeomorfismo,
- II. donde  $a(t)$  es continua,
- III. donde  $a(t)$  es inyectiva y
- IV. donde  $a(t)$  no es ni continua ni inyectiva.

El discernimiento de estas regiones de regularidad también tiene importancia por otras razones; en el espacio de parámetros se tienen diferentes zonas de sincronización (llamadas también *lenguas de Arnold*): por cada racional  $\frac{p}{q}$  se considera el conjunto de valores del parámetro tales que la función de fases de disparo tiene una órbita atractora de periodo  $q$  y *envolvencia*  $p$  (Guzmán, 1994). En cada una de las regiones de regularidad, la estructura de las lenguas de Arnold tiene propiedades distintas: de acuerdo con la teoría del número de rotación y sus generalizaciones: en las regiones I y III las lenguas no pueden intersectarse; en la región II las zonas de sincronización pueden intersectarse, reflejando el fenómeno de multiestabilidad; en la región IV conjeturamos la existencia de multiestabilidad

pero no tenemos ningún resultado teórico que respalde esta afirmación. Es interesante señalar aquí que el análisis de las lenguas de Arnold comúnmente sólo puede hacerse con la ayuda de computadoras, y, aún así, el esfuerzo computacional requerido es considerable.

## VII. TEOREMAS DE REGULARIDAD

Como ya se mencionó, para poder emplear la teoría del número de rotación en el estudio de la sincronización de neuronas de integración y disparo periódicamente forzadas, es muy importante determinar la continuidad y la inyectividad de las funciones de disparos. En Carrillo y Ongay (2001) se estudian condiciones bastante generales que determinan continuidad e inyectividad de la función de disparo. Hemos llamado a los resultados que caracterizan esas propiedades Teoremas de regularidad. Las demostraciones rigurosas de los teoremas de continuidad e inyectividad se pueden encontrar en la referencia citada o en Ongay (2000).

*Teorema (Teorema de inyectividad).* La función de disparo  $a$  es inyectiva en  $\text{int}(D_a)$  si y sólo si  $F(t, 0) > 0$ , para toda  $t \in \text{int}(D_a)$ .

*Teorema (Teorema de continuidad).* La función de disparo  $a$  es continua en  $\text{int}(D_a)$  si y sólo si  $F(t, 1) \geq 0$  para toda  $t$  tal que  $x(t, \tau, 0) = 1$  para alguna  $\tau \in \text{int}(D_a)$ .

### a) Aplicación al modelo KHR

Con el fin de ilustrar el uso de los teoremas de regularidad, en esta sección los aplicaremos al análisis de las regiones de regularidad del modelo KHR. Para este modelo  $F(t, x, \lambda) = -\sigma x + S + H \cos(2\pi t)$ , con  $\lambda = (\sigma, S, H)$  un vector de parámetros. Aquí,  $F(t, 0) \geq 0$  para toda  $t$  si y sólo si  $S - H \geq 0$  y  $F(t, 1) \geq 0$ , para toda  $t$  si y sólo si  $\sigma + S - H \geq 0$ . La ecuación diferencial del modelo KHR tiene una solución periódica asintóticamente estable, a saber

$$\phi(t) = \frac{s}{\sigma} + \frac{H\sigma}{\sigma^2 + 4\pi^2} \cos 2\pi t + \frac{2\pi H}{\sigma^2 + 4\pi^2} \sin 2\pi t$$

$$= \frac{s}{\sigma} + \frac{H}{\sqrt{\sigma^2 + 4\pi^2}} \sin(2\pi t + \beta)$$

donde  $\sin \beta = \frac{\sigma}{\sqrt{\sigma^2 + 4\pi^2}}$

En Mendoza *et al.*, (1998) se muestra que se tiene la siguiente dicotomía:

Si

$$\text{Max } \phi > 1,$$

el sistema dispara una vez que ha sido establecida cualquier condición inicial ( $D_a = R$ ) mientras que si

$$\text{Max } [\phi(t)] \leq 1,$$

para algunas condiciones, el sistema no dispara o produce secuencias finitas de disparos.

Para el modelo KHR

$$\text{Max}[\phi(t)] = \frac{s}{\sigma} + \frac{H}{\sqrt{\sigma^2 + 4\pi^2}}$$

La condición

$$\frac{s}{\sigma} + \frac{H}{\sqrt{\sigma^2 + 4\pi^2}} = 1$$

define una superficie contenida en el espacio de parámetros.

De acuerdo con el teorema de continuidad y la dicotomía, la región del espacio de parámetros donde la función de disparos  $a$  es continua en todo  $R$  queda delimitado por las condiciones:

$$\text{i) } S - H \geq \sigma$$

$$\text{ii) } \frac{s}{\sigma} + \frac{H}{\sqrt{\sigma^2 + 4\pi^2}} > 1$$

La igualdad

$$S - H = \sigma$$

determina un plano en el espacio de parámetros ( $\sigma$ ,  $S$ ,  $H$ ) que cruza el plano  $\sigma = 0$  a lo largo de la recta  $S = H$  y el plano  $H = 0$  a lo largo de la recta  $S = \sigma$ .

Para el modelo KHR, si  $H > 0$  entonces la condición  $S - H \geq \sigma$  implica que  $\text{Max}[\phi(t)] > 1$ . Así, si  $H > 0$  la condición  $S - H \geq \sigma$  es suficiente para tener la continuidad de la función de disparos  $a(t)$  en todo  $R$ .

Como  $F(t,0) \geq 0$  para toda  $t$ , si y sólo si

$$S \geq H$$

entonces los puntos  $(\sigma, S, H)$  que están sobre o bajo el plano  $S = H$  determinan, de acuerdo con el teorema de inyectividad, sistemas con  $a(t)$  inyectiva.

## b) Partición del espacio de parámetros

Como hemos visto, el espacio de parámetros puede ser dividido en regiones que se encuentran arriba o abajo de las superficies descritas anteriormente. Estas superficies, al intersectarse, determinan regiones en el espacio  $(\sigma, S, H)$ , en donde la función de disparos tiene características cualitativamente diferentes.

En la figura VII.4 se muestran cinco regiones [I, II, III, IV, V] determinadas por las superficies discutidas anteriormente. En estas regiones el sistema tiene las siguientes propiedades (Mendoza *et al.*, 1998):

I.  $a$  es un homeomorfismo en  $R$

$$\frac{S}{\sigma} + \frac{H}{\sqrt{\sigma^2 + 4\pi^2}} \geq 1 \text{ y } S - \sigma \geq H$$

II.  $a$  es discontinua e inyectiva en  $R$

$$\frac{S}{\sigma} + \frac{H}{\sqrt{\sigma^2 + 4\pi^2}} \geq 1 \text{ y } S \geq H$$

III.  $a$  es discontinua y no inyectiva en  $R$

$$\frac{S}{\sigma} + \frac{H}{\sqrt{\sigma^2 + 4\pi^2}} \geq 1 \text{ y } S < H$$

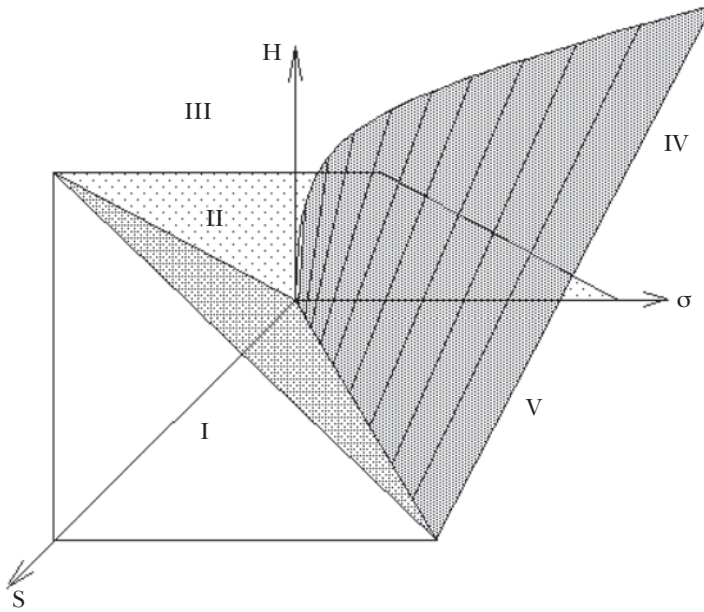
IV.  $D_a$  es un subconjunto propio de  $R$  (vacío o unión de intervalos)

$$\frac{S}{\sigma} + \frac{H}{\sqrt{\sigma^2 + 4\pi^2}} < 1 \text{ y } S < H$$

V.  $D_a = N$

$$\frac{S}{\sigma} + \frac{H}{\sqrt{\sigma^2 + 4\pi^2}} < 1 \text{ y } S \geq H$$

**Figura VII.4**  
**Regiones de regularidad del modelo KHR**



#### REFERENCIAS

- Baer, S. M., J. Rinzel y H. Carrillo (1995). "Analysis of an Autonomous Phase Model for Neuronal Parabolic Bursting", *J. Math. Biol.* 33(3), 309-333.
- Boyland, P. L. (1986). "Bifurcations of Circle Maps: Arnold Tongues, Bistability and Rotation Intervals" *Commun. Math. Phys.* 106(3), 353-381.
- Carrillo, H. (1984). Tesis doctoral, UNAM.
- Carrillo, H. y F. Ongay (2001). "On the Firing Maps of General Class of Forced Integrate and Fire Neurons", *Math Biosciences*, 172, 33-53.
- Cartwright, M. L. (1952). "Non-Linear Vibrations: A Chapter in Mathematical History", *Math. Gazette* 36, 80-88.
- Coddington, E. y N. Levinson (1955). *Theory of Ordinary Differential Equations*, McGraw-Hill, Nueva York.
- Guzmán, J. R. (1994). Tesis de maestría, UNAM.
- Hale, J. K. (1969). *Ordinary Differential Equations*, Wiley-Interscience, Nueva York.

- Ito, R. (1981). "Rotation Sets are Closed", *Math. Proc. Cambridge Philos. Soc.* 889, 107-111.
- Keener, J. P. (1980). "Chaotic Behavior in Piecewise Continuous Difference Equations", *Trans. AMS* 26, 589-604.
- Keener, J. P., F. C. Hoppensteadt y J. Rinzel (1981). "Integrate and Fire Models of Nerve Membrane Response to Oscillatory Input", *SIAM J. Appl. Math.* 41 (3), 503-517.
- Knight, B. W. (1972). "Dynamics of Encoding in a Population of Neurons", *J. Gen. Phys.* 59, 734-766.
- Martínez-Meckler, G. C., R. Mondragón y R. Pérez (1986). "Basin Structure Invariance of Circle Maps with Bistable Dynamics", *Phys. Review A* 3 (33), 2143-2145.
- Mendoza, M. A., H. Carrillo y F. Ongay (1998). "Dinámica de las neuronas de integración y disparo", *Reporte Técnico* 98-I, 763-775.
- Misiurewicz, M. (1982). "Periodic Points of Maps of Degree One of a Circle", *Ergodic Theory Dynamical Systems* 2, 221-227.
- Newhouse, S., J. Palis y F. Takens (1983). "Bifurcations and Stability of Families of Diffeomorphisms", *Inst. Hautes Etudes Sci. Publ. Math.* 57, 5-71.
- Nitecki, Z. (1971). *Differentiable Dynamics, an Introduction to the Orbit Structure of Diffeomorphisms*, MIT Press.
- Ongay Larios, F. (2000). Tesis doctoral, UNAM.
- Pérez, R. y L. Glass (1982). "Bistability. Period Doubling Bifurcations and Chaos in a Periodically Forced Oscillator", *Phys. Lett. A* 90, 441-443.
- Rescigno, A., R. B. Stein, R. L. Purple y R. E. Poppele (1970). "A Neuronal Model for the Discharge Patterns Produced by Cyclic Inputs", *Bulletin of Mathematical Biophysics* 32.
- Van der Pol, B. (1926). "On Relaxation Oscillations", *Philos. Mag.* 7, 2 (11), 978-992.

# ANÁLISIS CAÓTICO DE LAS TRES VARIABLES MACROSCÓPICAS DE TRÁFICO VEHICULAR

*Oscar A. Rosas-Jaimes\**

## Resumen

El tráfico de vehículos automotores puede ser modelado a través de tres bien conocidas variables: la densidad  $k$ , el flujo  $q$  y la velocidad  $v$ . Como muchos fenómenos, la naturaleza de este tipo de tráfico es no lineal, exhibiendo diversos aspectos de esta categoría, tales como el comportamiento caótico.

Este documento muestra un modelo matemático clásico obtenido de relaciones y representaciones gráficas entre estas variables, las cuales a su vez se usan para llevar a cabo análisis que transitan entre comportamientos estables, periódicos y caóticos.

**Palabras clave:** Sistemas no lineales, tráfico vehicular, comportamiento caótico, atractores, exponentes de lyapunov.

## INTRODUCCIÓN

El tráfico o circulación de vehículos tiene una presencia tal en cada aspecto de la movilidad de personas y productos que ha llegado a ser un fenómeno por sí mismo. La física de este tráfico puede ser modelada a través de diferentes aproximaciones. Debido a sus características probabilísticas, puede representarse como un conjunto de modelos estocásticos (Wohl y Martin, 1967), pero es innegable que el tráfico observa reglas y relaciones que muestran un carácter

\* Ingeniería de Transporte. Profesor de Tiempo Completo, Universidad Autónoma del Estado de México, Instituto Literario 100, Col. Centro. C.P. 50000. Toluca, México. Correo electrónico: <oscar.rosasjaimes@yahoo.com>.

determinístico (May, 1990). Esta naturaleza ambigua conduce a adaptar el mejor esquema dependiendo de las necesidades de planeación, investigación o diseño.

Muchos especialistas en tránsito o tráfico vehicular prefieren seguir categorías más prácticas que involucran mediciones y variables. En esta forma, el tráfico puede ser explicado a través de sus unidades individuales, los vehículos o grupos relativamente pequeños de estos mismos espaciamientos, a través de sus posiciones, velocidades o aceleraciones, obteniéndose modelos microscópicos (Zhang y Jarret, 1997). Este tipo de modelos tiene la ventaja de enfocarse en detalles finos, haciendo posible la distinción entre diferentes clases de vehículos, el modelado de características individuales y aun la simulación de estilos de conducción (Treiber y Helbing, 2001). Sin embargo, estas ventajas son al mismo tiempo obstáculos en otras circunstancias, dado que son difíciles de analizar bajo una perspectiva global.

Muchas veces los modelos de tipo acumulativo se usan en redes de caminos debido a que en ellos el tráfico necesita ser simulado o analizado como un continuo, usando variables promediadas o agregadas, cuyos valores puedan ser modelados, simulados, analizados y controlados macrocópicamente, como un fluido compuesto por vehículos (Lighthill y Whitham, 1955b; Richards, 1956). Este punto de vista permite el manejo de vastos conjuntos de caminos existentes hoy en día, tornándose en la principal razón de su uso extenso por muchos profesionales, desde sus primeros desarrollos hasta nuestros días (May, 1990).

Un primer aspecto de la naturaleza del tráfico, independientemente del tipo de modelo que se use, es que es no lineal. Como consecuencia, el tráfico puede exhibir un comportamiento característico, con nodos estables, saturaciones o trayectorias periódicas, las cuales aparecen cuando las condiciones o los parámetros varían, transitando a través de todos esos posibles estados hasta alcanzar uno donde el carácter determinístico de los modelos llega a hacerse impredecible debido a lo que se conoce como sensibilidad a las condiciones iniciales, es decir, se convierte en un estado caótico (Hillborn, 2000).

El presente trabajo se aboca al análisis de este último tipo de comportamiento, pero pone atención también a las posibles trayectorias que pueden ser observadas cuando un parámetro del modelo es modificado. El análisis propuesto aquí se basa en los así denominados diagramas fundamentales, que son representaciones gráficas de las principales variables macroscópicas de tráfico. La sección II describe estas representaciones en modo extenso.

Los diagramas fundamentales tienen una forma bien identificada, pero sus formas exactas dependen del modelo de tráfico sobre el que estén calculados. En este caso se usa un modelo clásico y bien conocido (Greenberg, 1959). Aun cuando se ha demostrado que dicho modelo no se ajusta a valores considerados

no congestionados o de circulación libre, para la variable velocidad provee una aproximación suficiente y sencilla para emular valores promedio de velocidad  $v$ , densidad  $k$  y flujo  $q$  de una corriente de vehículos circulando en una carretera. La sección III describe al modelo que se usa en este artículo, así como algunos aspectos generales de las variables involucradas.

La contribución principal de este escrito es el análisis del tráfico a través de las trayectorias que pueden ser logradas por la variación de alguno de sus parámetros nominales, lo que es equivalente a la modificación del sistema y de sus condiciones, como cuando la capacidad de una carretera permite un volumen más pequeño o más grande de vehículos, o como cuando las condiciones ambientales permiten o restringen las posibilidades de conducción.

Este análisis se lleva a cabo sobre los diagramas fundamentales que pueden obtenerse del modelo descrito. Muchos otros documentos (Low y Addison, 1998; Lo y Cho, 2005; Villalobos *et al.*, 2010) elaboran análisis de trayectorias caóticas sobre sistemas que modelan tráfico, pero están más bien relacionados con modelos microscópicos y no con macroscópicos. Por otra parte, y a diferencia de otros trabajos donde solamente es tomado en cuenta el diagrama fundamental flujo-densidad ( $q-k$ ), en el presente documento los diagramas fundamentales velocidad-densidad ( $v-k$ ) y velocidad-flujo ( $v-q$ ) también se han incluido, se realizan simulaciones sobre estos tres diagramas y sus respectivas discusiones son atendidas en la sección IV.

En la sección V se extienden estos análisis llevados a cabo midiendo el carácter caótico del fenómeno, así como la velocidad de divergencia de trayectorias cercanas, usando exponentes de Lyapunov, los que se encuentran directamente relacionados con la naturaleza no lineal del modelo usado, siendo al mismo tiempo un indicador cualitativo del comportamiento estable o caótico del sistema.

Al final se exponen algunos comentarios obtenidos como conclusión de los principales resultados a los que se llega con este trabajo.

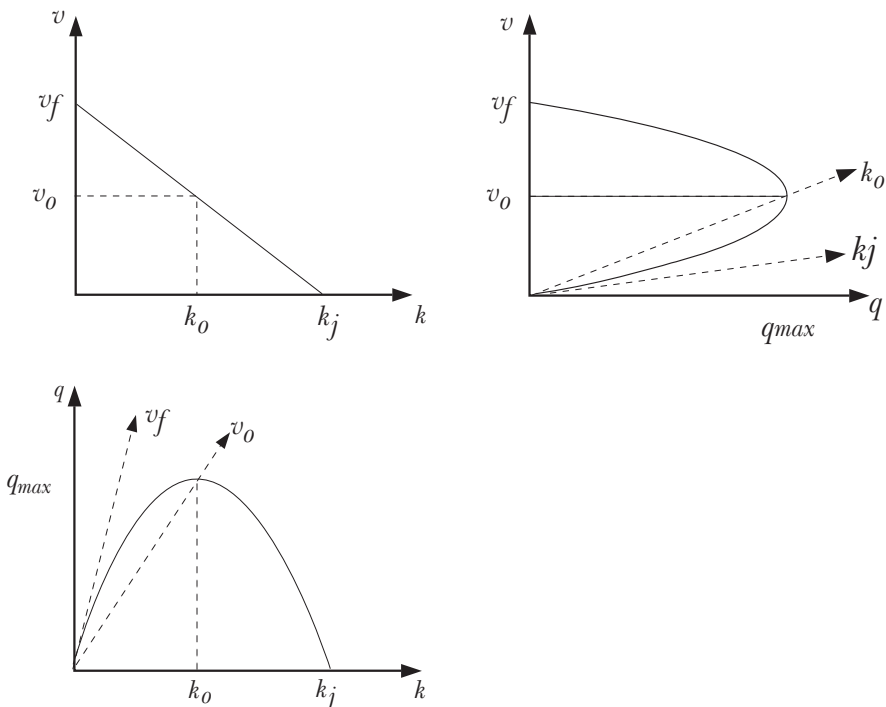
## I. DIAGRAMAS FUNDAMENTALES

Los modelos de tránsito o tráfico vehicular surgen principalmente de una necesidad de entender fenómenos que tienen fuertes implicaciones en aspectos económicos y sociales de la vida moderna, especialmente cuando la física de tales fenómenos tiene que ver con el incremento en la frecuencia de la congestión en vías de tránsito. Los primeros intentos para producirlos comenzaron con aproximaciones estadísticas y con analogías con otros fenómenos físicos (Wohl y Martin, 1967), (Lighthill y Whitham, 1955a; Lučić y Teodorović, 2002), con la intención

de que se ajustaran con datos reales, los cuales parecían seguir curvas que comenzaron a ser llamadas *diagramas fundamentales*.

La Figura I.1 muestra las gráficas flujo-densidad, velocidad-densidad y velocidad-flujo para un modelo de relación lineal entre velocidad y densidad, del cual las otras dos relaciones se pueden obtener (May, 1990). Estas funciones son idealizaciones de puntos verdaderos que representan los valores de las variables macroscópicas de tráfico de vehículos y su utilidad se basa en que son las mejores aproximaciones entendibles para cualquier conjunto de gráficas macroscópicas de tráfico.

En un modelo macroscópico para tráfico en circulación, la velocidad ( $v$ ) es definida como la velocidad del frente de onda de vehículos, muchas veces considerada como la máxima velocidad alcanzada por el promedio de los autos en movimiento. Cuando la congestión aparece, puede ser más valioso conocer la velocidad del frente en retroceso de la congestión ( $w$ ), una cantidad que mide un frente de onda que se mueve en la dirección opuesta al flujo del tráfico (Daganzo, 1994). Sin embargo, pocos trabajos hacen uso de esta cantidad (Rosas-Jaimes y Álvarez, 2007).



**Figura I.1.** Gráficas de relaciones fundamentales entre flujo ( $q$ ), velocidad ( $v$ ) y densidad ( $k$ ).

Los flujos  $q$  son cantidades relacionadas con un conjunto de vehículos que se mueven con respecto al tiempo. Puede ser un número de automóviles atravesando un punto en un periodo de tiempo o un conjunto de autos atravesando una sección de carretera en un intervalo de tiempo (Cal y Mayor y Cárdenas, 2007).

La densidad  $k$  es definida como el número de vehículos que ocupan una sección de un carril o tramo de una carretera o cualquier otro tipo de camino. La medición directa de la densidad puede obtenerse por fotografías aéreas, imágenes videograbadas o por observaciones *in situ*, limitando la longitud de un camino y contando los vehículos presentes en él en determinado momento (Thamizh y Dhivya, 2010).

Estas tres cantidades se encuentran relacionadas en una forma simple a través de

$$q = vk \quad (1)$$

y las relaciones entre parejas de ellas se representan en la Figura I.1. La forma de estas curvas es bastante descriptiva y al mismo tiempo idealizada. Depende de casos particulares de caminos y condiciones. Cada una de ellas muestra una función completa y continua pero es improbable encontrar el rango total de valores de cada variable en un punto de medición. Los datos obtenidos en la vida real tienen discontinuidades múltiples en las cuales muchas partes de estas curvas no se encuentran presentes (Kim y Zhang, 2004).

Sin embargo, estas curvas ilustran puntos significativos. Nótese, por ejemplo, que el flujo nulo ( $q = 0$ ) ocurre en dos condiciones diferentes:

1. Cuando en un camino no hay autos, la densidad  $k$  y el flujo  $q$  son cero. La velocidad es teórica y el valor del primer conductor que aparezca será presumiblemente alto. Esta velocidad se representa en el diagrama fundamental como  $v_f$  y se conoce como *velocidad de flujo o tránsito libre*.
2. Cuando la densidad  $k$  llega a ser tan grande que todos los vehículos son forzados a detenerse, el flujo  $q$  es cero otra vez, debido a que no hay movimiento. La densidad en esta situación es conocida como densidad de atascamiento y es referida como  $k_j$ .

Entre estos dos extremos hay muchas condiciones de tránsito de automóviles. Conforme la densidad se incrementa desde cero el flujo hace lo mismo, ya que hay más autos en el camino, pero la velocidad declina en valor, debido a la creciente interacción entre vehículos. Este decremento en la velocidad es imperceptible cuando las densidades y los flujos son bajos o incluso de valor mediano.

La velocidad decrece un poco antes de alcanzar el valor de flujo máximo. Esta condición se muestra en los diagramas de la Figura I.1 como la *velocidad crítica* o *velocidad óptima*  $v_o$ , la *densidad óptima*  $k_o$  y el *flujo máximo*  $q_{max}$ .

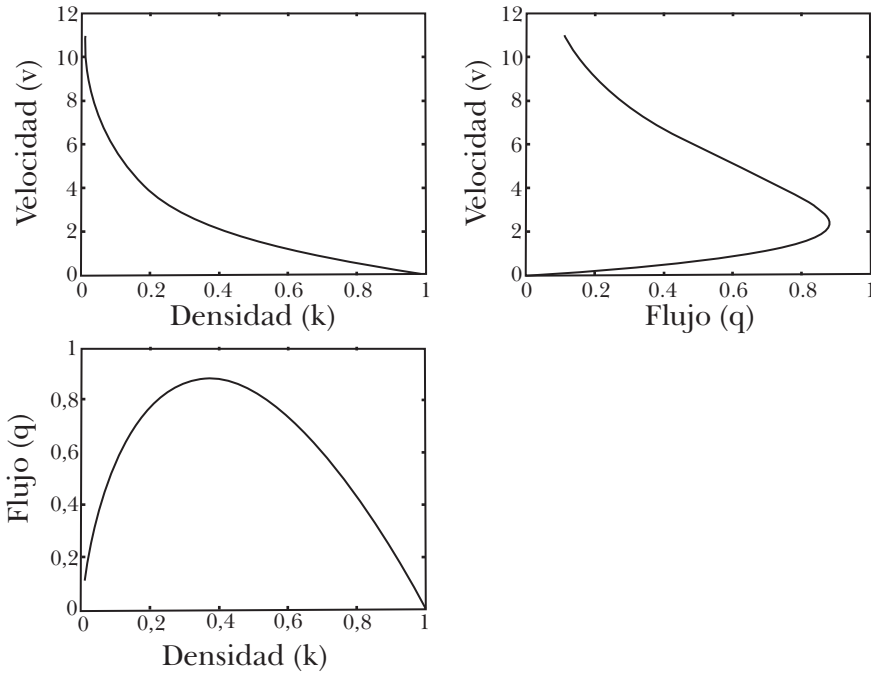
La pendiente de una línea recta dibujada desde el origen del diagrama velocidad-flujo hacia cualquier punto en la curva representa densidad. En forma similar, una línea recta dibujada desde el origen del diagrama densidad-flujo hacia cualquier punto sobre la curva representa velocidad. Estas pendientes pueden ser calculadas a partir de la ecuación (1). Otra pendiente importante es la que corresponde con la velocidad del frente de congestión  $w$ , no representado en la figura pero que sería dibujado desde el punto extremo de la densidad de atascamiento y con una pendiente negativa.

Nótese que estos tres diagramas son redundantes debido a que una vez establecida una relación entre una de las variables las otras dos quedan definidas. Cada una de estas relaciones (y su respectivo diagrama fundamental) tiene su propio campo de aplicación. Por ejemplo, el diagrama velocidad-densidad es la base para los modelos de flujo vehicular, debido a que para un valor de densidad corresponde solamente un valor de velocidad. La relación flujo-densidad es el punto de partida para control de tráfico, debido a que es posible identificar fácilmente aquellas regiones en las que el tráfico puede ser considerado libre, congestionado o en transición. La relación velocidad-flujo es útil porque muestra regiones de estos valores que pueden ser vinculados directamente con niveles de servicio en los caminos (May, 1990; Cal y Mayor y Cárdenas, 2007).

Como es posible observar en la Figura I.1, cualquier valor de flujo distinto del máximo puede ocurrir en dos condiciones diferentes, una con baja densidad y alta velocidad y otra con alta densidad pero baja velocidad. La porción de las curvas para este último caso representa la situación congestionada, con cambios súbitos en el tráfico, algunos de ellos periódicos, algunos caóticos, como se verá más adelante.

## II. MODELO DE GREENBERG

La ecuación (1) relaciona densidad, velocidad y flujo, pero dice poco acerca de la manera en la que los datos correspondientes se ajustan con el modelo usado, y diferentes esquemas de ajuste han sido propuestos. Infortunadamente, los datos de flujo de tráfico parecen ser lo bastante complejos como para que ningún modelo pueda lograr un ajuste perfecto. Algunos de estos modelos son mejores que otros para diferentes regímenes de tránsito.



**Figura II.2.** Diagramas fundamentales basados en el modelo de tráfico de Greenberg con valores normalizados.

El modelo de Greenberg fue el resultado de observar conjuntos de datos de velocidad y densidad para túneles, especialmente aquellos que describen congestionamiento (Greenberg, 1959).

Este autor correlaciona esta información con la analogía hidrodinámica (Lighthill y Whitham, 1955a, b; Richards, 1956) a partir de la ecuación de movimiento de un fluido unidimensional

$$\frac{dv}{dt} = \frac{v_o^2}{k} \frac{\partial k}{\partial x} \tag{2}$$

donde  $x$  es la coordenada de la distancia a lo largo del camino,  $t$  es tiempo, el parámetro  $v_o$  es la velocidad óptima o crítica,  $v$  y  $k$  son las variables ya conocidas de velocidad y densidad.

Ya que la velocidad es una función de la distancia y el tiempo, y por las propiedades de la derivación total, la ecuación (2) puede ser escrita en la forma

$$\frac{dv}{dt} \frac{\partial k}{\partial x} + v \frac{dv}{dk} \frac{\partial k}{\partial x} + \frac{v_0^2}{k} \frac{\partial k}{\partial x} = 0 \quad (3)$$

La cual puede ser dividida por  $dv/dk$  para obtener

$$\frac{\partial k}{\partial x} + v + \left( \frac{v_0^2}{k} \frac{dk}{dv} \right) \frac{\partial k}{\partial x} = 0 \quad (4)$$

Greenberg luego usa la expresión de conservación de masa

$$\frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0 \quad (5)$$

La cual, por (1), puede ser escrita como

$$\frac{\partial k}{\partial t} + \frac{\partial k}{\partial x} + k \frac{dv}{dk} \frac{\partial k}{\partial x} = 0 \quad (6)$$

En esta forma, con (4) y (6) es posible describir el comportamiento de los vehículos.

Ya que este par de ecuaciones constituye un sistema, el autor de este modelo obtiene una solución no trivial

$$\frac{dv}{dk} = -\frac{v_0}{k} \quad (7)$$

La ecuación diferencial (7) sigue la trayectoria dada por

$$v = v_0 \ln \left( \frac{k_j}{k} \right) \quad (8)$$

cuya función es mostrada en forma gráfica en la Figura II.2 en la porción que corresponde al diagrama fundamental de velocidad-densidad, el cual, como puede verse, calcula valores infinitos de velocidad. Aun cuando este hecho puede ser visto como un defecto para este modelo, Greenberg pudo demostrar que para regiones alejadas de estos valores de velocidad de flujo libre el modelo se ajusta bastante bien.

Combinando las ecuaciones (1) y (8) resulta una relación para la densidad  $k$  y el flujo  $q$

$$q = v_0 k \ln \left( \frac{k_j}{k} \right) \quad (9)$$

el cual es mostrado en la Figura II.2 en la respectiva porción del diagrama fundamental flujo-densidad.

La gráfica correspondiente a la relación velocidad-flujo es más fácil de lograr a través de un conjunto de valores obtenidos de los datos de flujo y densidad a través de la ecuación (1) en la forma

$$k = \frac{q}{v} \quad (10)$$

La cual es representada en la porción respectiva de la Figura II.2. Greenberg calculó y analizó valores paramétricos tales como el flujo máximo  $q_{max}$ , la velocidad óptima  $v_o$  y la densidad de atascamiento  $k_j$ . Estos dos últimos son difíciles de estimar en la práctica, y muchas veces son supuestos a través de observaciones directas de representaciones gráficas de los datos.

Colecciones reales de datos de tráfico muestran gran variación y dispersión y, como ya se ha mencionado, es difícil para cualquier modelo lograr un ajuste satisfactorio con ellos. Greenberg comentó sobre este hecho, puntualizando que este modelo representa un gran porcentaje de cualquier conjunto de datos de tráfico para una gran variedad de condiciones carreteras, lo que se ha confirmado por posteriores mediciones (May, 1990).

### III. ANÁLISIS DE COMPORTAMIENTOS NO LINEALES

El modelo de Greenberg posee otras características que lo hacen conveniente: su formulación teórica es sólida y relativamente sencilla, lo que la hace fácilmente discretizable.

Sea el conjunto de ecuaciones (8), (9) y (10) funciones que describen la relación entre pares de variables macroscópicas de tráfico. Con tal de generalizar un resultado, defínase la densidad máxima con el valor  $k_j = 1$ . De la función de flujo-densidad es posible notar que el flujo máximo  $q_{max}$  ocurre entonces cuando  $k_j / k = e$ , y que la densidad óptima es entonces  $k_o = e^1$ . Con estos valores puede entonces establecerse una normalización para este conjunto de funciones.

Por otro lado, la velocidad óptima  $v_o$  es un parámetro que puede ser usado para ajustar estos diagramas fundamentales para diferentes tipos de datos de tráfico, alargando o encogiendo sus contornos. Esto es equivalente a modificar las condiciones carreteras para el modelo, dado que a un valor mayor que  $v_o$  corresponde una mayor capacidad o alguna otra característica agrandada para el desplazamiento vehicular.

Para analizar el comportamiento de un sitio hipotético en algún camino o carretera, supondremos que los valores actuales en las funciones (8) – (10) condicionan el siguiente estado bajo un esquema discreto.

De esta forma, el esquema iterativo particular para la función (9) se da mediante una variable discreta de flujo  $q(i)$

$$q(i) = v_0 k(i) \ln \left( \frac{k_j}{k(i)} \right), \quad i = 1, 2, 3, \quad (11)$$

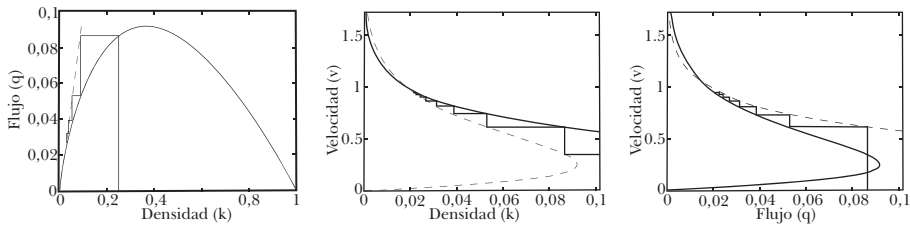
la cual, por la normalización efectuada, se iguala a los valores de densidad por medio de

$$k(i + 1) = 1(i) \quad (12)$$

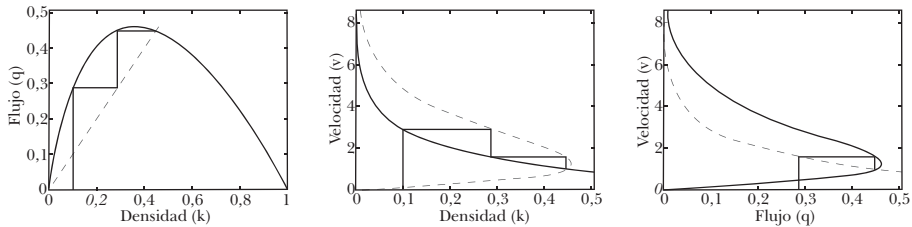
Es decir, comenzando desde una condición inicial y con los parámetros definidos hasta ahora, estas expresiones generarán un nuevo valor  $k(i + 1)$  a partir de una entrada previa  $q(i)$ , luego ese nuevo valor se utilizará para calcular otro valor de flujo  $q(i + 1)$ , el cual generará un siguiente valor de  $k$ , y así sucesivamente en una forma iterativa, y todos estos valores crearán un conjunto que puede ser registrado y mapeado. Físicamente, este sería el caso de una red cerrada de tráfico compuesta por arcos y nodos, todos incluidos en un volumen de control, pero tomando en cuenta sólo un punto para conocer el estado, siendo las interacciones y la geometría de dicha red reflejadas en ese solo punto.

De esta manera, si  $k = 0.25$  es una condición inicial con  $v$ , una primera serie de iteraciones puede ser ejecutada. En la Figura III.3 es posible seguir la trayectoria de las iteraciones calculadas. Para ésta y el resto de las iteraciones, se han llevado a cabo un número de 300 iteraciones. La línea  $q = k$  se ha dibujado sobre la gráfica flujo densidad, siendo la más comúnmente utilizada en la literatura. Es sumamente útil para visualizar la forma en la que las iteraciones se desarrollan, representadas por líneas verticales y horizontales que la intersectan. De esta forma, puede notarse que comenzando desde la condición inicial, el valor final se detiene en un punto cercano a  $q = k = 0.0183$ , indicando que el estado final se estabiliza para cualquier tiempo (iteración) futuro(a) en un valor de flujo libre, como el transitorio de una carretera que se vacía. La condición inicial elegida, así como otras que se verán en el resto de este trabajo, ha sido seleccionada debido a que exhibe un transitorio rápido, el cual permite visualizar mejor el comportamiento de las iteraciones.

Las gráficas correspondientes con los diagramas velocidad-densidad y velocidad-flujo muestran la convergencia en las iteraciones hacia  $v = 1$  y confirman el punto de estabilidad para  $k$  y  $q$ . Estos mapeos no repiten la misma idea de incluir algo similar a la línea de  $45^\circ$  para demarcar las iteraciones y más bien se ha preferido correr en paralelo las simulaciones para las otras variables.



**Figura III.3.** Comportamiento iterativo de los diagramas fundamentales normalizados. Velocidad óptima  $v_o = 0.25$ , condición inicial  $k = 0.25$ .



**Figura III.4.** Comportamiento iterativo de los diagramas fundamentales normalizados. Velocidad óptima  $v_o = 1.25$ , condición inicial  $k = 0.1$

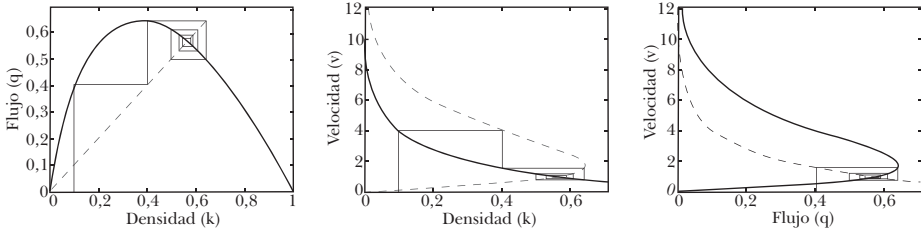
Hacerlo de esa manera permite observar que las iteraciones no solamente siguen el perfil del diagrama fundamental, sino que también lo hacen con el perfil de la variable faltante en el diagrama; es decir, si las iteraciones se efectúan sobre el diagrama fundamental velocidad-densidad, dichas iteraciones también tocan lo que sería el perfil del diagrama fundamental velocidad-flujo.

Esta es una consecuencia de la relación entre las tres variables macroscópicas, lo que constituye un resultado esperado, aunque no ha sido reportado en la literatura, hasta donde se sabe.

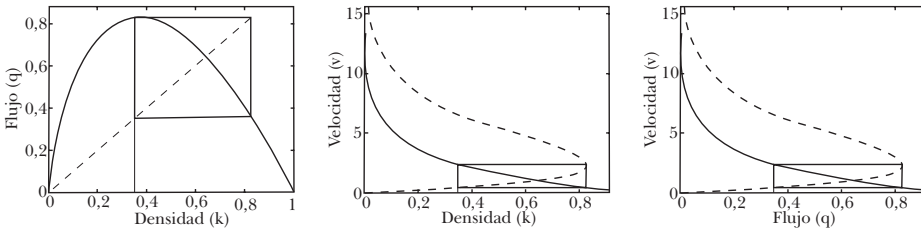
Si la condición inicial se cambia, el transitorio seguirá una trayectoria distinta, pero se alcanzará el mismo valor final, es decir, el estado estable de ambas situaciones es el mismo. Se dice que el punto  $q = k = 0.0183$  ( $v - 1$ , simultáneamente) es un *atractor* para las trayectorias del sistema (Hillborn, 2000).

Conforme el parámetro  $v_o$  varía, un conjunto de valores finales resulta, los cuales pueden ser dibujados como en la Figura III.11, conocida como *mapa de*

*bifurcaciones*. En él se dibujan líneas verticales que corresponden con los valores  $v_o$  usados y que cruzan con el valor final alcanzado.



**Figura III.5.** Comportamiento iterativo de los diagramas fundamentales normalizados. Velocidad óptima  $v_o = 1.75$ , condición inicial  $k = 0.1$ .



**Figura III.6.** Comportamiento iterativo de los diagramas fundamentales normalizados. Velocidad óptima  $v_o = 2.25$ , condición inicial  $k = 0.35$ .

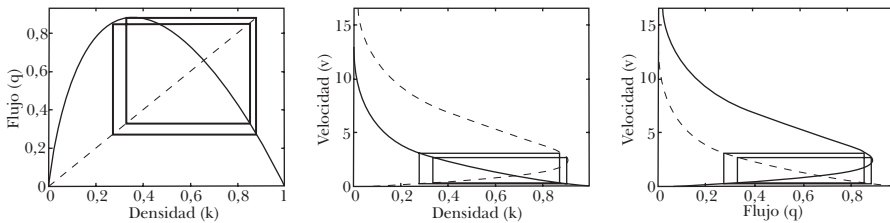
La primera de tales líneas señala el valor del primer conjunto de iteraciones que ha sido descrito. Dado que este trabajo toma en consideración las tres variables macroscópicas de tráfico, el mapa de bifurcaciones  $v - v_o$  también es presentado (Figura III.12). Los valores finales de  $q = k$  y  $v$  son mostrados en ambos mapas de bifurcaciones para cada grupo de iteraciones efectuada con diferentes valores de  $v_o$ .

La Figura III.4 muestra otro conjunto de iteraciones para los tres diagramas fundamentales cuando la velocidad óptima se ajusta a  $v_o = 1.25$  y la condición inicial es  $q = k = 0.1$ . Otra vez, un punto estable se alcanza, pero al contrario de la primera serie de iteraciones, donde se alcanzaron valores de régimen de tráfico

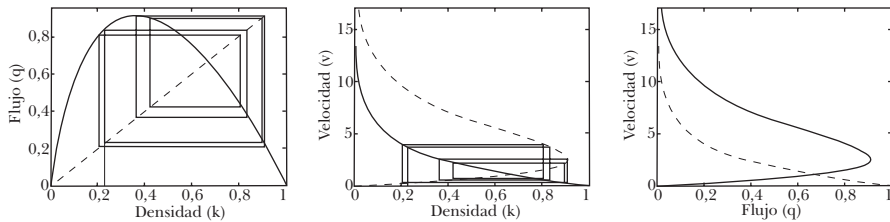
libre, en este caso se obtienen valores de tráfico congestionado, aun cuando el valor inicial de densidad pertenece a valores de tráfico libre.

Esta situación simula un proceso de congestión en una ruta, debido a características sobre el camino, representadas por  $v_0$ . En esta ocasión los valores logrados son  $q = k = 0.4493$  y la velocidad es otra vez  $v = 1$ . Los mapas de bifurcaciones (Figuras III.11 y III.12) exhiben estos valores en una manera análoga. Al igual que en la pasada simulación, si la condición inicial se cambia sólo el transitorio es diferente, pero el estado estable alcanzado al final llega a ser el mismo que en el caso anterior.

Manteniendo la condición inicial pero cambiando al parámetro  $v_0 = 1.75$  la situación corre hacia otro punto de estabilidad, pero las iteraciones se muestran cíclicas aunque atenuadas al final (Figura III.5).



**Figura III.7.** Comportamiento iterativo de los diagramas fundamentales normalizados. Velocidad óptima  $v_0 = 2.405$ , condición inicial  $k = 0.275$ .



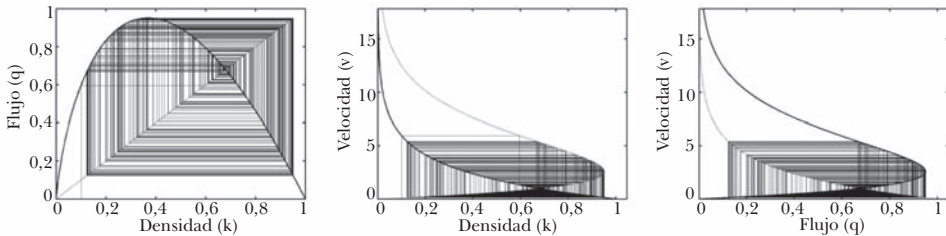
**Figura III.8.** Comportamiento iterativo de los diagramas fundamentales normalizados. Velocidad óptima  $v_0 = 2.48$ , condición inicial  $k = 0.23$ .

Como se verá en breve, este comportamiento marca un límite importante en los mapas de bifurcaciones. Los valores finales para las tres variables son  $q = k = 0.5647$  y  $v = 1$ .

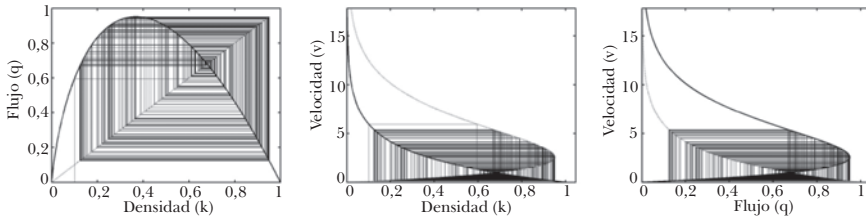
Si  $v_o = 2.25$  el atractor de las variables de tráfico se comporta en forma puramente cíclica (Figura III.6). De hecho, los valores para la densidad y el flujo fluctúan entre  $q = k = 0.3533$  y  $q = k = 0.8271$ , y los valores para la velocidad lo hacen entre  $v = 2.3409$  y  $v = 0.4272$ , una vez que los transitorios han desaparecido. Esto se corresponde en los mapas de bifurcaciones, donde este valor de velocidad óptima marca la región donde dos ramas señalan un conjunto de pares de valores finales. El significado físico es una respuesta oscilatoria de los vehículos, que pasan de valores cercanos a los óptimos a valores de congestiónamiento de un tiempo (iteración) al siguiente.

Si la velocidad óptima es variada nuevamente a  $v_o = 2.405$  (Figura III.7), con una condición inicial para este caso de  $k = 0.275$ , el tráfico se convierte en un ciclo de periodo 4, con valores  $q = k = 0.8496$ ,  $q = k = 0.3330$ ,  $q = k = 0.8806$ ,  $q = k = 0.2692$  y  $v = 3.1560$ ,  $v = 0.3919$ ,  $v = 2.6446$ ,  $v = 0.3057$ . Si se pone atención en los mapas de bifurcaciones, las ramas que corresponden con conjuntos de 4 valores para cada variable de tráfico corren en un intervalo que va desde aproximadamente  $v_o = 2.395$  hasta  $v_o = 2.440$ .

Este proceso puede continuar. Si ahora  $v_o = 2.48$ , y la densidad inicial es  $k = 0.23$  las iteraciones son las que se muestran en la Figura III.8, y se relacionan con la correspondiente velocidad óptima en los mapas de bifurcaciones. El comportamiento cíclico se ha vuelto una oscilación de periodo 8 con  $q = k = [0.8094, 0.4244, 0.9021, 0.2305, 0.8389, 0.3654, 0.9123, 0.2076]$  y  $v = [3.8987, 0.5243, 2.1256, 0.2555, 3.6392, 0.4356, 2.4965, 0.2276]$ .



**Figura III.9.** Comportamiento iterativo (caótico) de los diagramas fundamentales normalizados. Velocidad óptima  $v_o = 2.585$ . Condición inicial:  $k = 0.1$ .



**Figura III.10.** Comportamiento iterativo (caótico) de los diagramas fundamentales normalizados. Velocidad óptima  $v_o = 2.585$ . Condición inicial:  $k = 0.101$ .

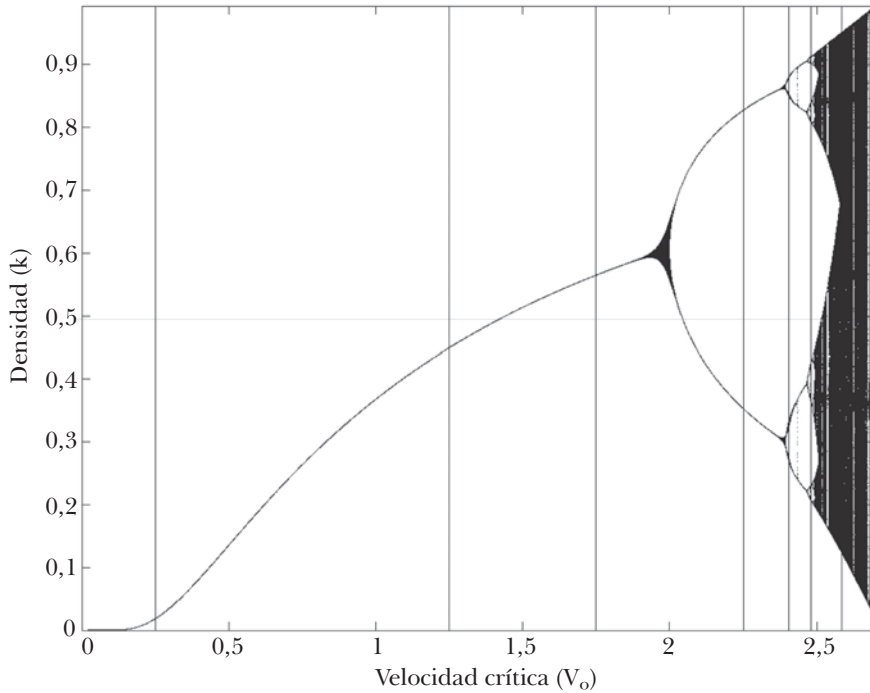
Es importante recalcar que la distancia entre dos valores adyacentes de  $v_o$  decrece para cada conjunto de bifurcaciones, y que un conjunto de bifurcaciones de periodo  $2^n$  alcanza más rápidamente el siguiente conjunto de bifurcaciones de periodo  $2^{n+1}$ , hasta que son muy difíciles de distinguir. Esta situación causa que una muy pequeña variación en el parámetro  $v_o$  pueda exhibir inestabilidades entre trayectorias.

Aunque el parámetro  $v_o$  permaneciese fijo, un gran número de valores finales pueden resultar con una gran sensibilidad a la elección de las condiciones iniciales.

Un ejemplo de esta situación se muestra en las Figuras III.9 y III.10, donde  $v_o = 2.585$  y  $k = 0.1$  para la primera de ellas, la cual muestra lo que se conoce como *trayectorias caóticas*. El comportamiento cíclico es sólo aparente (algunas veces llamado *cuasi-periódico*) y, estrictamente hablando, no existe, ya que los distintos valores no exhiben un periodo en el cual series de ellos puedan repetirse. Esta clase de atractor es conocido como *extraño*.

Los valores finales alcanzados luego de 300 iteraciones son  $q = k = 0.8451$  y  $v = 4.0519$ . Si la velocidad óptima  $v_o$  se deja sin cambios pero se cambia ligeramente la condición inicial a  $q = k = 0.101$ , la Figura III.10 muestra una imagen que a primera vista parece igual a la anterior, pero que observándola detenidamente revela que la trayectoria es bastante diferente. De hecho, los valores finales después del mismo número de iteraciones son ahora  $q = k = 0.2020$  y  $v = 0.2199$ .

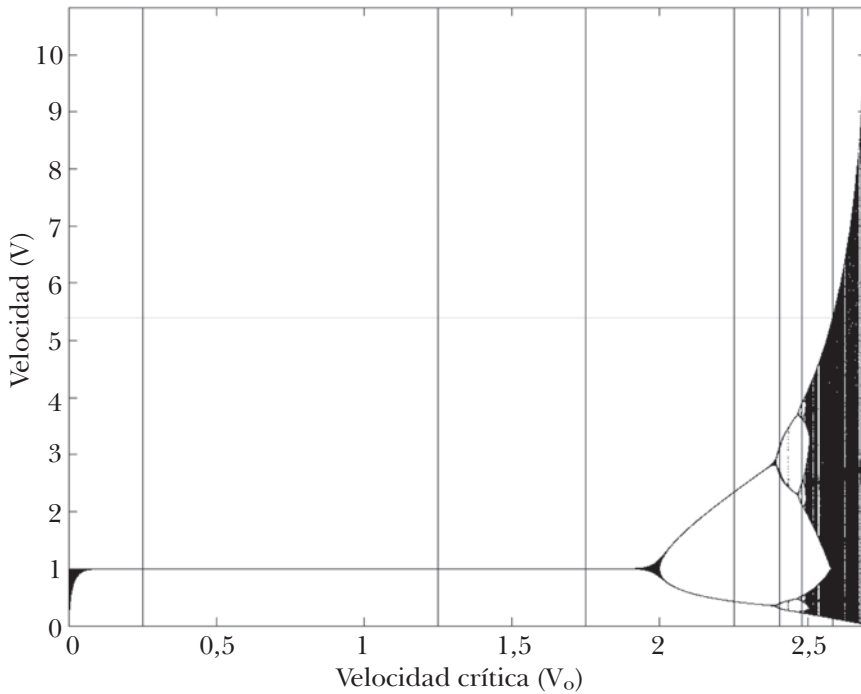
Esta característica consistente en una divergencia creciente entre dos trayectorias que tienen un estado inicial muy cercano es determinista pero impredecible, propia de muchos sistemas no lineales bajo cierto conjunto de parámetros. En un conjunto más lineal de trayectorias, se espera que para diferencias pequeñas en las condiciones iniciales, las trayectorias generadas permanezcan suficientemente cercanas como para no tomar en cuenta dicha diferencia, o que al menos las diferencias sean proporcionales. En otras palabras, para estados iniciales muy cer-



**Figura III.11.** Diagrama de bifurcaciones  $k - v_0$ .

canos, las trayectorias permanecen lo suficientemente juntas entre sí como para realizar predicciones en los comportamientos de los sistemas que representan.

Sin embargo, como se ha visto en las últimas iteraciones que aquí se han mostrado, esta diferencia se vuelve importante para los llamados sistemas caóticos, como en el caso de este esquema de tráfico de vehículos, ya que en estas situaciones un conjunto de estados muy cercanos divergirá entre sí. Esto significa físicamente que es imposible llevar a cabo una simulación de tales sistemas con la intención de predecir estados futuros, ya que los estados reales al comienzo de dicha simulación no pueden medirse y transcribirse a los cálculos con infinita precisión. Estas diferencias infinitesimales entre los valores medidos y los reales crecerán rápidamente, en una forma exponencial, según transcurre el tiempo (las iteraciones). Las predicciones solamente serían válidas, bajo cierta tolerancia, para unos pocos estados futuros.



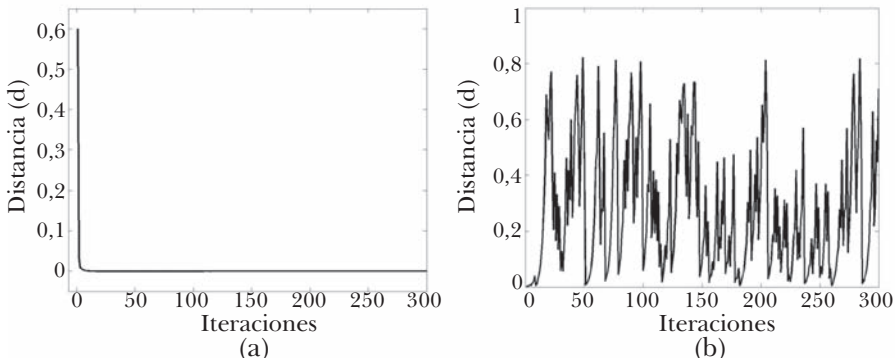
**Figura III.12.** Diagrama de bifurcaciones  $v - v_0$ .

#### IV. DIVERGENCIA DE TRAYECTORIAS CERCANAS

Traducir un sistema a partir de sus principios físicos en un modelo matemático implica, la mayor parte del tiempo, que es posible sustituir valores numéricos en los parámetros y variables que lo conforman y con ello obtener un resultado que es equivalente a la medida del comportamiento del sistema estudiado, es decir, cuando los valores reales de dicho sistema (temperatura, presión o humedad, si se habla del clima [Lorenz, 1996]; o densidad, flujo y velocidad, si es el tráfico)

se encuentran lo suficientemente cercanos a los valores obtenidos mediante cálculos hechos en el modelo, se puede decir que el modelo es una representación aceptable de dicho sistema.

De esta manera, muchos modelos pueden predecir las trayectorias que representan las variables del sistema, con un margen de error aceptable por un tiempo indefinido. Sin embargo, esto es verdad principalmente para sistemas relativamente simples, modelos de sistemas lineales en su mayoría.



**Figura IV.13.** Distancias entre dos trayectorias. (a) Con densidades iniciales  $k = 0.1$  y  $k = 0.7$  respectivamente, ambos con  $v_0 = 1.75$ . (b) Con densidades iniciales  $k = 0.1$  y  $k = 0.101$  respectivamente, ambos con  $v_0 = 2.585$ .

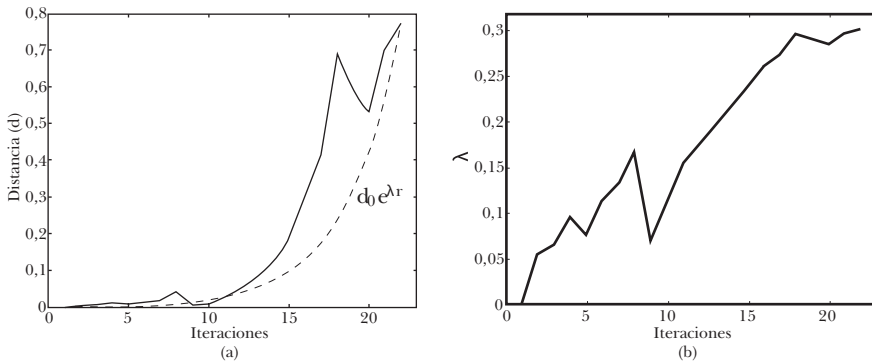
Conforme un mayor número de sistemas, naturales o artificiales, ha sido estudiado, y conforme un rango más amplio de valores, situaciones y complejidades ha sido tomado en cuenta, los modelos han tenido que incorporar características no lineales que aún son capaces de predecir estados futuros bajo cierto conjunto de valores paramétricos. Pero, para otro conjunto de tales parámetros, esta predictibilidad comienza a desaparecer, como se ha ilustrado ya en la sección pasada.

Esto significa que una discrepancia muy pequeña entre dos estados iniciales  $k_m$  y  $k_n$  con valores iniciales muy cercanos entre sí podrían terminar en puntos muy diferentes, según puede verse en las Figuras III.9 y III.10, y que sus trayectorias,

aunque comiencen muy cerca una de la otra, divergirán en una forma creciente, una característica que es propia de los sistemas caóticos (Hillborn, 2000). Estas aproximaciones geométricas constituyen un aspecto cualitativo del sistema de tráfico vehicular modelado. Para una aproximación cuantitativa de esta característica caótica, una medida de la divergencia de estas trayectorias podría llevarse a cabo a través de la aplicación de las iteraciones a partir de dos condiciones iniciales ligeramente distintas  $k_{m0}$  y  $k_{n0} = k_{m0} + \epsilon$ , con lo que se obtendrían a su vez  $q(k_{m0})$  y  $q(k_{n0}) = q(k_{m0} + \epsilon)$  respectivamente. Y en forma similar y paralela se pueden obtener  $v(k_{m0})$  y  $v(k_{n0}) = v(k_{m0} + \epsilon)$  también respectivamente.

Sea la distancia  $d_r$  entre los valores calculados para cada iteración  $m$  y  $n$  de cada conjunto de valores de flujo (o de valores de velocidad) calculados

$$\begin{aligned}
 d_0 &= |q(k_{m0}) - q(k_{n0})| = |q_m - q_n| \\
 d_1 &= |q_{m+1} - q_{n+1}| \\
 d_2 &= |q_{m+2} - q_{n+2}| \\
 &\dots \\
 d_r &= |q_{m+r} - q_{n+r}|
 \end{aligned}$$



**Figura IV.14.** Análisis de distancias entre dos trayectorias caóticas, para  $N = 22$ . (a) Las distancias crecen con cada iteración, ajustándose a una función exponencial con parámetro  $\lambda$ . (b) Cálculo de  $\lambda$  para cada  $d_r$ .

Esta distancia se reducirá para trayectorias convergentes y crecerá para aquellas que diverjan entre sí. Por ejemplo, considérense las iteraciones para  $v_0 = 1.75$  y condiciones iniciales  $k = 0.1$  (Figura III.5) y  $k = 0.7$ , que son valores iniciales bastante separados entre sí dada la normalización. Es sencillo obtener las trayectorias que son tanto oscilatorias como convergentes, y entonces observar que las distancias entre puntos análogos según las iteraciones se reducen rápidamente (Figura IV.13a).

Esto es verdad para la mayoría de las series iterativas que se han mostrado, excepto para aquéllas llamadas caóticas, donde dicha distancia se comporta en una forma distinta. De hecho, procediendo en forma similar, para trayectorias caóticas (Figuras III.9 y III.10), condiciones iniciales distintas pero cercanas implican distancias crecientes entre tales trayectorias (Figura IV.13b).

Es evidente que un sistema como el de tráfico aquí expuesto, se encuentra acotado, esto es, las distancias  $d_r$  crecen pero no para siempre, y su dinámica caótica muestra que el estado del sistema cambia con el tiempo barriendo un rango amplio de valores de tal forma que cualquier trayectoria eventualmente retornará a una región ya visitada previamente. Sin embargo, es posible demostrar (Hillborn, 2000) que dicha trayectoria nunca cruza exactamente por el mismo punto dos veces, es decir, ninguna trayectoria se cruza a sí misma. Esto es difícil de mostrar en una gráfica bidimensional, donde más bien es importante observar los puntos calculados en lugar de las líneas que los conectan.

Tiene sentido tomar un número limitado de distancias  $d_r$  para observar cómo incrementan sus tamaños debido a las trayectorias caóticas con las que se relacionan. Si se propone un número de iteraciones tomadas en cuenta para dicho cálculo, por ejemplo  $N = 22$ , el cual es un número promedio de iteraciones que exhiben este comportamiento de distanciamiento incremental; y en la Figura IV.14a, se observa un incremento exponencial en las distancias  $d_i$ , lo cual hace posible suponer que éstas siguen una función como

$$d_r = d_0 e^{\lambda r} \quad (14)$$

o más explícitamente

$$\frac{dr}{\epsilon} = \frac{|q_r(k_{mo} + \epsilon) - q_r(k_{mo})|}{\epsilon} = e^{\lambda r} \quad (15)$$

De la cual su tasa de crecimiento  $\lambda$  de divergencia puede ser calculada

$$\lambda = \frac{1}{r} \ln \left( \frac{|q_r(k_{mo} + \epsilon) - q_r(k_{mo})|}{\epsilon} \right) \quad (16)$$

Estos valores de  $\lambda$  se dibujan en la Figura IV.14b.

Dado que  $\lambda$  depende de los valores iniciales, para caracterizar al atractor es mejor calcular su media

$$\lambda = \frac{1}{N} \sum_{r=1}^N \lambda_r \quad (17)$$

La ecuación (17) define al *exponente de Lyapunov*, una medida para la divergencia de dos trayectorias distintas aunque cercanas, y una cantidad para caracterizar al caos. Si es negativo, las distancias  $d_r$  decrecerán y entonces corresponderán con un sistema cuyos parámetros alcanzan un estado estable, con un valor estable convergente (Figuras III.3 - III.5) o con un comportamiento cíclico (Figuras III.6 - III.8). Si es positivo (Figuras III.9 y III.10) se tiene un sistema caótico y, por tanto, un atractor extraño. De hecho  $\lambda_{N=22} = 0.3022$ .

Textos tanto teóricos como técnicos (Cal y Mayor y Cárdenas, 2007; Board, 1994) describen diversos comportamientos en un mismo camino dependiendo de las condiciones que lo afecten, o diferentes valores de densidad, flujo y velocidad para caminos con diferentes geometrías. De esta forma, es bastante común encontrar situaciones en tramos de carreteras reales donde las condiciones del tránsito se vuelven estados estables convergentes o cíclicos en una forma predecible, situaciones que reportan condiciones de paro y arranque para los vehículos, y también casos donde valores de tránsito libre y congestionado se alternan en una naturaleza cuasi-periódica. Las simulaciones aquí mostradas son aproximaciones cualitativas de tales fenómenos vehiculares. Análisis cuantitativos más profundos en datos de tráfico dan razones para extender estos resultados hacia trabajos relacionados.

## CONCLUSIONES

El tránsito o tráfico vehicular es modelado como un sistema no lineal, lo cual puede ser efectuado a través de esquemas microscópicos y macroscópicos, siendo los primeros los relacionados con variables que tiene que ver con unidades individuales, mientras que los segundos se relacionan con cantidades promedio, especialmente la densidad  $k$ , el flujo  $q$  y la velocidad  $v$ . A diferencia de otros trabajos que se pueden encontrar en la literatura, éste presenta análisis acerca del comportamiento del tráfico como un sistema macroscópico no lineal basado en los conocidos diagramas fundamentales que relacionan a dichas variables por parejas.

Existen diferentes modelos que dan detalle de la relación entre estas variables macroscópicas, como es el de Greenberg, el cual se encuentra bastante bien pro-

bado para datos reales, y aun cuando no se ajusta satisfactoriamente para velocidades en régimen de tráfico libre, se ha probado en un rango amplio de valores prácticos, además de ser fácilmente derivado de expresiones básicas, haciéndolo ideal para llevar a cabo un análisis sobre él.

Una de las principales contribuciones de este escrito es la de incluir los diagramas fundamentales de flujo-densidad y de velocidad-densidad, para llevar a cabo las respectivas iteraciones y análisis en ellos, en lugar de usar solamente el diagrama fundamental flujo-densidad.

La normalización se efectúa en las variables densidad  $k$  y flujo  $q$  con tal de generalizar resultados. La velocidad  $v$  se obtiene directamente de la relación entre las variables de tráfico. Del modelo de Greenberg es posible manipular la velocidad óptima  $v_o$  como el parámetro que escala los diagramas fundamentales, lo cual equivale a cambiar las condiciones que afectan la capacidad de la carretera hipotética, el valor de velocidad máxima que en ella se podría desarrollar, y otras características.

Las iteraciones en los valores de densidad  $k$  sustituidos en el modelo de Greenberg para obtener el flujo  $q$  son ejecutadas, y mediante la normalización efectuada, dichos valores son tomados directamente como nuevos valores de densidad, los cuales son otra vez sustituidos para lograr nuevos. Este conjunto de valores es dibujado con la ayuda de una línea recta  $q = k$ , la cual emula la evolución en el tiempo de estos valores de tráfico.

Cuando  $v_o$  es cambiada de valores bajos a altos, las iteraciones dibujan trayectorias que convergen a valores estables, cíclicos o que muestran resultados caóticos. Es posible obtener un mapa de los valores finales luego de cierto número de iteraciones para cada modificación hecha a  $v_o$ . Dado que esta clase de mapeo muestra los puntos y las divisiones de las trayectorias de valores oscilatorios con un cierto periodo, se les conoce como mapas de bifurcaciones. Es fácil reconocer en estas gráficas cuáles rangos de valores observan un comportamiento estable, cíclico o caótico. Otro punto importante abordado en este documento es mostrar el mapa de bifurcaciones elaborado con la velocidad  $v$  y la velocidad óptima  $v_o$ , obtenido a través de los esquemas de iteraciones aquí efectuados.

La característica principal de las bifurcaciones es que pueden dividirse en ciclos de periodo  $2^n$ , es decir, para la primera ocasión en la que ello ocurre, el mapa se divide en dos ramas, luego se divide en cuatro, después en ocho, en dieciséis y así sucesivamente. Pero esto también se hace en intervalos cada vez más cortos y no puede continuar indefinidamente. Pronto, el periodo  $2^n$  se yuxtapone al siguiente  $2^{n+1}$  y cualquier variación en el parámetro  $v_o$  o en las condiciones iniciales dará una situación inestable entre las ramas, lo cual implica grandes diferencias entre dos estados finales para valores iniciales muy cercanos de tales estados.

Esta sensibilidad a las condiciones iniciales es una característica bastante conocida en los sistemas caóticos. Dos trayectorias de valores que comienzan desde condiciones iniciales muy cercanas divergen entre sí. Una medida de dicha divergencia es el exponente de Lyapunov, el cual es negativo para trayectorias estables o cíclicas, pero es positivo para las caóticas.

El tráfico es un sistema no lineal, el cual puede ser analizado en la forma cualitativa presentada en este documento. Para un estudio más cuantitativo es necesario tratar datos de tráfico con herramientas de análisis de señales, análisis fractal y otras técnicas relacionadas, una tarea que da pie a otros trabajos.

## REFERENCIAS

- Board T. R. (1994). *Highway Capacity Manual*, 3a ed., National Research Council, Washington, D.C., Special Report 209.
- Cal y Mayor, R. y Cárdenas, J. (2007). *Ingeniería de tránsito*, 8a ed., Alfaomega, México.
- Daganzo, C. F. (1994). "The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory", *Transportation Research B*, vol. 28B, núm. 4, pp. 269-287.
- Greenberg, H. (1959). "An Analysis of Traffic Flow", *Operations Research*, vol. 7, núm. 1, pp. 79-85, enero-febrero.
- Hillborn, R. C. (2000). *Chaos and Nonlinear Dynamics*, 2a ed., Oxford University Press, Oxford.
- Kim, T. y Zhang, H. M. (2004). "An Empirical Study on Gap Time and its Relation to the Fundamental Diagram of Traffic Flow", en *VII International IEEE Conference on Intelligent Transportation Systems*, Washington, D.C., pp. 94-99.
- Lighthill, M. J. y Whitham, G. B. (1955a). "On Kinematic Waves. I. Flood Movement in Long Rivers", *Proc. Royal Soc. A*, vol. 229, pp. 281-316.
- (1955b). "On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads", *Proc. Royal Soc. A*, vol. 229, pp. 317-345.
- Lo, S. C. y Cho, H. J. (2005). "Chaos and Control of Discrete Dynamic Model", *Journal of the Franklin Institute*, vol. 342, pp. 839-851.
- Lorenz, E. N. (1996). *The Essence of Chaos*. University of Washington Press, Washington, D. C.
- Low, D. J. y Addison, P. S. (1998). "A Nonlinear Temporal Headway Model of Traffic Dynamics", *Nonlinear Dynamics*, vol. 16, núm. 2, pp. 127-151.
- Lučić, P. y Teodorović, D. (2002). "Transportation Modeling: An Artificial Life Approach", en *XIV IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2002*, 4-6 noviembre, pp. 216-223.

- May, A. D. (1990). *Traffic Flow Fundamentals*. Prentice Hall, Nueva Jersey.
- Richards, P. I. (1956). "Shock Waves on the Highway", *Operation Research*, vol. 4, pp. 42-51.
- Rosas-Jaimes, O. y Álvarez, L. (2007). "Traffic Density and Velocity Estimation", *Nonlinear Dynamics*, vol. 49, núm. 4, pp. 555-566.
- Thamizh V. y Dhivya, G. (2010). "Measuring Heterogeneous Traffic Density", *International Journal of Engineering and Applied Sciences*, vol. 6, núm. 3, pp. 144-148.
- Treiber, M. y Helbing, D. (2001). "Microsimulations of Freeway Traffic Including Control Measures", *Automatisierungstechnik*, vol. 49, pp. 478-484.
- Villalobos, J., Toledo, B. A., Pastén, D., Muñoz, V., Rogan, J., Zarama, R., Lammoglia, N. y Valdivia, J. A. (2010). "Characterization of the Nontrivial and Chaotic Behavior that Occurs in a Simple City Traffic Model", *Chaos*, vol. 20.
- Wohl, M. y Martin, B. (1967), *Traffic System Analysis*, McGraw-Hill (Series in Transportation), Nueva York.
- Zhang, X. y Jarret, D. (1997). "Stability Analysis of the Classical Car-Following Model", *Transportation Research B*, vol. 31, núm. 6, pp. 441-462.

# DINÁMICA DE LOS PRECIOS FORWARD COMO PROCESO ESTOCÁSTICO CON REVERSIÓN A LA MEDIA

*Vladimir G. Makhankov\**  
*Ma. de Lourdes Nájera-López\*\**  
*Máximo A. Agüero\*\*\**

## Resumen

Se consideran procesos estocásticos con reversión a la media y se construyen modelos auto-consistentes en la dinámica de precios forward y algunas aplicaciones en la industria. Estos modelos están contruidos con conceptos y ecuaciones de geometría diferencial estocástica, con el fin de crear un sistema de ecuaciones de forma cerrada para los precios forward y su volatilidad. Algunas de las soluciones analíticas se presentan para el caso de un factor y en específico para la relación del precio forward/volatilidad de las tasas de interés. Estos modelos también desempeñan una función en las condiciones iniciales para un proceso estocástico que describe el precio forward y la volatilidad de las tasas de interés.

**Palabras clave:** Geometría diferencial estocástica, procesos estocásticos con reversión a la media y estructura temporal de instrumentos específicos de economía y finanzas.

\* 4148 Chaparron Pl. Santa Fe, Nuevo México 87507, Estados Unidos, <vladgala@hotmail.com>. BusinessMath, Santa Fe, Nuevo México, Estados Unidos.

\*\* Estudiante del Doctorado en Ciencias Económico-Administrativas, UAEM. Investigación como estudiante adjunta a la Facultad de Ciencias, Plantel Nezahualcóyotl de la Escuela Preparatoria, Universidad Autónoma del Estado de México, Instituto Literario 100, Toluca 50000, México, <mlnajeral@uaemex.mx>.

\*\*\* Investigación de estancia sabática en el Laboratorio Nacional de Los Álamos, T-División, Nuevo México, Estados Unidos. Facultad de Ciencias, Universidad Autónoma del Estado de México, Instituto Literario 100, Toluca 50000, México.

## INTRODUCCIÓN

En este trabajo se presentan modelos auto-consistentes para describir y analizar dos diferentes instrumentos económicos que juegan un papel fundamental en la evaluación correspondiente de derivados y, por ende, en la gestión de riesgos. Ambos tienen la misma estructura temporal, es decir, vencimiento del contrato en un tiempo definido. El primer instrumento son las tasas de interés de bonos de descuento en el mercado financiero, y el segundo son los precios forward de una mercancía. Con base en las ecuaciones de la geometría diferencial estocástica (DSG), que describen la dinámica del precio spot/tasa de interés y la de largo plazo (todas las curvas de los precios forward/tasas de interés), se construyen modelos matemáticos que resultan ser muy similares.

El presente estudio, al menos en su parte económica, se ha visto influenciado por las obras de varios autores: Cortázar y Schwartz (1994), Schwartz (1997), Hillard y Reis (1998), Clewlow y Strickland (1999), por mencionar algunos. Más tarde se dan referencias más detalladas en los apartados oportunos.

El trabajo se organiza de la siguiente manera: En la sección I se da una breve introducción a la geometría diferencial estocástica, es decir, el Movimiento Browniano geométrico para múltiples curvas (por ejemplo, una esfera), los procesos de Makhankov (1995 y 1997), que permiten tener en cuenta el comportamiento estocástico de los precios forward/volatilidad de las tasas de interés y cerrar el sistema de ecuaciones. La sección II presenta el estudio de la dinámica de los precios forward y las posibles soluciones para la curva forward, su media y su volatilidad. La sección III se dedica al estudio de la dinámica de las tasas de interés y algunas soluciones. Finalmente, la sección IV contiene las conclusiones.

### I. ECUACIONES DE LA GEOMETRÍA DIFERENCIAL ESTOCÁSTICA

Para una mejor comprensión de los resultados obtenidos se expone en forma sucinta los conceptos claves y el formalismo matemático de los procesos estocásticos necesarios en el desarrollo del trabajo.

Considerando solamente un movimiento browniano en una esfera primero se construye un “frame bundle” sobre la misma, considerando un punto  $X_1$  en  $S^2$  y una fracción de un plano tangente en este punto, que se denota como  $T_{X_1}S^2$ . A esto se le denomina fibra en el punto  $X_1$ . A continuación, se procede con un punto próximo  $X_2$  y, haciendo lo mismo, se tiene  $T_{X_2}S^2$ . De este modo, se puede resguardar todo el ámbito con estos fragmentos que se pegan a lo largo de las líneas de sus intersecciones, como un balón de fútbol que da un poliedro. Así se tiene un

ejemplo del haz fibrado con la esfera, siendo la base de la misma y el poliedro un paquete de fibras (de hecho, un paquete de marcos en este caso). Llamando a este poliedro “cubierta” de la esfera. Así que finalmente se tiene:

- a) la esfera que está curvada y
- b) el revestimiento es un espacio euclideo.

Un proceso de Wiener, es decir, una martingala, que satisface las ecuaciones

$$\langle dW^q dW^p \rangle = \delta^{qp} dt \tag{I.1}$$

se produce en el mundo euclideo. De igual manera, se lleva a cabo para la semi-martingalas (aproximadamente), pues sólo en un espacio euclideo es posible representar un proceso estocástico, en forma semi-martingala

$$d\tilde{W}^q = \alpha^q dt + dW^p \tag{I.2}$$

Esto significa que el proceso de Wiener sólo puede aparecer en la cubierta, mientras que una partícula se mueve en la esfera. Ahora se deben ajustar ambos fenómenos, considerando una “perturbación” con las fuerzas fluctuantes (procesos de Wiener) y una partícula en el punto  $X_1$  de la esfera; este punto también pertenece a la superficie  $S^2$ , de ahí que la partícula se somete a un choque aleatorio, como:

$$d\vec{X}_1 = \hat{\sigma}_1 d\tilde{W}_1 \tag{I.3}$$

y salta a un punto  $X_2$  en la esfera, en este punto de nuevo se somete a un choque

$$d\vec{X}_2 = \hat{\sigma}_2 d\tilde{W}_2$$

y así sucesivamente. La matriz  $\hat{\sigma}$  define la movilidad de las partículas (sensibilidad). Aquí hay que destacar que todas las diferenciales mencionadas arriba, son del tipo Stratonovich, que permite utilizar el cálculo diferencial estándar.

Ahora se supone que todos  $d\vec{W}_i$  son iguales en la distribución (movimiento browniano); además, se deben relacionar  $\vec{\sigma}_i$  y  $\vec{\sigma}_2$ . Nótese que la matriz  $\hat{\sigma}$ , siendo de hecho un operador rotatorio, puede ser construida a partir de dos vectores  $\vec{j}_i$  y  $\vec{j}_2$ :

$$\hat{\sigma} = \begin{pmatrix} j_1^x & j_2^x \\ j_1^y & j_2^y \end{pmatrix} \quad (\text{I.4})$$

lo que da un marco natural en la fracción de la superficie.

Mientras se mueven de un espacio a otro, este marco cambia su orientación. Así que el cambio total de un vector  $\vec{B}$ , debido a que se mueve de un punto a otro, se compone de dos partes

$$\vec{\delta B} = d\vec{B} + \Gamma \vec{B} dX \quad (\text{I.5})$$

Donde el primer término es el diferencial a lo largo de la trayectoria

$$d\vec{B}(t) = \frac{\partial \vec{B}}{\partial X^i} dX^i(t)$$

y el segundo permite un cambio de la orientación del marco. Dado que la matriz  $\hat{\sigma}$  consta de  $n$  vectores, siguiendo la misma regla se transforma en:

$$\delta \hat{\sigma} = d\hat{\sigma} + \vec{\Gamma} \hat{\sigma} dX(t) \quad (\text{I.6})$$

El siguiente supuesto es que “Las reglas no deben cambiar de juego a juego” (“fair game” o ninguna oportunidad de arbitraje), significa que el cambio total de  $\hat{\sigma}$  debe desaparecer

$$\delta \hat{\sigma} = 0$$

o

$$d\hat{\sigma} = -\vec{\Gamma} \hat{\sigma} dX(t) \quad (\text{I.7})$$

junto con la ecuación para el elemental shock

$$dX(t) = \hat{\sigma} dW \quad (\text{I.8})$$

lo anterior da las ecuaciones de la geometría diferencial estocástica de la esfera. La generalización a otras variedades curvas es sencilla.

## II. PRECIOS FORWARD Y SU MODELAJE

Para una mejor comprensión se deben tener presentes los siguientes conceptos.

**Definición 1.** Un contrato forward es un derivado particularmente simple y es un acuerdo para comprar (posición larga) o vender (posición corta) un activo subyacente en una fecha futura pactada (fecha de vencimiento,  $T$ ) para un determinado precio (el precio de entrega,  $k$ ).

En el momento del contrato se indica el precio de entrega, que debe ser tal que el valor del contrato; para ambas partes sea igual a cero. El contrato es obligatorio.

**Definición 2.** El precio forward  $F(t, T)$  para un contrato determinado se define como el precio de entrega que hará que el contrato tenga un valor cero.

El precio forward y el precio de entrega son iguales en el momento de la celebración del contrato; conforme pasa el tiempo, están a solas desde tiempos pre-específicos; inicialmente la entrega de precios es constante. Por tanto, se puede pensar en el precio forward como el precio de entrega en el momento actual  $t$ .

Se presenta la expresión que relaciona el precio forward y el precio spot (Hull, 1993):

$$F(t, T) = S(t) e^{(r+u-y)(T-t)} \quad (\text{II.1})$$

donde  $r$  es la tasa de interés libre de riesgo,  $u$  es la tasa stock por unidad (es una proporción constante), mientras que  $y$  es el rendimiento promedio anual. Tanto la tasa de stock como el rendimiento son por lo general funciones desconocidas, pero este último puede ser un proceso estocástico.

Siguiendo a Cortázar y Schwartz (1994), se describirá la dinámica de los precios forward mediante la ecuación

$$\frac{dF(t, T)}{F} = \sum_{p=1}^n \Sigma_p(t, T) dW^p(t) \quad (\text{II.2})$$

siendo  $\Sigma_p(t, T)$  la volatilidad correspondiente a un factor aleatorio  $p$ -ésimo descrito por un proceso de Wiener  $dW^p(t)$ . Entonces el modelo (II.2) describe la dinámica  $n$ -factor de la curva forward  $F(t, T)$ .

Asumiendo que las tasas de interés son deterministas y los precios futuros son iguales a los precios forward (véase, por ejemplo, Hull, 1993). En (II.2) se tiene  $n$  fuentes independientes de incertidumbre que impulsan la evolución de la curva forward  $F(t, T)$ .

Mediante la integración de (II.2), usando el lema de Ito, se obtiene:

$$F(t, T) = F(0, T) \exp \left\{ \sum_{j=1}^n \left[ -\frac{1}{2} \int_0^t \Sigma_j(\tau, T)^2 d\tau + \int_0^t \Sigma_j(\tau, T) dW^j(\tau) \right] \right\} \quad (\text{II.3})$$

Entonces, para el precio de contado, por definición es  $S(t) = F(t, T)$  y con  $T=t$  se tiene:

$$S(t) = F(0, t) \exp \left\{ \sum_{i=1}^n \left[ -\frac{1}{2} \int_0^t \Sigma_i(\tau, t)^2 d\tau + \int_0^t \Sigma_i(\tau, t) dW^i(\tau) \right] \right\} \quad (\text{II.4})$$

Esto significa que el logaritmo natural del precio spot (al igual que el precio forward) se distribuye normalmente en el tiempo  $T$ , dado el precio forward en el momento cero, de tal manera que:

$$\ln(S(T)) \approx N \left[ \left\{ \ln(F(0, T)) - \frac{1}{2} \sum_{i=1}^n \int_0^T \Sigma_i(\tau, T)^2 d\tau \right\}, \sum_{i=1}^n \int_0^T \Sigma_i(\tau, T)^2 d\tau \right]$$

Al diferenciar (II.4) en  $\tau$  se tiene la ecuación diferencial estocástica para el precio spot

$$\frac{dS(t)}{S(t)} = \left\{ \frac{\partial \ln F(0, t)}{\partial t} - \sum_{i=1}^n \left[ \int_0^t \Sigma_i(\tau, t) \frac{\partial \Sigma_i(\tau, t)}{\partial t} d\tau - \int_0^t \frac{\partial \Sigma_i(\tau, T)}{\partial t} dW^i(\tau) \right] \right\} dt \quad (\text{II.5})$$

$$+ \sum_{i=1}^n \Sigma_i(t, t) dW^i(t)$$

El término entre llaves es interpretado como el equivalente de la suma determinista de la tasa de interés  $r(t)$  libre de riesgo y un rendimiento  $y(t)$ , *que en general debe ser estocástico*. Muchos modelos conocidos no son más que casos especiales de este enfoque general.

Ahora se conoce que las volatilidades en (II.2) son procesos estocásticos. ¿Qué tipo de procesos estocásticos podrían ser? Para responder a esta pregunta se puede recurrir a la geometría diferencial estocástica (Makhankov, 1995; 1997) que se ha descrito anteriormente. Como resultado se obtiene un modelo auto-consistente descrito por un sistema de ecuaciones estocásticas. Para resolverlo se deben especificar las condiciones iniciales:  $F(0,T)$ ,  $S(0)$  y  $\Sigma_i(0,T)$ .

Por tanto, la dinámica del logaritmo del precio forward está dada por la ecuación (II.6) en los diferenciales de Ito, tal como se muestra:

$$dLn(F(t,T)) = -\frac{1}{2} \sum_{q=1}^n \Sigma_q^2(t,T)dt + \sum_{q=1}^n \Sigma_q(t,T)dW^q(t) \quad (II.6)$$

Si se asume que el término espacio interno del modelo debe ser discreto (lo que es verdadero en realidad) por:

$$F(t,T) = F(t,kT) = F^k(t)$$

ahora denotando

$$X^k(t) = \ln F^k(t) \quad (II.7)$$

se llega a la ecuación

$$dX^i(t) = -\frac{1}{2} \sum_{q=1}^n \Sigma_q^{i^2}(t)dt + \sum_{q=1}^n \Sigma_q^i dW^q(t), \quad i \in (1, \dots, m) \quad (II.8)$$

empleando los diferenciales de Ito.

Si se quiere que el modelo dinámico del precio forward corresponda al movimiento browniano puro en el colector de curvas, se debe igualar la ecuación (II.7) con la primera ecuación del sistema (II.1), dado que también se enuncian en los diferenciales de Ito

$$dX^i(t) = \sum_{q=1}^n \left( \frac{1}{2} d\Sigma_q^i + \Sigma_q^i \right) dW^q(t) = -\frac{1}{2} \sum_{j,k,q} \Gamma_{jk}^i \Sigma_q^k \Sigma_q^j dt + \sum_{q=1}^n \Sigma_q^i dW^q(t) \quad (II.9)$$

Por lo que se tiene la ecuación para la auto-consistencia del modelo

$$\sum_{q=1}^m \Sigma_q^{i^2}(t)dt = \sum_{q=1}^n \sum_{j,k} \Gamma_{jk}^i \Sigma_q^k \Sigma_q^j dt \quad (II.10)$$

Resolviendo la ecuación respecto a

$$A_{k,q}^i = \Gamma_{j,k}^i \Sigma_q^j$$

se obtiene

$$A_{k,q}^i = \sum_j \delta_k^j$$

Sustituyendo esta ecuación en la ecuación estocástica para precios “forward”

$$dX^i(t) = f^i(t) + \sum_{p=1}^n \sigma_p^i(t) dW(t)$$

se llega a

$$d\Sigma_p^i(t) = -\Sigma_p^i \sum_{q=1}^n \Sigma_q^i dW^q(t) \tag{II.11}$$

Ahora el sistema es cerrado y consistente, ya que la curvatura del espacio se define por el “término de fuerza” (la tendencia) en la ecuación de la dinámica de los precios forward.

Resolviendo el siguiente sistema de ecuaciones:

$$dX^i(t) = \sum_{q=1}^n \Sigma_q^i dW^q(t)$$

$$d\Sigma_p^i(t) = -\Sigma_p^i \sum_{k=1}^n \Sigma_k^i dW^k(t)$$

donde los diferenciales de Ito son

$$dX^i(t) = -\frac{1}{2} \sum_{q=1}^n \Sigma_q^{i2}(t) dt + \sum_{q=1}^n \Sigma_q^i dW^q(t)$$

$$d\Sigma_q^i = \Sigma_q^i \sum_p (\Sigma_p^i)^2 dt - \Sigma_q^i \sum_p \Sigma_p^i dW^p \tag{II.12}$$

De la primera ecuación de (II.12) se infiere que  $X$  tiene una distribución normal. ¿Qué pasa con la volatilidad?

### III. REDUCCIÓN A MODELO DE UN FACTOR

Considerando ahora una reducción del modelo a un solo factor, esto tiene sentido ya que, como es bien conocido, algunos modelos pueden ser los mercados de energía, así como los financieros que muestran comportamiento de modelos de un solo factor, donde el análisis de los componentes principales da entre

80% hasta 90% de la contribución total al componente principal de acuerdo con Clewlow y Strickland (1999), y Wilmott (2001).

Por tanto, para un solo factor se tiene  $dW^p = dW$  y  $\Sigma^i = Y$  entonces

$$dY = Y^3 dt - Y^2 dW \tag{III.1}$$

y la ecuación de Fokker-Plank para la ecuación de la probabilidad de transición  $\rho$  se enuncia como:

$$\partial_t \rho = \partial_y (-y^3 + \frac{1}{2} \partial_y y^4) \rho \tag{III.2}$$

Ahora se consideran soluciones estacionarias de (III.2). Entonces se tiene que

$$\partial_y (-y^3 + \frac{1}{2} \partial_y y^4) \rho = 0$$

con una solución

$$\rho = \frac{cy + a}{y^3} \tag{III.3}$$

Tomando en cuenta el proceso relacionado con Stratonovich

$$d\Sigma_q^i = -\Sigma_q^i \sum_p^n \Sigma_p^i dW^p$$

y para un solo factor de un solo término el proceso  $Z = \Delta^i$  es

$$dZ = -Z^2 dW \tag{III.4}$$

con la ecuación de Fokker-Planck

$$\partial_t \rho = \frac{1}{2} \partial_z (z^2 \partial_z z^2) \rho$$

y las soluciones estacionarias

$$\rho = \frac{a}{z^3} \tag{III.5}$$

Por lo que se observa, ambos procesos tienen una distribución similar, y es lo suficientemente pequeño

$$y \ll \frac{a}{c}$$

tal como se puede ver, lo anterior es un proceso de Stratonovich.

Ahora, por medio del proceso de Ito se tiene la ecuación (II.8), donde la media puede estimarse a partir del término de tendencia por el siguiente razonamiento: tomando la media de dicha ecuación se tiene que para  $\Sigma = \langle \Sigma \rangle + s$  y  $\langle s \rangle = 0$  es

$$\langle d\Sigma \rangle = \langle \Sigma^3 \rangle = \langle S + \langle \Sigma \rangle \rangle^3 = 3 \langle S^2 \rangle \langle \Sigma \rangle + \langle \Sigma \rangle^3$$

donde la varianza es  $\langle s^2 \rangle \approx t \langle \Sigma \rangle^4$  y el primer término de la ecuación puede ser omitido. Ahora bien, como  $\langle d\Sigma \rangle = \langle \Sigma_{t+1} - \Sigma_t \rangle = d \langle \Sigma \rangle$  entonces se llega a

$$d \langle \Sigma \rangle = \langle \Sigma \rangle^3 dt$$

con la solución

$$\langle \Sigma \rangle = \sigma \sqrt{\frac{1}{1 - t\sigma^2}} \approx \sigma \left( 1 + \frac{1}{2} t\sigma^2 \right)$$

tomando lo anterior, se puede calcular la curva forward. Con el fin de obtener estimación analítica, se limita a un horizonte de “corto plazo”:

$$\sigma^2 t \ll 1 \tag{III.6}$$

y considerando sólo los dos primeros términos iniciales de la expansión asintótica. Lo anterior contempla las propiedades estadísticas del modelo y sus horizontes de corto plazo. A continuación, se estudia la dinámica del modelo con más detalle.

#### IV. ECUACIONES DE LA GEOMETRÍA DIFERENCIAL ESTOCÁSTICA EN EL PRECIO FORWARD Y VOLATILIDAD

Analizando la versión del modelo de un solo factor (II.2), para  $F(t, T)$  se tiene

$$\frac{dF(t, T)}{F(t, T)} = \Sigma(t, T) dW(t) \quad \text{y} \quad F(t, t) = S(t) \tag{IV.1}$$

y utilizando el lema de Ito

$$d \ln F(t, T) = -\frac{1}{2} \Sigma^2(t, T) dt + \Sigma(t, T) dW(t)$$

integrando una vez, se obtiene

$$\text{Ln} \frac{F(t,T)}{F(0,T)} = -\frac{1}{2} \int_0^t \Sigma^2(u, T) du + \int_0^t \Sigma(u,T) dW(u)$$

o

$$F(t,T) = F(0,T) \exp \left\{ -\frac{1}{2} \int_0^t \Sigma^2(u, T) du + \int_0^t \Sigma(u,T) dW(u) \right\} \quad (\text{IV.2})$$

Esta solución se define de manera precisa con una exactitud de una función arbitraria  $F(0,T)$ . Para limitar esta libertad, se puede especificar un proceso aleatorio para el precio spot  $S(t)$ . Ahora bien, como  $S(t) = F(t,t)$  y sabiendo que la ecuación  $S(t)$  proporciona la  $F(0,T)$  a través de los parámetros que intervienen en la ecuación  $S(t)$ . Tomando ahora un proceso de reversión a la media para  $S(t)$ , es decir:

$$\frac{dS}{S} = \alpha(\mu - \ln S) dt + \sigma(t) dW(t) \quad (\text{IV.3})$$

Este tipo de procesos es muy popular en la econometría, ya que, por ejemplo, los precios forward, así como las tasas de interés tienden hacia su valor esperado a largo plazo. Este fenómeno se conoce como reversión a la media (Hull, 1993: 388). También el modelo de un factor con reversión a la media de Schwartz (1997) es utilizado para la dinámica de los precios; de hecho, la ecuación del tipo (IV.3) apareció en la física desde hace mucho tiempo y se supone que describe el llamado proceso de Ornstein-Uhlenbeck.

Por otro lado, la ecuación (IV.2) proporciona

$$S(t) = F(0,t) \exp \left\{ -\frac{1}{2} \int_0^t \Sigma^2(u, t) du + \int_0^t \Sigma(u,t) dW(u) \right\} \quad (\text{IV.4})$$

Donde  $\ln S$  se distribuye normalmente con

$$\text{Media} = \text{Ln } F - \frac{1}{2} \int_0^t \Sigma^2(u,T) du$$

$$\text{Distribución} = \Sigma(t,t)$$

A partir de la ecuación (IV.4), aplicando logaritmos se obtiene

$$\text{Ln } S(t) = \text{Ln } F(0,t) - \frac{1}{2} \int_0^t \Sigma^2(u,t) du + \int_0^t \Sigma(u,t) dW(u) \quad (\text{IV.5})$$

A continuación, diferenciando respecto a  $t$  se tiene

$$dLn S(t) = \left[ \frac{\partial Ln F(0,t)}{\partial t} - \frac{1}{2} \Sigma^2(t,t) - \int_0^t \Sigma(u,t) \Sigma_t(u,t) du + \int_0^t \Sigma_t(u,t) dW(u) \right] dt + \Sigma(t,t) dW(t)$$

Aplicando el lema de Ito se comprueba que

$$dLn S(t) + \frac{1}{2} \Sigma^2(t,t) dt = \frac{dS}{S}$$

por tanto

$$\frac{dS(t)}{S(t)} = \left[ \frac{\partial Ln F(0,t)}{\partial t} - \int_0^t \Sigma(u,t) \Sigma_t(u,t) du + \int_0^t \Sigma_t(u,t) dW(u) \right] dt + \Sigma(t,t) dW(t) \tag{IV.6}$$

Ahora bien, si en lugar del proceso del activo subyacente la dinámica del precio forward se define por la ecuación (IV.3) se obtiene el sistema auto-consistente de ecuaciones, tales como:

$$\sigma(t) = \Sigma(t,t) \tag{IV.7a}$$

$$\alpha(\mu - Ln S) = \frac{\partial Ln F(0,t)}{\partial t} - \int_0^t \Sigma(u,t) \Sigma_t(u,t) du + \int_0^t \Sigma_t(u,t) dW(u) \tag{IV.7b}$$

Reescribiendo la ecuación (II.23) de la forma

$$\int_0^t \Sigma(u,t) dW(u) = \{ Ln S(t) - Ln F(0,t) \} + \frac{1}{2} \int_0^t \Sigma^2(u,t) du \tag{IV.8}$$

Fácilmente se puede resolver el sistema de ecuaciones (IV.6), (IV.7) y (IV.8) si

1) El primer grupo de modelos es

$$\partial_t \Sigma_1(u,t) = -\alpha \Sigma_1(u,t) \tag{IV.9}$$

para el caso más general

2) El segundo grupo de modelos es

$$\partial_t \Sigma_2(u,t) = -\alpha \Sigma_2(u,t) + f(t) \tag{IV.10}$$

donde  $f(t)$  es una función conocida de  $t$ . Esas condiciones permiten cancelar integrales estocásticas en las ecuaciones (similar a la condición del menor riesgo)

y son necesarias para dar solución a todo el problema. Lo que significa que la volatilidad de los precios forward decae de un nivel a otro, o cero.

Para el primer grupo de modelos, si se sustituye la ecuación (IV.9) en (IV.7) y usando (IV.8) se llega a

$$\frac{\partial \text{Ln}F_1(0,t)}{\partial t} + \alpha \text{Ln}F_1(0,t) = \alpha \left( \mu - \frac{1}{2} \int_0^t \Sigma_1^2(u,t) du \right) \equiv \Phi_1(t) \tag{IV.11}$$

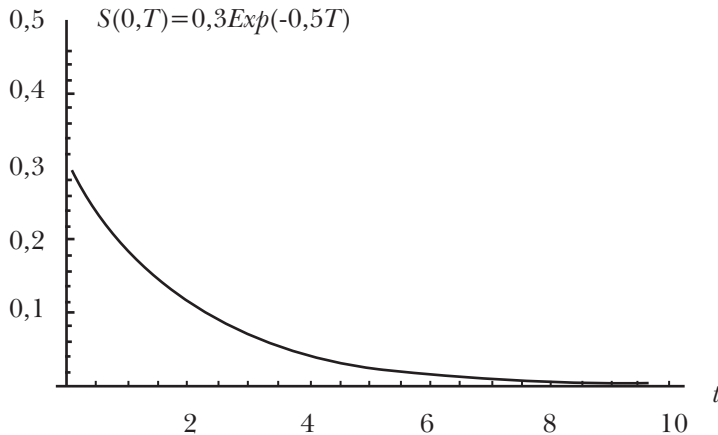
Para el caso general (II.28) se tiene

$$\Phi_2(t) = \alpha \left( \mu - \frac{1}{2} \int_0^t \Sigma_2^2(u,t) du \right) + \int_0^t \Sigma_2(u,t) f(u) du - \int_0^t f(u) dW(u) \tag{IV.12}$$

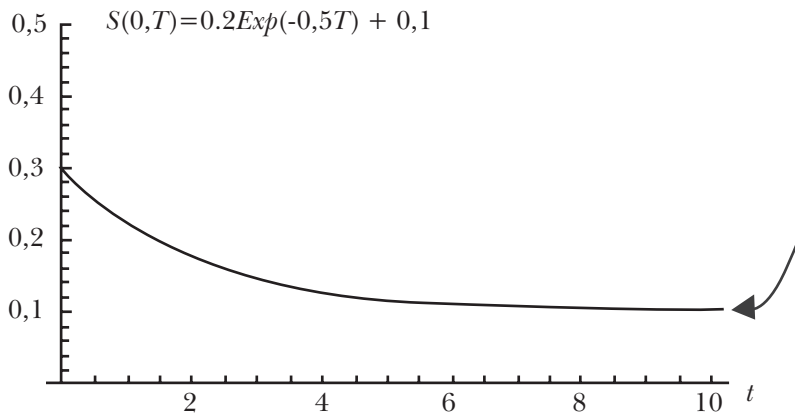
La ecuación (IV.11) junto con la (IV.12) puede resolverse fácilmente sabiendo la volatilidad de los precios forward  $\Sigma(t,T)$ .

A continuación se muestran las curvas del modelo real y de la volatilidad:

**Figura IV.1**  
**Representación asintótica para vencimiento igual a cero**

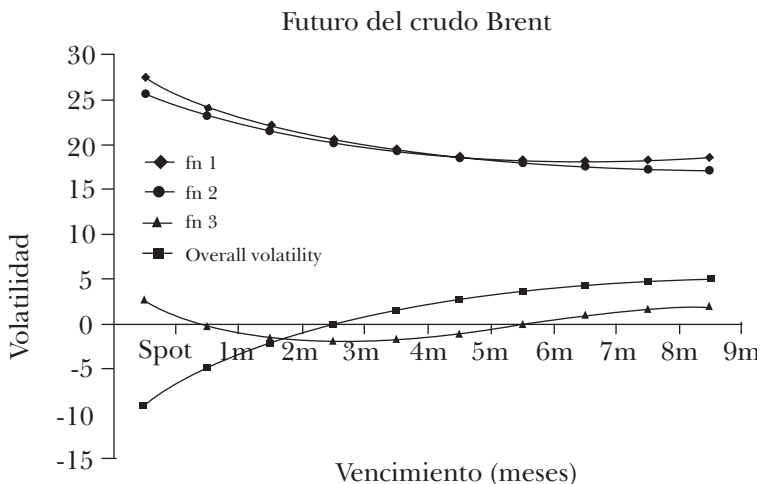


**Figura IV.2**  
**Representación asintótica para vencimiento diferente de cero**

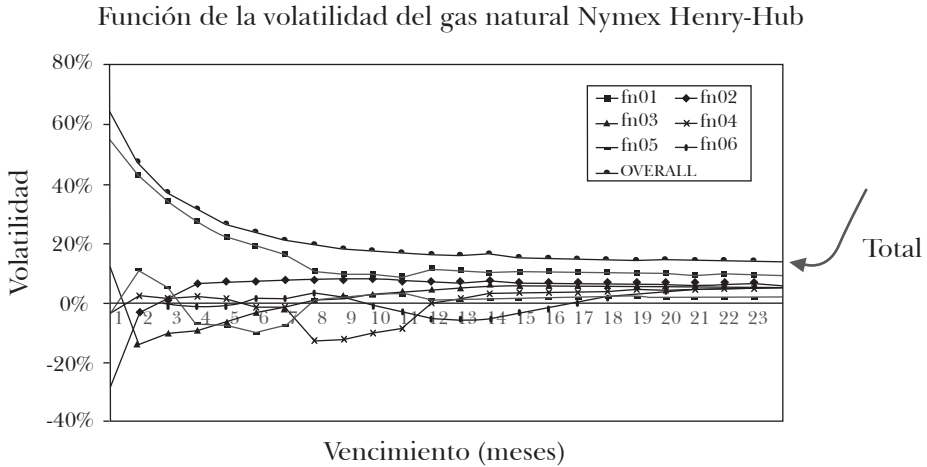


En realidad, se tiene la situación muy cerca de la segunda curva (ver Clewlow y Strickland, 1999).

**Figura IV.3**  
**Análisis de los principales componentes del petróleo crudo**



**Figura IV.4**  
**Análisis de los principales componentes del gas natural**



La solución de (IV.11) (obtenida por el método de varianza constante) es

$$\text{Ln } F_i(0,t) = e^{-\alpha t} \left\{ \int_0^t e^{\alpha u} \Phi_i(u) du + \text{const} \right\}$$

ya que en  $t = 0$ ,  $\text{Ln } F_i(0,0) = \text{Ln } S_i(0)$  entonces

$$\text{Ln } F_i(0,t) = e^{-\alpha t} \left\{ \int_0^t e^{\alpha u} \Phi_i(u) du + \text{Ln } S_i(0) \right\} \tag{IV.13}$$

Donde  $\mu$  no necesariamente es una constante, puede ser una función del tiempo.

Por ejemplo: para el proceso de volatilidad de los precios forward se puede tener las siguientes posibilidades:

1) es una función regular del tiempo; por ejemplo, una primera clase de tres parámetros del modelo exponencial

$$\Sigma_1(t,T) = \sigma_1 e^{-\alpha(T-t)} \tag{IV.14}$$

y

2)  $\mu(t) = \mu_1 e^{kt}$  (la constante  $k$  puede ser positiva o negativa)

Entonces la integral se evalúa exactamente como

$$F_1(0, T) = \exp \left\{ e^{-\alpha T} \text{Ln} S_1(0) + \frac{\alpha \mu_1}{\alpha + k} (e^{kT} - e^{-\alpha T}) - \frac{\sigma_1^2}{4\alpha} (1 - e^{-\alpha T})^2 \right\} \quad (\text{IV.15})$$

Ahora mediante  $K = 0$  se llega al resultado conocido (Clewlow y Strickland, 1999):

$$F_{1R}(0, T) = \exp \left\{ e^{-\alpha T} \text{Ln} S_1(0) + \mu_1 (1 - e^{-\alpha T}) - \frac{\sigma_1^2}{4\alpha} (1 - e^{-\alpha T})^2 \right\} \quad (\text{IV.16})$$

Es fácil ver en este caso que la curva del precio forward es regular cuando la volatilidad es cero.

Se requiere de cálculos más avanzados para obtener la curva de los precios forward para el modelo con cuatro parámetros:

$$\Sigma_2(t, T) = \sigma_1 e^{-\alpha(T-t)} + \sigma_0 \quad (\text{IV.17})$$

En este caso se tiene

$$\langle \text{Ln} F_{2R}(0, T) \rangle = e^{-\alpha T} S_2(0) + \mu (1 - e^{-\alpha T}) - \frac{1}{4\alpha} [\sigma_1^2 (1 - e^{-\alpha T})^2 - 2\sigma_0^2 (\alpha T - 1 + e^{-\alpha T})] \quad (\text{IV.18})$$

y

$$\text{Var}[\text{Ln} F_{2R}(0, T)] = \frac{\sigma_0^2}{2\alpha} [2e^{-\alpha T} (1 - \alpha T - \alpha^2 T^2) - e^{-2\alpha T} (2 + \alpha^2 T^2)] \quad (\text{IV.19})$$

Es decir, en el segundo caso, por lo regular la volatilidad del proceso de los precios forward  $\Sigma(t, T)$  y la curva  $F_{2R}(t, T)$  se convierten en procesos estocásticos con volatilidad diferente de cero (IV.19).

Más cálculos avanzados son necesarios para tomar en cuenta la aleatoriedad del proceso de la volatilidad de los precios forward, para ello se resuelve la segunda ecuación de la geometría diferencial estocástica (II.12). Esta ecuación es autosuficiente y puede ser analizada por separado.

Bajo el mismo supuesto (del modelo de un solo factor) por el proceso de Stratonovich se tiene,

$$d\Sigma_{1Sr}^{sr}(t, T) = -(\Sigma_{1Sr}^{sr})^2(t, T) dW(t) \quad (\text{IV.20})$$

y por el proceso de Ito:

$$d\Sigma_{1Sr}^i = (\Sigma_{1Sr}^i)^3 dt - (\Sigma_{1Sr}^i)^2 dW \quad (\text{IV.21})$$

A partir de las expresiones (III.3) y (III.5) se puede ver que las soluciones de ambas ecuaciones de los precios forward (FP) son idénticas si  $c \ll a$ , o por ejemplo, cuando  $c = a$ . Esto significa que en este caso el proceso de distribución de Stratonovich es una subclase de la distribución más general de Ito.

A partir del proceso de Stratonovich se tiene el cálculo convencional:

$$\Sigma_{1st}^{str}(\tau, t) = \frac{\sigma(0, t)}{1 + \sigma(0, t)W(\tau)}$$

o

$$\Sigma_{1st}^{str}(t, T) = \frac{\sigma(0, T)}{1 + \sigma(0, T)W(t)} \tag{IV.22}$$

y

$$\partial_t \Sigma_{1st}^{str}(\tau, t) = \frac{\sigma_t(0, t)}{(1 + \sigma(0, t)W(\tau))^2}$$

En la ecuación (IV.22)  $W(t)$  es un proceso estándar de Wiener con media inferior y varianza igual a uno de distribución Gaussiana. Por tanto, se puede expresar  $W(t)$  en función de la  $\Sigma^i$

$$W(t) = \frac{1}{\Sigma^{str}(t)} - \frac{1}{\Sigma^{str}(0)} \tag{IV.23}$$

También es fácil calcular la media y la varianza de  $\Sigma^i(t, T)$  para  $t$  pequeño (plazos cortos)

$$\begin{aligned} E[\Sigma^i(t, T)] &\approx \sigma^i(0, T) \{1 + t\sigma^i(0, T)^2\} \\ \text{Var}[\Sigma^i(t, T)] &\approx t\sigma^i(0, T)^4 \{1 + 6t\sigma^i(0, T)^2\} \end{aligned}$$

Cabe mencionar que la serie mayor de  $t$  es asintótica y, en principio, divergente, que en estricto sentido el único punto de convergencia es  $t = 0$ , incluso sin su frontera. También, como era de esperarse, los resultados son independientes del término aleatorio.

Ahora, para el proceso de Stratonovich se puede ir aún más lejos. Con (IV.23) se calcula la distribución  $\Sigma^{str}(t, T)$  por medio de la fórmula

$$\rho(w)dW = \rho(1/\Sigma^{str}) \frac{dW}{d\Sigma^{str}} d\Sigma^{str} \equiv \rho(\Sigma^{str}) d\Sigma^{str}$$

con  $\rho(w)$  siendo una distribución Gaussiana, por lo que se tiene

$$\rho(\Sigma^{str}) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(1/\Sigma^{str} - 1\Sigma_0^{str})^2}{2t}\right) \frac{d\Sigma^{str}}{(\Sigma^{str})^2}$$

una distribución para el recíproco de  $W(t)$ .

### V. PROPIEDADES DINÁMICAS DE LAS CURVAS FORWARD

Seguidamente se aborda el estudio de las propiedades dinámicas de las curvas forward.

En el primer caso, es decir, cuando,  $\sigma(0,t) = \alpha_1 e^{-\alpha t}$ , se tiene

$$\partial_1 \Sigma_{1St}^{str}(\tau, t) = -\alpha \frac{\sigma(0, t)}{(1 + \sigma(0, t)W(\tau))^2} = -\frac{\alpha \Sigma_{1St}^{str}(\tau, t)}{(1 + \sigma(0, t)W(\tau))} \tag{V.1}$$

asumiendo que

$$\sigma(0, T)W(T) \ll 1 \tag{V.2}$$

La última condición es esencial para las evaluaciones. Esto se debe a las siguientes relaciones

$$\sigma(0, T)W(t) = \sigma_1 e^{-\alpha T} W(t) \Rightarrow^{t \leq T} \sigma_1 e^{-\alpha T} W(t) e^{-\alpha(T-t)} \leq \sigma_1 e^{-\alpha t} W(t) \approx \sigma_1 e^{-\alpha t} \sqrt{t}$$

Que se pueden inferir trivialmente, por tanto se tienen dos casos:

- 1)  $\alpha T \ll 1$  entonces  $\sigma_1^2 T \ll 1$  y
- 2)  $\alpha T \ll 1$ ,  $\sigma_1^2 T$  es arbitraria

Ahora por consiguiente se tiene el sistema de ecuaciones

$$\begin{aligned} \partial_t \Sigma_{1St} &= -\alpha \Sigma_{1St}(u, t) + f_1(u) \\ f_1(u) &= \alpha W(u) \sigma^2(0, t) [1 - 2\sigma(0, t) W(u)] \end{aligned} \tag{V.3}$$

y

$$\frac{1}{\alpha} \Phi_1(u) = \mu + 2\sigma^3(0, u) \int_0^u W(\tau) d\tau - \frac{1}{2} \sigma^2(0, u) W^2(u) - \frac{9}{2} \sigma^4(0, u) \int_0^u W^2(u) du \tag{V.4}$$

Finalmente, se puede resolver el conjunto de ecuaciones (IV.11) y (V.4) para obtener  $LnF(0, T)$  siendo un proceso estocástico con media y volatilidad, respectivamente, como se muestra a continuación

$$\begin{aligned} \langle LnF_{1St}(0,T) \rangle &= \mu(1-e^{-\alpha T}) + e^{-\alpha T} \{LnS(0) - \frac{1}{2\alpha} \sigma_1^2 [1-(1+\alpha T) e^{-\alpha T}]^2 \\ &- \frac{\sigma_1^4}{6\alpha^2} [1-(1+3\alpha T + \frac{1}{2}(3\alpha T)^2) e^{-3\alpha T}] \} \end{aligned} \quad (V.5)$$

y

$$\begin{aligned} Var[LnF_{1St}(0,T)] &= 4\alpha^2 \sigma_1^6 e^{-2\alpha T} \int_0^T e^{-2\alpha(u+z)} du \int_0^T dz \int_0^u \langle W(\tau) d\tau \int_0^z W(t) \rangle dt \\ &= \frac{\sigma_1^6}{12\alpha^2} Te^{-4\alpha T} \{-3 + 2\alpha^2 T^2 + e^{-2\alpha T} (3 + 6\alpha T + 4\alpha^2 T^2)\} \end{aligned} \quad (V.6)$$

En seguida se compara el resultado de este modelo con el caso convencional

$$\mu = const \text{ y } \Sigma_{1R}(t,T) = \sigma_1 e^{-\alpha(T-t)}$$

y

$$LnF_{1R}(0,T) = \mu(1-e^{-\alpha T}) + e^{-\alpha T} Ln S_1(0) - \frac{\sigma_1^2}{4\alpha} (1-e^{-2\alpha T})^2$$

para valores específicos de los parámetros involucrados.

Al observar (IV.4) y (IV.6) se tiene que las medias difieren en los términos proporcionales a alguna potencia de  $\sigma_1$ .

Como es preciso subrayar, la volatilidad del proceso de precio spot  $S(t)$  se define como  $\sigma(t) = \Sigma(t,t)$ , por tanto:

$$\sigma_1(t) = \frac{\sigma_1 e^{-\alpha t}}{1 + \sigma_1 e^{-\alpha t} W(t)} \approx \sigma_1 e^{-\alpha t} [1 - \sigma_1 e^{-\alpha t} W(t)] \approx \sigma_1 e^{-\alpha t} \quad (V.7)$$

Entonces para la segunda clase del modelo de cuatro parámetros se tiene:

$$\partial t \Sigma_{2St}(0,t) = -\alpha (\Sigma_{2St}(0,t) + \sigma_0), \quad i.e. \quad \sigma(0,t) = \sigma_1 e^{-\alpha t} + \sigma_0 \quad (V.8)$$

y  $\sigma_2(t) \approx \frac{\sigma_0}{1 + \sigma_0 W(t)}$ , cuando  $\alpha T \gg 1$

es decir, la volatilidad de los precios spot decae con el tiempo, lo que es posible para el proceso de reversión a la media. Ahora, en lugar de

$$f_1(u) = \alpha W(u) \sigma^2(0,t) [1 - 2\sigma(0,t)W(u)]$$

se tiene

$$f_2(u) = \alpha\sigma_0 + \alpha\sigma(0,t) W(u)[\sigma(0,t) - 2\sigma_0 - 2\sigma(0,t)(\sigma(0,t) - \frac{3}{2}\sigma_0)W(u)]$$

y

$$\Phi_2(t) = \alpha\mu - \alpha \frac{\sigma(0,t)}{2} \int_0^t \{(\sigma - 2\sigma_0) - 2\sigma(2\sigma - 3\sigma_0)W(u) + 3\sigma^2(3\sigma - 4\sigma_0)W^2(u)\} du$$

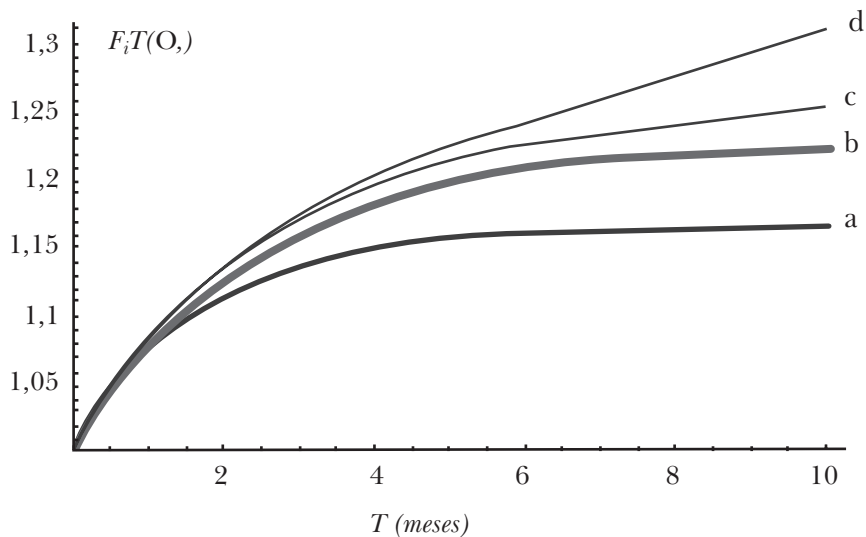
En seguida, con una exactitud de menor orden con respecto a  $W(t)$  se asume que

$$\begin{aligned} \langle Ln[F_{2St}(0,T)] \rangle &= e^{-\alpha T} Ln S_2(0) + \mu(1 - e^{-\alpha T}) + \frac{1}{2\alpha} \{(\sigma_1^2 - \sigma_0^2)\} e^{-\alpha T} \\ &- [\sigma_1^2 (1 + \alpha T)e^{-2\alpha T} - \sigma_0^2 (1 - \alpha T)] \end{aligned} \tag{V.9}$$

y

$$Var[LnF_{2St}(0,T)] = \frac{\sigma_0^2}{2\alpha} \{e^{-2\alpha T}(4e^{-\alpha T} - 1) + (2\alpha T - 3)\} \tag{V.10}$$

Los términos estructurales de estos modelos son:  $F_{1R}$ ,  $F_{1St}$ ,  $F_{2R}$  y  $F_{2St}$ , los valores específicos de los parámetros  $\alpha$ ,  $\mu$ ,  $\sigma_0$ ,  $\sigma_1$  y  $S(0)$  se dan en las figuras V.1 a la V.4.



**Figura V.1.** Familia de funciones  $F_i(0,T)$  con vencimiento  $\tau$ . La más baja (a) representada por  $F_{1R}$ , la siguiente (b) por  $F_{1St}$ , la c por  $F_{2R}$ , y la superior (d) por  $F_{2St}$ , para  $S(0) = 1$ .

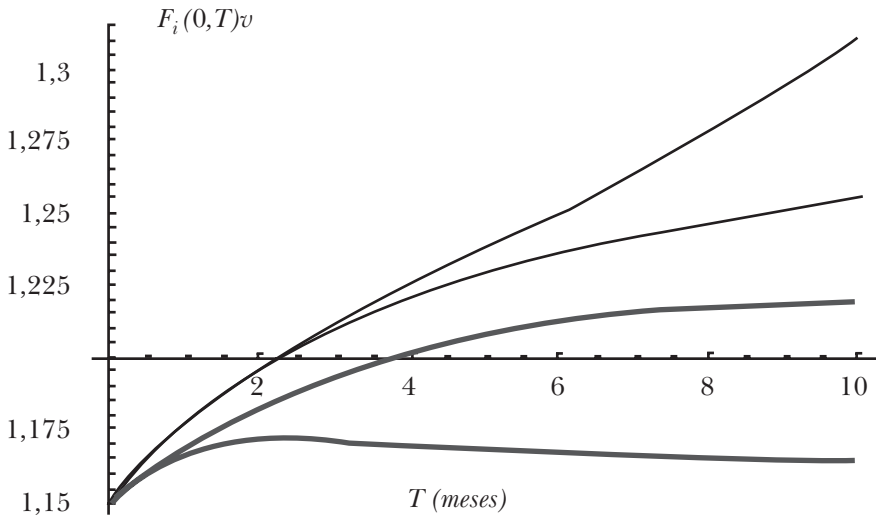


Figura V.2. Familia de funciones  $F_i(0, T)$  con vencimiento  $\tau$ , para  $S(0) = 1.15$ .

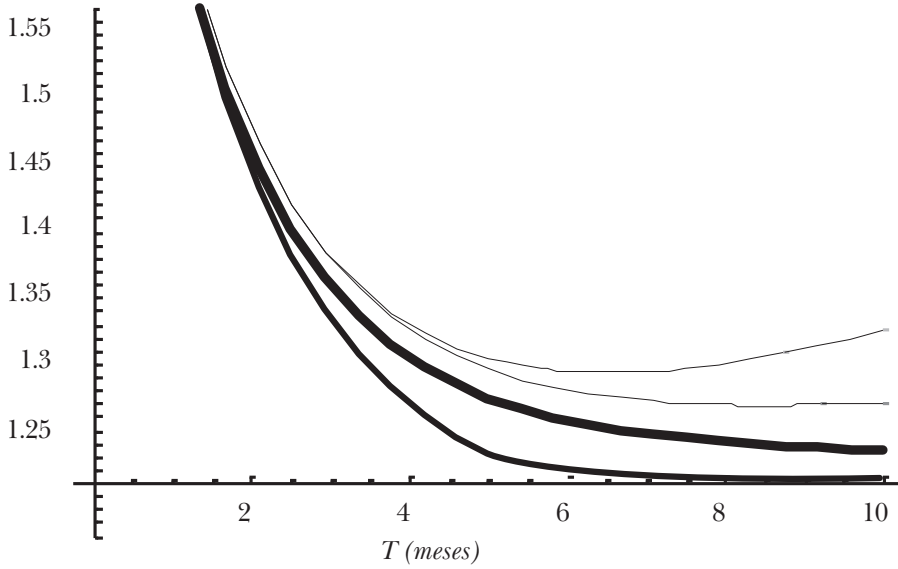
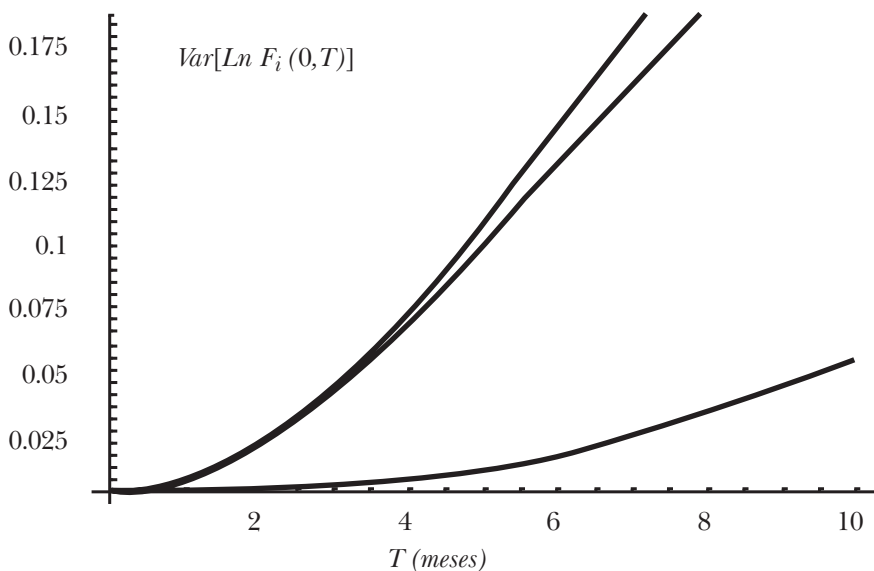


Figura V.3. Familia de funciones  $F_i(0, T)$  con vencimiento  $\tau$ , para  $S(0) = 2$ .

En todos los cálculos se consideró  $\alpha=0.5$ ,  $\mu=0.2$ ,  $\sigma_1=0.14$ ,  $\sigma_0=0.17$  y , lo más próximo a los valores del mercado del crudo Nymex. Se observa en la Figura V.3, que para el plazo corto (alrededor de dos meses), y fuera de las inmediaciones de  $S(0) = 1.15$ , las cuatro curvas se ven muy cerca una de otra, posteriormente se dispersan aproximadamente en 7-10% al final de un año. Alrededor del punto  $S(0) = 1.15$  en específico, la función  $F_{1R}(T)$  tiene una concavidad más pronunciada, esencialmente diferente de las otras tres curvas que son más lisas.

Para el primer modelo exponencial, las varianzas de las variables ordinarias y estocásticas no son significativas. Para las familias del segundo modelo, la varianza estocástica (curva superior), el “regular” (curva del medio) y la diferencia entre ellos (la curva de abajo) son las siguientes:



**Figura V.4.** Varianzas de las variables ordinarias y estocásticas del segundo modelo y su diferencia.

## CONCLUSIONES

La importante influencia que pueden causar las condiciones de frontera en distribuciones estacionarias estocásticas fueron formuladas en los documentos de (Makhankov *et al.*, 1995) y acompañadas de estudios numéricos (Makhankov, 1997). Lo obtenido en este trabajo se puede resumir con las siguientes observaciones: A). Las condiciones de frontera e iniciales deben especificar los datos de mercado sobre la base de los métodos conocidos, tales como series de tiempo, ARCH y GARCH y así sucesivamente. B) Los experimentos numéricos que están en función de las condiciones de frontera, muestran una posible inestabilidad de la solución conducente a la imprevisibilidad del comportamiento del mercado. Se han encontrado analíticamente las distribuciones estacionarias para la volatilidad del precio forward y las curvas calculadas analíticamente y comparadas con los datos de mercado muestran que para el corto plazo; es decir, de aproximadamente dos meses y fuera de las inmediaciones de  $S(0) = 1.15$ , los resultados son casi idénticos porque las cuatro curvas se ven relativamente cercanas una de otra. Este cuadro ya no es similar y no se repite a medida que pasa el tiempo, las curvas se dispersan y al final del año se encuentran dispersas entre 7-10%; por consiguiente, no se ven muy cerca. Uno de los modelos que podría ser empleado es el bi-paramétrico exponencial puro y, por ende, podría ser utilizado para el caso incluso de tiempos de largo plazo. Se ha estudiado la cuestión de cuál es la dimensión del proceso de Wiener que genera el comportamiento estocástico en el mercado, y se afirma que un proceso de Wiener de un factor sólo puede coincidir con la volatilidad de las tasas de interés a corto plazo. Después de un supuesto típico de que las volatilidades de las tasas de interés son constantes en la dinámica de las tasas forward, está completamente descrito por dos procesos independientes de Wiener; sin embargo, en el caso general de largo plazo requiere de la dinámica de más de dos procesos de Wiener, ya que, de acuerdo con la teoría anterior, la volatilidad cambia en el tiempo y la solución que se da es una estimación simple para el cambio en el tiempo. Así pues, en general, la dinámica a largo plazo es descrita por un proceso de Wiener multidimensional. Evidentemente, esto debe contar con al menos tres dimensiones; sin embargo, debido a la degeneración definitiva de los parámetros de mercado, el número mínimo de procesos estocásticos que operan en condiciones de mercado diferentes deben ser extraídos de los datos de éste. Esta imagen debe ser real debido al análisis de los componentes principales (véanse figuras IV.3. y IV.4) (Wilmott, 2001; Clewlow y Strickland, 1999).

## REFERENCIAS

- Clelow, C. y Strickland Ch. (1999). "Power Pricing – Making it Perfect". *Internet, Power: Continuing the Electricity Forward Curve Debate*.
- Cortazar, G. y Schwartz E. (1994). "The Valuation of Commodity Contingent Claims". *The Journal of Derivatives* 1(4): 27-39.
- Dubrovin, B., Fomenko, A. y Novikov, S. (1984). "Modern Geometry. Methods and Applications". Parte I, *The Geometry of Surfaces, Transformation Groups, and Fields*. Springer, Heidelberg.
- Hillard, J. y Reis, J. (1998). "Valuation of Commodity Futures and Options under Stochastic Convenience Yields, Interest Rates, and Jump Diffusions in the Spot". *Journal of Financial and Quantitative Analysis*, 33 (1): 61-86.
- Hull, J.C. (1993). *Options, Futures, and Other Derivative Securities*. Prentice Hall, Nueva Jersey.
- Kendal, W. (1987). "Stochastic Differential Geometry: an Introduction". *Acta Applicandae Mathematica*, 9: 29-60.
- Makhankov, V. G. (1997). "Stochastic Differential Geometry in Finance Studies". En E. Infield, R. Zelazny y A. Garl Kowski (eds.), *Nonlinear Dynamics, Chaotic and Complex Systems*. Cambridge University Press, Cambridge.
- Makhankov, V., Taranenko, Yu, Gomez, C. y Jones, R. (1995). *Geometrical Setting of the Term Structure of Interest Rate*. Los Álamos National Laboratory, Los Álamos.
- Makhankov, V. y Pashaev, O. (1992). *Integrable Pseudospin Models in Condensed Matter*. Harwood Academy Publishers, Londres.
- Schwartz, E. (1997). "The Stochastic Behavior of Commodity Prices: Implications for Pricing and Hedging". *The Journal of Finance*, LII (3): 923-973.
- Stratonovich, R. (1968). *Conditional Markov Processes and their Application to the Theory of Optimal Control*. Elsevier, Nueva York.
- Vasicek, O. (1977). "An Equilibrium Characterization of the Term Structure". *Journal of Financial Economics*, 5: 177-188.
- Wilmott, P. (2009). *The Mathematics of Financial Derivatives*. Cambridge University Press, Cambridge.
- (2001). *Quantitative Finance*. John Wiley & Sons.

# MÉXICO, TIPO DE CAMBIO Y CRECIMIENTO: UN ANÁLISIS CONTRAFACTUAL, 1950-2010\*

*Eduardo Loría\*\**

*Jorge Ramírez\*\*\**

## **Resumen**

De acuerdo con el enfoque de Thirlwall y Hussein (1982) sobre la restricción externa al crecimiento (REC), el papel del tipo de cambio real se vuelve fundamental en la explicación del lento crecimiento económico en los países que tienden a apreciar sistemáticamente su tipo de cambio, como México. Bajo este argumento se estiman dos modelos de vectores autorregresivos estructurales (SVAR): uno con los datos observados y otro alterno con una serie de tipo de cambio real de equilibrio generado aleatoriamente. Los resultados nos indican que un tipo de cambio competitivo, alto y estable disuelve el histórico círculo vicioso crecimiento-apreciación cambiaria-crisis de balanza de pagos-depreciación-recesión que ha afectado a la economía mexicana desde los años setenta.

**Palabras clave:** Restricción externa al crecimiento, tipo de cambio real, crecimiento económico.

\* Este artículo ha contado con el financiamiento del proyecto: “Recesión y mercados laborales en México. Un análisis estructural, 1985-2020”. PAPIIT IN-305711, DGAPA, UNAM. Los autores son los únicos responsables de lo que aquí se dice o se omite.

\*\* Doctor en Economía, Universidad Nacional Autónoma de México, UNAM. Profesor de Tiempo Completo (Teoría económica y Macroeconometría) y Coordinador del Centro de Modelística y Pronósticos Económicos, Facultad de Economía, UNAM. Facultad de Economía, Edificio B, 3er. Piso, Circuito Interior, Ciudad Universitaria, México, D.F., C.P. 04510 (52) (55) 5622 2142, <www.eduardoloria.name, <eduardol@servidor.unam.mx>.

\*\*\* Maestro en Economía, Universidad Autónoma del Estado de México, UAEM. Profesor de la Facultad de Economía de la Universidad Autónoma del Estado de México en las áreas de Teoría económica, Matemáticas y Econometría, <jramirez@uaemex.mx>.

## INTRODUCCIÓN

Gran parte del debate sobre las causas del lento crecimiento que afecta a México en los últimos 30 años culpa a la estrategia de crecimiento seguida por el gobierno. Los sectores de izquierda acusan a los gobiernos de haber seguido una estrategia neoliberal, mientras que los grupos de derecha también lo cuestionan por no haber profundizado las reformas estructurales iniciadas en los años ochenta. En ambos casos parece omitirse una variable que desde nuestro punto de vista es fundamental en la explicación y que tiene que ver con el comportamiento del tipo de cambio real.

Tradicionalmente en nuestro país la estabilidad cambiaria frente al dólar se ha considerado como un sinónimo de fortaleza nacional en todos los planos, no sólo en el económico. Por esto, las autoridades han tratado de mantener a toda costa un valor nominal estable con las menores variaciones (depreciaciones) posibles respecto del dólar estadounidense restando importancia a su valor real (apreciación) y a las negativas consecuencias sobre la oferta, la demanda, la balanza de pagos y, finalmente, sobre el crecimiento económico.

El esfuerzo de las autoridades económicas por mantener firme la paridad nominal ha generado episodios de apreciaciones cambiarias reales que generalmente han provocado crisis de balanza de pagos que a la postre han llevado a devaluaciones importantes, las cuales se han traducido en fuertes caídas del nivel de actividad, del empleo, del salario real y a regresiones distributivas. Las razones de estos efectos contractivos se encuentran en los altos coeficientes de insumos importados a producto y a la alta elasticidad ingreso de las importaciones (Loría, 2001), por lo que las devaluaciones generan fuertes efectos ingreso y precio sobre productores y consumidores y muy bajos efectos sustitución. Las consecuencias de las devaluaciones remediales son las que los gobiernos han querido evitar, y por ello le han conferido a la estabilidad cambiaria el papel de ancla antiinflacionaria, propósito que no ha cambiado aún en el enfoque de objetivos de inflación que el banco central adoptó oficialmente en la segunda mitad de los años noventa.

Sin embargo, la evidencia empírica demuestra contundentemente que esta política ha generado procesos que en el largo plazo no son sostenibles y que hacen que las fases de crecimiento económico se autoderroten.

Nuestro objetivo es explorar la importancia del tipo de cambio real en el crecimiento de la economía mexicana para el periodo 1950-2010 utilizando dos modelos de vectores autorregresivos estructurales (SVAR).

Los resultados empíricos demuestran –en línea con el enfoque postkeynesiano– que mantener un tipo de cambio real competitivo es un factor importante de crecimiento económico.

El trabajo se divide en tres apartados: en el primero se recoge la evidencia empírica del tipo de cambio en la economía nacional. Continuamos con un análisis teórico sobre la restricción externa al crecimiento (REC) propuesta por Thirlwall, y se revisa la metodología de los SVAR. Posteriormente se presentan y se discuten los principales resultados del modelo. Finalmente se puntualizan las principales conclusiones y se hacen algunas recomendaciones de política económica.

## I. EL PAPEL DEL TIPO DE CAMBIO EN LA ECONOMÍA MEXICANA

La evidencia empírica demuestra un fuerte efecto del tipo de cambio nominal en la formación de precios, por lo que los fenómenos inflacionarios y desinflacionarios tradicionalmente se han vinculado de manera estrecha a las fluctuaciones del tipo de cambio nominal y real (Castro *et al.*, 1997 y Loría, 2003).

Por sus características estructurales la economía mexicana presenta un problema central, el cual radica en que tratar de preservar el equilibrio interno de esta manera (apreciando el tipo de cambio) generalmente afecta el equilibrio externo, que tarde o temprano afectará al primero y así sucesivamente hasta conformar un círculo vicioso que afecta negativamente la dinámica del desarrollo económico. Por tanto, este *trade off* entre inflación y equilibrio del tipo de cambio ha marcado el gran problema del crecimiento de la economía mexicana, por lo menos desde hace poco más de 50 años.

Bajo los argumentos anteriores resulta claro que un gran esfuerzo de política económica consiste en encontrar cómo romper ese círculo vicioso.

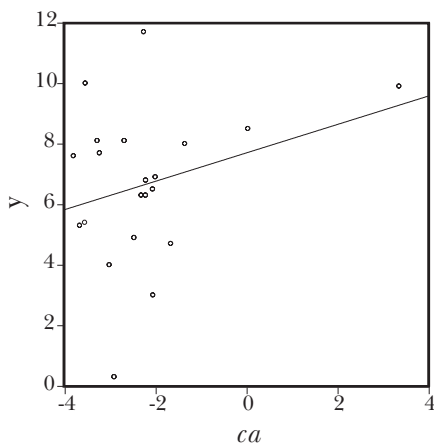
Es plausible entonces considerar que en la apreciación endémica del tipo de cambio real reside la naturaleza del ciclo económico de la economía mexicana. Asimismo, por el solo hecho de que la economía mexicana tiene una pesada carga financiera derivada de su deuda externa histórica, existe –de principio– un piso básico de déficit de cuenta corriente que de 1950 a 2010 osciló como porcentaje de las exportaciones en 28.7% y como porcentaje del PIB en 3.8%.

Si bien un déficit de cuenta corriente se asocia positivamente con el crecimiento de corto plazo, a lo largo del tiempo se vuelve en contra cuando se enfrentan las obligaciones financieras por el pago del servicio de la deuda externa. Los casos de países de la eurozona y de Estados Unidos son un excelente ejemplo de ello.

Hay que advertir que este *trade off* entre crecimiento y déficit de cuenta corriente ha cambiado notablemente en el tiempo. Para el periodo 1950-1970, en que prevaleció un régimen de protección comercial muy elevado y un bajo endeu-

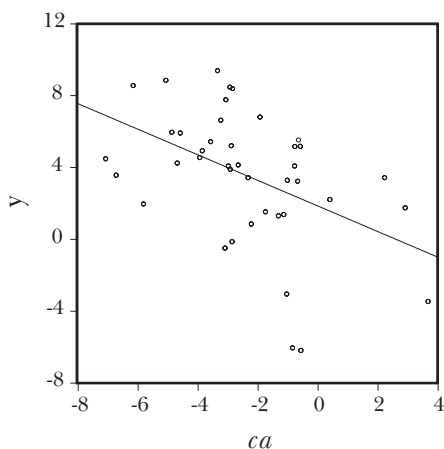
damiento externo en el marco institucional de *Bretton Woods*, prácticamente no existió. La Gráfica I.1a muestra la existencia de una ligera relación positiva entre ambas variables. Básicamente podemos referir que las altas tasas de crecimiento del producto ( $y$ ) no necesariamente se asociaban a altos déficits de cuenta corriente como porcentaje del producto ( $ca$ ). Sin embargo, en la medida que comenzó a fracturarse el sistema *Bretton Woods* y la economía mexicana comenzó a sufrir choques externos, acompañados por apreciaciones cambiarias y altos niveles de deuda desde principios de los años setenta, esta relación se invirtió dominando para todo el periodo 1950-2010 (Gráfica I.1b).

**Gráfica I.1a**  
**México: crecimiento anual del PIB ( $y$ ) y déficit de cuenta corriente como % del PIB ( $ca$ ). Ajuste lineal, 1950-1970**



Al evaluar conjuntamente la evolución de  $y$  y de  $ca$  (véase Cuadro I.1) resalta que desde 1950 y hasta 1975 ambas variables mantuvieron desempeños esencialmente estables; esto es, altas tasas de crecimiento (6.61%) se acompañaban de moderados déficits externos (2.51% en promedio respecto al PIB). Luego de las grandes devaluaciones de mediados de los setenta fue cuando ambas variables iniciaron caminos muy erráticos. Respecto a la efímera fase de alto crecimiento (1978-1981), destaca el gran déficit de  $ca$  que –al combinarse con choques externos adversos– llevó más tarde a una larga fase de maxi devaluaciones y de estancamiento con inflación (estancflación). La cuasifijación cambiaria (1988-1994) fue muy exitosa para bajar la inflación, pero también para apreciar el tipo de cambio

**Gráfica I.1b**  
**México: crecimiento anual del PIB ( $y$ ) y déficit de cuenta corriente como % del PIB ( $ca$ ). Ajuste lineal, 1970-2010**



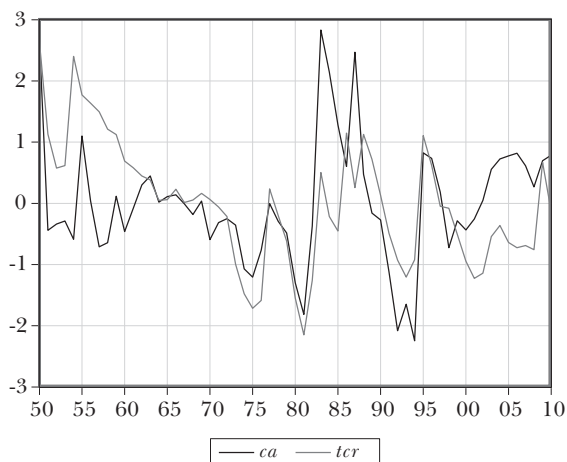
real que, al combinarse con el reinicio del crecimiento, llevó nuevamente a otra serie de devaluaciones entre diciembre de 1994 y fines de 1995.

El argumento central del trabajo establece la importancia fundamental del tipo de cambio real sobre la cuenta corriente y, por tanto, sobre el crecimiento económico –a reserva de probarse econométricamente–, intuitivamente esta relación se puede observar con claridad en la Gráfica I.2.

**Cuadro I.1**  
**México: crecimiento y cuenta corriente a PIB. Estadísticas básicas por periodos, 1950-2010**

Periodo	$y$		$ca$	
	Media	Desviación estándar	Media	Desviación estándar
1950-1970	6.66	2.55	-2.24	1.56
1970-2010	3.53	3.68	-2.36	2.35
1950-1975	6.61	2.38	-2.51	1.55
1978-1981	8.79	0.41	-4.36	1.50
1981-1988	1.23	3.83	-0.31	3.31
1989-1994	3.91	1.05	-4.94	1.91
1995-2010	2.48	3.92	-1.60	1.09

**Gráfica I.2**  
**México: Cuenta corriente y tipo de cambio real (tcr),**  
**1950-2010 (series normalizadas)**



\*  $tcr = E \cdot (PUS/PMEX)$ , donde  $E$  = Tipo de cambio nominal (pesos por dólar de EU);  $PUS$  = INPC de Estados Unidos;  $PMEX$  = INPC de México.

En resumen, las relaciones anteriormente planteadas indican que el déficit de cuenta corriente de la economía mexicana tiene tres componentes principales que son sistemáticos: *a)* el diferencial de precios con Estados Unidos, *b)* la mayor elasticidad ingreso de las importaciones respecto a las exportaciones (Loría, 2001) y *c)* el endeudamiento externo histórico. Hay que advertir que este último factor –por ser de carácter financiero– tiene una dinámica autorregresiva muy importante.<sup>1</sup>

## II. LA RESTRICCIÓN EXTERNA AL CRECIMIENTO

La restricción externa al crecimiento (REC) planteada originalmente por Thirlwall en los años setenta plantea que cualquier economía, independiente de su grado de desarrollo, no puede crecer a una tasa indefinida en el largo plazo, sino a la tasa que corresponde al equilibrio de la balanza comercial.

Esto es así debido a que en el largo plazo no puede tener un déficit creciente porque eventualmente caerá en un problema de balanza de pagos.

<sup>1</sup> Esto es, por demás, obvio y se desprende del principio del interés compuesto.

En su versión original el modelo de Thirlwall (1979) sólo considera a la balanza comercial, aunque en una versión posterior (Thirlwall y Hussein, 1982) se amplía el concepto de la REC al considerar a los movimientos de la balanza servicios factoriales y no factoriales incluyendo así a la cuenta corriente.

A continuación se presenta la formalización del modelo.

Consideremos el equilibrio comercial:

$$PX = P^*EM \tag{1}$$

Donde:  $P$  = INPC doméstico;  $X$  = Volumen de las exportaciones;  $P^*$  = INPC del resto del mundo, pero para el caso de México son los precios de Estados Unidos;  $E$  = tipo de cambio nominal y  $M$  = volumen de importaciones.

Aplicando logaritmos y sus variaciones, tenemos el equilibrio comercial en términos dinámicos:

$$p + x = p^* + c + m \tag{2}$$

Las variables en minúsculas denotan las tasas de crecimiento.

Es necesario especificar las funciones de exportaciones e importaciones:

$$X = (P / EP^*)^\eta Z^\varepsilon \tag{3}$$

$$M = (EP^*/P)^\psi Y^\pi \tag{4}$$

Interesa tener estas ecuaciones también en términos dinámicos, por lo que aplicamos logaritmos y diferenciamos respecto al tiempo:

$$x = \eta(p - p^* - e) + \varepsilon z \tag{5}$$

$$m = \psi(p^* + e - p) + \pi y \tag{6}$$

Donde:  $\eta$  = elasticidad precio de las exportaciones;  $\varepsilon$  = elasticidad ingreso de las exportaciones;  $z$  = producto real de Estados Unidos;  $\psi$  = elasticidad precio de las importaciones;  $\pi$  = elasticidad ingreso de las importaciones;  $y$  = producto real doméstico.

Por teoría:  $\eta < 0$ ;  $\varepsilon > 0$ ;  $\psi < 0$  y  $\pi > 0$ .

Resolviendo el sistema al sustituir (5) y (6) en (2) y haciendo operaciones obtenemos la ecuación principal de crecimiento económico restringido por el equilibrio de balanza comercial:

$$y_b = [1 + \eta + \psi] (p - p^* - e) + \varepsilon z / \pi \quad (7)$$

Que expresa que  $y_t$  puede crecer si:

- a) Se deprecia el tipo de cambio real.
- b) Se cumple la condición Marshall-Lerner:  $|\eta + \psi| > 1$ .
- c) Aumenta el crecimiento externo (Estados Unidos).
- d) Aumenta la elasticidad ingreso de las exportaciones.
- e) Disminuye la elasticidad ingreso de las importaciones.

Un supuesto importante de este modelo es que el tipo de cambio real se mantiene constante en el largo plazo, lo cual implica que:  $(p - p^* - e) = 0$ , con lo cual (7) se simplifica sensiblemente:

$$y_b = \varepsilon z / \pi \quad (8)$$

o bien

$$y_b = x / \pi \quad (8')$$

debido a que, al considerar fijo al tipo de cambio real, el crecimiento de las exportaciones se determina exclusivamente por  $\varepsilon z$ , que corresponde a la tasa de crecimiento de las exportaciones.

Por el objetivo del trabajo, que es ver la REC determinada por la cuenta corriente, reexpresamos (1) incorporando ahora los flujos netos de capital (F) que expresa el equilibrio de la cuenta corriente de la balanza de pagos:

$$PX + EF = P^*EM \quad (9)$$

Reexpresando al lado izquierdo en términos porcentuales (de las exportaciones netas):

$$\theta = \frac{PX}{PX - EY} \quad (10)$$

$$1 - \theta = \frac{EY}{PX - EY} \quad (11)$$

Donde  $\theta$  es la proporción de las exportaciones de bienes y servicios en las exportaciones netas y  $1 - \theta$  es la proporción de los flujos de capital respecto a las exportaciones netas.

Dinamizando (9), y siguiendo el mecanismo ya descrito, tenemos:

$$\theta(p + x) + (1 - \theta)(c + f) = p^* + c + m \quad (12)$$

Resolviendo nuevamente el sistema llegamos a la siguiente expresión fundamental:

$$y_{bf} = [(\theta\eta + \psi + 1)(p - p^* - c) + \theta\epsilon z + (1 - \theta)(f + c - p)]/\pi \quad (13)$$

De lo que resultan los factores que permiten acelerar (reducir) el crecimiento económico restringido por el equilibrio de la cuenta corriente: *a*) el tipo de cambio real; *b*) el crecimiento del producto exterior; *c*) la entrada neta de capitales (servicios factoriales y no factoriales), y *d*) la elasticidad ingreso de las importaciones.

De nuevo, si se supone que el tipo de cambio real es constante en el tiempo, se elimina el primer término del lado derecho y queda una expresión algebraica más sencilla:

$$Y_{bf} = [\theta\epsilon z + (1 - \theta)(f + e - p)]/\pi \quad (14)$$

Resulta claro que una política de apreciación cambiaria con objetivos desinflacionarios aumentará la REC y, por tanto, reducirá el crecimiento económico y viceversa.

Lo que hacemos en este trabajo no es proponer depreciaciones continuas, sino evitar apreciaciones reales y mantener un tipo de cambio real alto y estable.

### III. METODOLOGÍA ECONOMÉTRICA

#### III.1 El modelo SVAR

Una gran bondad de los modelos SVAR es que en el análisis de las series de tiempo permite evaluar causalidad, sensibilidad y respuestas dinámicas eliminando perturbaciones indeseadas al identificar el sistema con base en la estructura de los datos y en los argumentos provenientes de una teoría económica relevante.

Un VAR irrestricto (estándar) estima un modelo a partir únicamente de los rezagos de las variables endógenas y de las variables exógenas asignadas:

$$y_t = d_t + Cy_{t-i} + v_t \quad (15)$$

Donde  $y_t$  es un vector de variables endógenas,  $d_t$  es un vector de componentes determinísticos (constante, tendencia y variables *dummy* de intervención o estacionales) y  $v_t$  es el vector de innovaciones.

En (15) no se explican los efectos contemporáneos entre las variables, pero tales efectos están contenidos en la matriz de varianzas y covarianzas generada del vector  $v_t$ .

Un análisis exhaustivo de un VAR primitivo nos conduce a una mejor comprensión. Consideremos la siguiente expresión (Enders, 2004):

$$By_t = d_t + Ay_{t-1} + \varepsilon_t \quad (16)$$

El VAR en su forma reducida (16) es simplemente una reparametrización de la especificación más general dada por (15). De hecho, es fácil denotarlo dado que  $C = B^{-1}A$  y  $v_t = B^{-1}\varepsilon_t$ . Lo anterior implica que los errores del modelo final (16) son combinaciones lineales de los *shocks* no correlacionados  $\varepsilon_t$ .

Para recuperar las interacciones contemporáneas de interés contenidas en la matriz  $B$ , tradicionalmente se impone una estructura triangular estándar de Cholesky (Sims, 1980). Sin embargo, una identificación proveniente de la teoría económica y de la estructura de los datos exige imponer restricciones *ad hoc* para computar con el mayor sentido económico posible –además del estrictamente estadístico– las funciones de impulso-respuesta. Este procedimiento permite alcanzar la condición de identificación, que establece que el número de elementos

diferentes a cero en la matriz  $B$  debe ser igual o menor que  $\left(\frac{n^2 - n}{2}\right)$  (Enders, 2004: 272 y Hamilton, 1994: 334).

### III.2 Estimación

Con objeto de probar con rigor nuestra hipótesis principal, estimamos dos modelos de vectores autorregresivos estructurales con series anuales para el periodo 1950-2010.

El primero con las series históricas observadas y el segundo cambiando solamente la serie de tipo de cambio real a partir de generar una serie aleatoria tomando como punto de referencia el valor que adquirió en 1954. El objetivo es probar que un tipo de cambio real más alto y estable al que se observó en adelante, opera positivamente sobre el crecimiento económico.

### III.2.1 El modelo base

De acuerdo con los criterios tradicionales,<sup>2</sup> inicialmente estimamos un VAR(3)<sup>3</sup> irrestricto con constante y una variable *dummy* para 1995 que captura la caída más abrupta del crecimiento que ha tenido la economía mexicana desde la gran depresión de los años treinta. Como se muestra en el Cuadro III.2A (del Anexo) nuestra estimación cumple con todos los supuestos de correcta especificación.

El orden de las variables se eligió tomando como referencia la causalidad en el sentido de Granger (véase Cuadro III.4A), logrando obtener una especificación exactamente identificada cuya expresión  $B^{-1}\mathbf{v}_t = \boldsymbol{\varepsilon}_t$  es:

$$\begin{bmatrix} 1 & 0 & 0 \\ b_{21} & 1 & 0 \\ b_{31} & b_{32} & 1 \end{bmatrix} \begin{bmatrix} tcr \\ ca \\ y \end{bmatrix} = \begin{bmatrix} \varepsilon_{tcr} \\ \varepsilon_{ca} \\ \varepsilon_y \end{bmatrix}$$

Al resolver el modelo se obtuvieron dos relaciones fundamentales con los signos correctos, los cuales reflejan solamente el sentido de las relaciones no elasticidades.<sup>4</sup>

$$tcr = \varepsilon_{tcr} \tag{17}$$

$$ca = 0.41 * \varepsilon_{tcr} + \varepsilon_{ca} \tag{18}$$

$$y = -0.48 * \varepsilon_{tcr} - 0.44 + \varepsilon_{ca} + \varepsilon_y \tag{19}$$

Del sistema anterior es posible observar que la variable más exógena es el tipo de cambio real (*tcr*) la cual sólo es explicada por sus innovaciones, seguido por la balanza de cuenta corriente como proporción del PIB (*ca*), que es explicada únicamente por el tipo de cambio real. Esta relación refleja el efecto correctivo que tiene una depreciación sobre la cuenta corriente.

<sup>2</sup> Schwarz, Akaike, Hannan-Quinn, Error Final de Predicción y estadístico LR.

<sup>3</sup> Es importante destacar que todas las series son estacionarias en sus valores originales (véase Anexo, Cuadro III.1A).

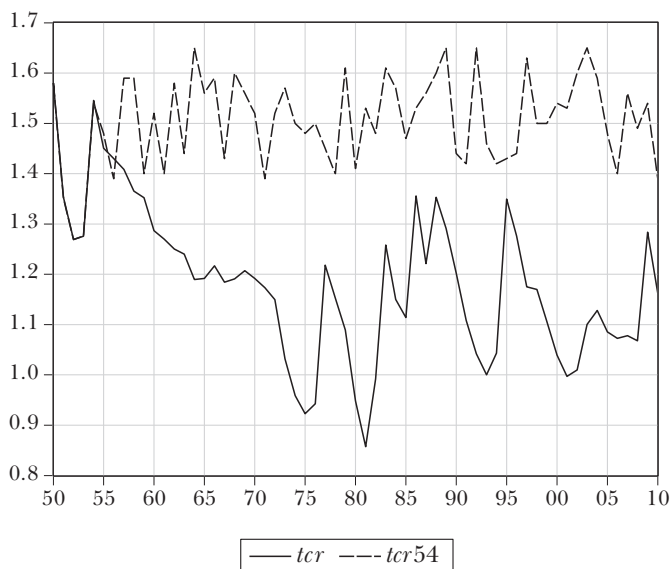
<sup>4</sup> Como se sabe, lo que arrojan las siguientes ecuaciones son las innovaciones estructurales, por lo que no pueden leerse directamente los parámetros, sino solamente los signos. El análisis de impulso respuesta aporta elementos algebraicos en cuanto a la magnitud y duración de las relaciones entre variables.

La última ecuación presenta el efecto negativo que tiene una depreciación remedial o correctiva sobre el crecimiento del producto ( $y$ ), así como el efecto negativo del aumento del desequilibrio externo sobre el crecimiento económico de largo plazo.

### III.2.2 El modelo alternativo

Con el objetivo de probar con rigor nuestra hipótesis principal estimamos un modelo SVAR alternativo suponiendo que el tipo de cambio real se mantuvo en niveles altos y estables, tomando como referencia la observación de 1954, debido a que desde entonces se comienzan a gestar las grandes y constantes sobrevaluaciones, además de ser la única devaluación que ha registrado efectos expansivos inmediatos. Con esta observación generamos una serie aleatoria ( $tcr54$ ) teniendo como límite de fluctuaciones valores de  $\pm$  una desviación estándar (véase Gráfica III.1).<sup>5</sup>

**Gráfica III.1**  
**Contraste del tipo de cambio real**



<sup>5</sup> Al igual que las demás, esta serie es estacionaria en niveles (véase Anexo Cuadro III.1A).

Bajo el mismo procedimiento estimamos un VAR(3)<sup>6</sup> que siguió el mismo ordenamiento de las variables, resultando el siguiente sistema de ecuaciones:

$$tcr54 = \varepsilon_{tcr51} \quad (20)$$

$$ca = 0.04 \cdot \varepsilon_{tcr54} \varepsilon_{ca} \quad (21)$$

$$y = 0.47 \cdot \varepsilon_{tcr54} - 0.61 \cdot \varepsilon_{ca} + \varepsilon_y \quad (22)$$

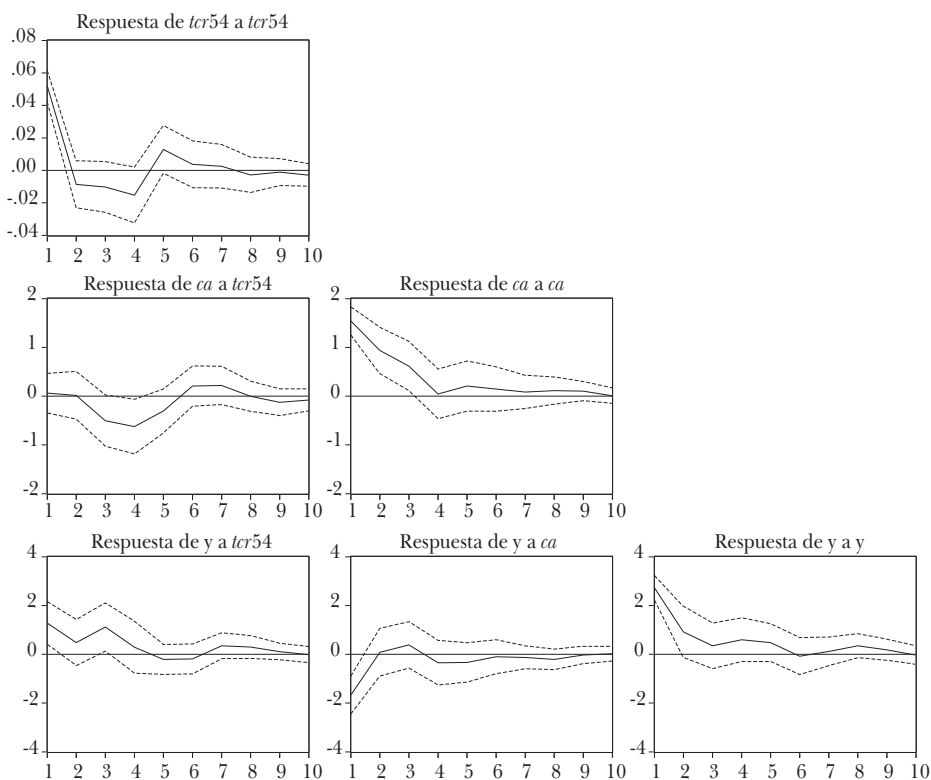
Es de destacar que los signos se conservan, con excepción del que corresponde al tipo de cambio respecto al crecimiento, el cual ahora es positivo. Lo anterior prueba nuestra hipótesis central. Al evaluar el análisis de impulso respuesta se observa claramente el efecto positivo de un choque en las innovaciones de la paridad cambiaria sobre el crecimiento del PIB. Respecto del análisis de descomposición de varianza, es posible destacar el carácter exógeno del tipo de cambio y su carácter explicativo de la cuenta corriente.

Con los resultados anteriores podemos probar nuestra hipótesis central en el sentido de que el tipo de cambio real es fundamental para explicar el crecimiento económico, además de verificar el hecho de que mantener un tipo de cambio competitivo en niveles altos y estables impacta positivamente al crecimiento económico. A un resultado similar llega Ibarra (2011: 212) quien afirma: “Un tipo de cambio competitivo no sólo estimula un crecimiento más acelerado de las exportaciones de manufacturas no vinculado a la maquila, sino que también amplifica el efecto positivo de las exportaciones sobre el PIB”.

<sup>6</sup> Con una variable *dummy* para 1955 y 1956 que refleja el cambio estructural derivado de la nueva serie de tipo de cambio. Las pruebas de correcta especificación pueden verse en el Cuadro III.2A.

### Gráfica III.2 Análisis de impulso respuesta<sup>7</sup>

Response to Structural One S.D. Innovations  $\pm$  S.E.



<sup>7</sup> En principio se debe considerar que todos los efectos son significativos al 95% de confianza, de acuerdo con el rango que determinan las bandas de confianza para  $\pm 2$  errores estándar (Valdés, 1997; Gulli, 2005; Calderón y Méndez, 2009).

**Cuadro III.1**  
**Descomposición de varianza**

<i>Descomposición de varianza de tcr54:</i>				
<i>Periodo</i>	<i>S.E.</i>	<i>tcr54</i>	<i>ca</i>	<i>y</i>
1	0.05	100.00	0.00	0.00
5	0.06	80.38	8.22	11.41
10	0.07	77.74	9.40	12.86
15	0.07	77.72	9.39	12.89
20	0.07	77.71	9.40	12.89
<i>Descomposición de varianza de ca:</i>				
<i>Periodo</i>	<i>S.E.</i>	<i>tcr54</i>	<i>ca</i>	<i>y</i>
1	1.54	0.16	99.84	0.00
5	2.19	15.42	76.36	8.22
10	2.24	17.01	74.15	8.83
15	2.24	17.09	74.03	8.88
20	2.24	17.10	74.02	8.88
<i>Descomposición de varianza de y:</i>				
<i>Periodo</i>	<i>S.E.</i>	<i>tcr54</i>	<i>ca</i>	<i>y</i>
1	3.45	13.84	23.53	62.64
5	3.93	21.09	20.64	58.27
10	3.99	22.04	20.44	57.52
15	3.99	22.04	20.49	57.47
20	3.99	22.05	20.49	57.47

Factorización: Estructural.

## CONCLUSIONES Y RECOMENDACIONES DE POLÍTICA

En el documento se prueba econométricamente nuestra hipótesis central que afirma que el crecimiento económico de largo plazo de la economía mexicana depende centralmente del tipo de cambio real, debido a que actúa de manera determinante sobre el equilibrio de la cuenta corriente y con ello reduce la restricción externa al crecimiento.

En la teoría postkeynesiana de economía abierta se acepta que el crecimiento económico de largo plazo de cualquier país –independientemente de su grado de desarrollo económico (Atesoglu, 1993, 1993-1994, 1995, 1997)– depende centralmente de sus elasticidades del comercio exterior y tiende a despreciar la importancia del tipo de cambio en la restricción externa al crecimiento, porque considera que es esencialmente fijo en el largo plazo.

Aquí no proponemos devaluaciones continuas, sino evitar apreciaciones sistemáticas que deterioran la cuenta corriente y, por tanto, reducen la capacidad de crecimiento.

Se demostró que la elección de las autoridades económicas entre tener baja inflación y un tipo de cambio de equilibrio, tradicionalmente se han inclinado por obtener el primer objetivo, con lo cual se han generado ciclos caracterizados por crecimiento económico acompañados de apreciación cambiaria que, invariablemente, suscitan desequilibrios en el sector externo que desembocan en violentas devaluaciones remediales que originan estados de estanflación particularmente desde 1976 y hasta 1994 en que se mantuvo fijo (con variantes) el tipo de cambio nominal. A pesar de que se obtengan eventualmente las metas de inflación, la apreciación cambiaria combinada con los orígenes estructurales del déficit de la cuenta corriente han provocado crisis recurrentes de balanza de pagos. Esta política es la que explica –en última instancia– los procesos de alto y siga de los últimos 50 años.

Con el propósito de probar empírica y rigurosamente estas argumentaciones se especificaron dos modelos SVAR: el primero estima el sistema económico tal y como lo refieren los datos observados, mientras que el segundo supone un nivel de tipo de cambio real alto y estable. Ambos modelos de series de tiempo prueban que el tipo de cambio real es fundamental en la explicación del déficit de la cuenta corriente y, por tanto, del crecimiento económico.

La prueba de causalidad en el sentido de Granger corrobora la exogeneidad del tipo de cambio real, por lo que puede utilizarse como un instrumento y predictor del crecimiento económico o, visto inversamente, de su restricción.

El análisis de impulso-respuesta y de descomposición de varianza del modelo alternativo prueba satisfactoriamente la hipótesis de que el crecimiento económi-

co de México depende de la restricción externa y que ella, a su vez, está directamente asociada al tipo de cambio real.

Por tanto, una conclusión central es la aceptación de que esta variable es mucho más importante de lo que muchos creen. No es un precio más dentro del sistema económico. *Es el precio de todos los precios*, ya que actúa simultáneamente sobre el equilibrio interno y el externo, más aun en las condiciones actuales en que los tradicionales instrumentos de política económica se han debilitado sensiblemente.

El tipo de cambio es la variable que puede contrarrestar los ciclos económicos, así como los costos sociales de las reformas estructurales. Por tanto, dentro del régimen de flotación cambiaria el objeto central de las autoridades económicas debería enfocarse a mantener un tipo de cambio real competitivo muy cercano a su valor de equilibrio de largo plazo.

Podría argumentarse *a contrario sensu* que seguir esta política generaría una espiral inflacionaria. Sin embargo, creemos que un tipo de cambio real competitivo, como aquí lo hemos probado, tendría un fuerte efecto estimulante sobre la producción y la demanda internas, con lo cual se contrarrestarían los efectos desindustrializadores que se generan de las apreciaciones cambiarias. De esta suerte, a partir de estimular el crecimiento económico es plausible considerar que se podrían generar cadenas productivas que reducirían los altos coeficientes de importaciones de productos y la elasticidad ingreso de las importaciones, con lo cual las presiones inflacionarias debieran reducirse. En suma, lo que se debe buscar a toda costa es una política monetaria favorable al crecimiento que evite apreciaciones y las consecuentes depreciaciones cambiarias remediales que, por ser de ese carácter, siempre han tenido fuertes costos sociales y económicos tal como lo han demostrado Krugman y Taylor (1978).

El objetivo de mantener un tipo de cambio competitivo, por el contrario, evitaría estas fuertes depreciaciones correctivas con lo cual podrían también evitarse los choques inflacionarios resultantes.

El caso más reciente que refiere este esquema (Damil y Frenkel, 2009) es el de Argentina, la cual luego de su crisis política de 2001 y 2002 puso en marcha una política económica sustentada en:

- a) Un tipo de cambio alto (competitivo) y estable, el cual evita al máximo las apreciaciones.
- b) Superávit fiscal y de cuenta corriente, lo cual, aunado al refinanciamiento de su deuda, provocó que en promedio esta economía creciera en 7.5% durante 2003-2010, y aún durante la crisis de 2009 mostró un crecimiento moderado.

Una forma de avanzar en ese sentido es que, además del objetivo de control inflacionario –que por indicación constitucional le corresponde al Banco de México–, ahora podría asignársele el de impulsar el crecimiento económico. Stiglitz (2002) indica que una de las lecciones que ha dejado la crisis argentina es que enfocar la política económica al objetivo de inflación –sin tomar en cuenta explícitamente el objetivo del crecimiento– es muy arriesgado.

Ello podría conseguirse si se persigue un objetivo de tipo de cambio con base en reglas activistas. De esta manera, el Banco de México seguiría así reglas de operación de tasas de interés de referencia que actuarían simultáneamente sobre agregados monetarios (crédito) y sobre el tipo de cambio nominal y, por tanto, real.

De hecho, en un régimen de tipo de cambio flexible, es la política monetaria la que debe comandar el crecimiento económico a través de sus efectos sobre las tasas de interés y, por tanto, sobre el tipo de cambio.

En tal sentido, debe dársele la mayor prioridad a la recuperación del crédito interno a través de políticas monetarias menos conservadoras y dejar entonces que el tipo de cambio real juegue el papel que le corresponde en un régimen de tipo de cambio flexible: ser un factor de expansión de la demanda interna y, al mismo tiempo, de protección del equilibrio externo.

## REFERENCIAS

- Atesoglu, H. S. (1993). “Balance of Payments Constrained Growth: Evidence from the United States”. *Journal of Post Keynesian Economics*, núm. 15, Summer.
- (1993-1994). “Exports, Capital Flows, Relative Prices, and Economic Growth in Canada”. *Journal of Post-Keynesian Economics*, vol. 16, núm. 2, Winter.
- (1995). “An Explanation of the Slowdown in US Economic Growth”. *Applied Economic Letters*, 2.
- (1997). “Balance of Payments Constrained Growth Model and its Implications for the United States”. *Journal of Post-Keynesian Economics*, vol. 19, núm. 3, Spring.
- Calderón, G. A. y J. J. Méndez (2009). “Relación del ciclo de crecimiento de los Estados Unidos de América con el ciclo económico de Guatemala”. *Notas Monetarias Banco de Guatemala*. Recuperado de <[www.banguat.gob.gt/inveco/notas/articulos/envolver.asp?karchivo=3201&kdisc=si](http://www.banguat.gob.gt/inveco/notas/articulos/envolver.asp?karchivo=3201&kdisc=si)>.
- Castro, C.; E. Loría y M. A. Mendoza (1997). *Eudoxio: modelo macroeconómico de la economía mexicana*. 1ª reimpresión (2000). México: UNAM.
- Damill, M. y R. Frenkel (2009). *Las políticas macroeconómicas en la evolución reciente en la economía argentina*. Banco Central de la República Argentina. Recuperado

- de <[http://www.bcra.gov.ar/pdfs/investigaciones/PaperFrenkel\\_Damill.pdf](http://www.bcra.gov.ar/pdfs/investigaciones/PaperFrenkel_Damill.pdf)> (noviembre 2011).
- Enders, W. (2004). *Applied Econometric Time Series*, 2<sup>nd</sup> Ed. Wiley & Sons, INC.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.
- Gulli, I. (2005). “Ley de Okun descomposición de las fluctuaciones económicas”. Documento de trabajo, Asociación Argentina de Economía Política. Recuperado de <[www.aeep.org.ar/espa/anales/workso5/gulli.pdf](http://www.aeep.org.ar/espa/anales/workso5/gulli.pdf)>.
- Ibarra, C. (2011). “México: la maquila, el desajuste monetario y el crecimiento impulsado por las exportaciones”. *Revista CEPAL*, núm. 104, agosto.
- Krugman, P y L. Taylor (1978). “Contractionary Effects of Devaluations”. *Journal of International Economics*, 8: 445-56.
- (1989). *La inestabilidad de la tasa de cambio*. Massachusetts: MIT Press.
- (1997). *Currency Crises*. Recuperado de <<http://web.mit.edu/krugman/www/crises.html>>.
- Loría, E. (2001). “La restricción externa dinámica al crecimiento de México a través de las propensiones del comercio, 1970-1999”. *Estudios Económicos*, vol. 16, núm. 2, julio-diciembre. El Colegio de México.
- (2003). “The Mexican Economy: Balance of Payment Constrained Growth Model. The Importance of the Exchange Rate, 1970-1999”. *Journal of Post Keynesian Economics*, vol. 25, núm. 4. Summer.
- Thirlwall, A. P. (1997). *Macroeconomics Issues from a Keynesian Perspective Selected of A.P. Thirlwall*, vol. II. Edward Edgar, GB.
- Thirlwall, A. P y M. N. Hussein (1982). “The Balance of Payments Constraint, Capital Flows and Growth Rates Differences between Developing Countries”. *Oxford Economic Papers*, vol. 34, pp. 498-509.
- Sims, C. A. (1980). “Macroeconomics and Reality”. *Econometrica*, vol. 48, núm. 1. January.
- Stiglitz, J. (2002). “Argentina’s Lessons”. *Project Syndicate*. January. Recuperado de <<http://www.project-syndicate.org>>.
- Valdés, R. (1997). *Transmisión de la política monetaria en Chile*. Documento de trabajo No. 16. Banco Central, Chile.

## ANEXO

**Cuadro III.1A**  
**Pruebas de raíces unitarias**

	<i>ADF</i>	<i>DF-GLS</i>	<i>PP</i>	<i>KPSS</i>
<i>y</i>	-3.10 <sup>1</sup>	-3.91 <sup>1</sup>	-2.92 <sup>5</sup>	0.06 <sup>9</sup>
<i>ca</i>	-2.73 <sup>1</sup>	-2.30 <sup>3</sup>	-2.75 <sup>6</sup>	0.08 <sup>10</sup>
<i>tcr</i>	-3.72 <sup>2</sup>	-1.71 <sup>1</sup>	-3.76 <sup>7</sup>	0.53 <sup>11</sup>
<i>tcr54</i>	-8.09 <sup>2</sup>	-6.69 <sup>4</sup>	-8.25 <sup>8</sup>	0.12 <sup>12</sup>

Pruebas válidas al 99% de confianza.

ADF y PP sin intercepto ni tendencia; DF-GLS y KPSS con intercepto. <sup>1</sup>sin rezagos; <sup>2</sup>con intercepto y sin rezagos; <sup>3</sup>con intercepto y sin rezagos válida al 95% de confianza; <sup>4</sup>con intercepto, tendencia y sin rezagos; <sup>5</sup>con un rezago; <sup>6</sup>con dos rezagos; <sup>7</sup>con intercepto y dos rezagos; <sup>8</sup>con intercepto y diez rezagos; <sup>9</sup>con intercepto, tendencia y dos rezagos; <sup>10</sup> con tres rezagos; <sup>11</sup> con cinco rezagos; <sup>12</sup>con siete rezagos.

**Cuadro III.2A**  
**Pruebas conjuntas del VAR irrestricto**

	<i>Normalidad</i>			<i>Autocorrelación</i>	<i>Heteroscedasticidad</i>
	<i>Sesgo</i>	<i>Kurtosis</i>	<i>Jarque-Bera</i>	<i>LM (4)</i>	<i>Términos no cruzados</i>
Conjunta	5.18 (0.16)	0.37 (0.94)	5.55 (0.47)	8.48 (0.49)	115.45 (0.44)

**Cuadro III.3A**  
**Pruebas conjuntas del VAR irrestricto alternativo**

	<i>Normalidad</i>			<i>Autocorrelación</i>	<i>Heteroscedasticidad</i>
	<i>Sesgo</i>	<i>Kurtosis</i>	<i>Jarque-Bera</i>	<i>LM (4)</i>	<i>Términos no cruzados</i>
Conjunta	3.43 (0.32)	14.58 (0.02)	36.65 (0.06)	4.57 (0.87)	225.82 (0.41)

**Cuadro III.4A**  
**Causalidad en el sentido de Granger**

<i>Variable dependiente TCR</i>			
<i>Excluidos</i>	$\chi^2$	<i>Df*</i>	<i>Prob.</i>
<i>ca</i>	2.335001	3	0.5058
<i>y</i>	0.494328	3	0.9201
Conjunta	4.059548	6	0.6686
<i>Variable dependiente ca</i>			
<i>Excluidos</i>	$\chi^2$	<i>Df*</i>	<i>Prob.</i>
<i>tcr</i>	20.03220	3	0.0002
<i>y</i>	3.053138	3	0.3835
Conjunta	28.34862	6	0.0001
<i>Variable dependiente y</i>			
<i>Excluidos</i>	$\chi^2$	<i>Df*</i>	<i>Prob.</i>
<i>tcr</i>	15.55839	3	0.0014
<i>ca</i>	6.861845	3	0.0764
Conjunta	22.49321	6	0.0010

\* Df: Grados de libertad.



# CUANTIFICACIÓN DE VARIABLES PARA EVALUAR RIESGOS DE FALLA EN PRESAS DE TIERRA\*

*David Joaquín Delgado Hernández\*\**, *David De León Escobedo\*\**,  
*Oswaldo Morales Nápoles\*\*\**, *Benjamín Pérez Pliego\*\**  
*y Juan Carlos Arteaga Arcos\*\**

## Resumen

Un sistema complejo es aquel cuyos elementos interactúan mutuamente de distintas formas. Cuando las interacciones son no lineales, el análisis del sistema requiere la identificación de los patrones de comportamiento entre ellos mediante variadas estrategias. Las presas de tierra, obras de infraestructura relevantes para el desarrollo de un país, útiles para el abastecimiento de agua potable, riego, generación de energía eléctrica, actividades recreativas y control de inundaciones, representan un sistema complejo no lineal. Así, es necesario evaluar sus riesgos de falla, identificando tanto las variables que intervienen en sus mecanismos de colapso como sus interacciones. En este capítulo se presentará la metodología de cuantificación de cuatro variables: a) frecuencia sísmica, b) precipitación, c) tubificación y d) desbordamiento, a través de dos estrategias: a) información estadística y b) juicio estructurado de expertos. Los resultados muestran que ambas son útiles en la construcción de distribuciones de probabilidad que pueden ser utilizadas como insumos de modelos matemáticos como las Redes Bayesianas.

**Palabras clave:** Desbordamiento, experto, juicio, precipitación, presas, Red Bayesiana, riesgo, sismo, tubificación.

\* Los autores agradecen por el apoyo brindado a los cuatro expertos que tomaron parte en este estudio; de igual manera, por el soporte económico, a la Universidad Autónoma del Estado de México, al Consejo Mexiquense de Ciencia y Tecnología (COMECyT) y al Consejo Nacional de Ciencia y Tecnología (Conacyt, por medio de los proyectos 158225 & 145462). Oswaldo Morales Nápoles agradece por el financiamiento parcial recibido por parte de TU Delft y TNO.

Los autores también agradecen a los alumnos David Carlos Pérez-Flores y José Emmanuel Rivero-Santana por su participación en el proyecto del cual se desprende este capítulo de libro.

\*\* Profesor de Tiempo Completo, Facultad de Ingeniería, Universidad Autónoma del Estado de México, Cerro de Coatepec s/n, C.P. 50130, Toluca. E-mails: <david.delgado@fi.uaemex.mx>, <jcartegaa@uaemex.mx>, <daviddeleonescobedo@yahoo.com.mx> y <benja13\_min99@hotmail.com>.

\*\*\* Investigador de Tiempo Completo. Organización de los Países Bajos para la Investigación Científica Aplicada (TNO-Delft). Faculty of Electrical Engineering, Mathematics and Computer Science TU Delft Delft Institute of Applied Mathematics Mekelweg 4 2628 CD Delft, Netherlands. B06.160 +31 15 27 84563. E-mail: <oswaldo.moralesnapoles@tno.nl>.

## INTRODUCCIÓN

Cuando se hace referencia a “sistemas no lineales complejos” se conciben sistemas que incluyen una gran cantidad de agentes que interactúan entre sí de múltiples maneras y que, desde luego, presentan un comportamiento no lineal (Waldrop, 1993). A manera de ejemplo, y tomando como referencia el sistema utilizado como caso de estudio que se presentará a lo largo de este capítulo, se observa lo siguiente: en una presa de tierra existen diversas incertidumbres como la composición y resistencia de sus materiales de construcción, el tipo de suelo donde se ha colocado, y las fuerzas a que estará sometida durante su vida útil, lo que constituye un sistema complejo.

Por ello, es necesario realizar evaluaciones del riesgo que permitan cuantificar la magnitud de las consecuencias que se tendrían en la eventualidad de una falla. Las metodologías que se han desarrollado para tal fin son también sistemas complejos, donde un conjunto de variables se integran en un modelo para simular el comportamiento del sistema real. En virtud de que la perspectiva dada al capítulo se orienta a la construcción de distribuciones de probabilidad útiles para ser empleadas en Redes Bayesianas,<sup>1</sup> se resalta que dichas redes implican la cuantificación de variables que generalmente son no lineales (ejemplo: frecuencia sísmica y desbordamientos), el cálculo de correlaciones entre ellas, el uso de estadística inferencial, la aplicación del Teorema de Bayes y el uso de cópulas, por lo que se trata de sistemas complejos no lineales.

Es oportuno aclarar desde ahora que los conceptos que aquí serán expuestos no pretenden establecer una metodología formal para la evaluación del riesgo, lo cual implicaría la estimación de indicadores como la pérdida anual esperada (PAE) o la pérdida máxima probable (PMP) derivada de la falla de una presa. Más bien el enfoque pretende mostrar una forma simplificada que brinde una primer idea de cómo cuantificar algunas variables involucradas en la evaluación del riesgo en presas de tierra, para que estas estructuras puedan ser identificadas como posibles candidatas para análisis de mayor detalle; las cuantificaciones de las variables que aquí se presentarán son insumos necesarios para la evaluación formal del riesgo, el cual se encuentra fuera del alcance del presente trabajo.

Con estas ideas en mente, se señala que las presas son estructuras naturales o construidas por el hombre, a las cuales se les dan diversos fines en beneficio del ser humano (ejemplo: generación de energía eléctrica, riego y consumo humano

<sup>1</sup> Las Redes Bayesianas son sólo una de las metodologías disponibles para la evaluación del riesgo, pero existen otras como el FORM y el SORM (Métodos del Primero y Segundo Orden del Momento Estadístico) y la simulación de Monte Carlo, mismos que también pueden considerarse como sistemas complejos.

de agua). En el caso de las estructuras artificiales, los métodos de diseño y construcción han evolucionado a lo largo del tiempo, así como las actividades para las que se usan.

Consecuentemente, resulta indispensable cuantificar los riesgos en presas de tierra para generar planes de evacuación, y proponer estrategias de mantenimiento que permitan que estas estructuras se mantengan operando adecuadamente. En este capítulo se presentarán dos metodologías<sup>2</sup> para cuantificar algunas variables que pueden producir la falla de presas, con la intención de emplearlas posteriormente en la evaluación numérica de riesgos, aunque se insiste que dicha evaluación cae fuera del alcance del presente trabajo.

Para ello, primero se establecerán algunos conceptos relacionados tanto con el tema de presas como con el de riesgos. Acto seguido, se describirá la cuantificación simplificada de cuatro variables: a) frecuencia sísmica, b) precipitación, c) tubificación y d) desbordamiento. El análisis de siete presas en el Estado de México ha sido fundamental para realizar la estimación de las distribuciones de probabilidad de ocurrencia de cada una de esas variables. Similarmente, el uso del Juicio Estructurado de Expertos (JEE) ha jugado un rol determinante en dicha cuantificación, como se describirá más adelante. Finalmente se proponen ideas para utilizar las distribuciones encontradas en análisis cuantitativos, para estimar el riesgo de falla de una presa de tierra.

## I. CONCEPTOS Y DEFINICIONES: PRESAS

La literatura reporta varias definiciones de lo que es una presa. Por ejemplo, la Comisión Internacional de Grandes Presas (ICOLD) la define como:

[...] una obstrucción artificial al flujo natural de agua que se construye con uno o más propósitos específicos como la acumulación de agua para la irrigación, la generación de energía eléctrica, la creación de un lago artificial para la navegación o actividades recreativas, el abastecimiento de agua a las ciudades o a la industria, la prevención de inundaciones, la desviación de ríos hacia canales, y para tener una reserva de agua fresca (2008).

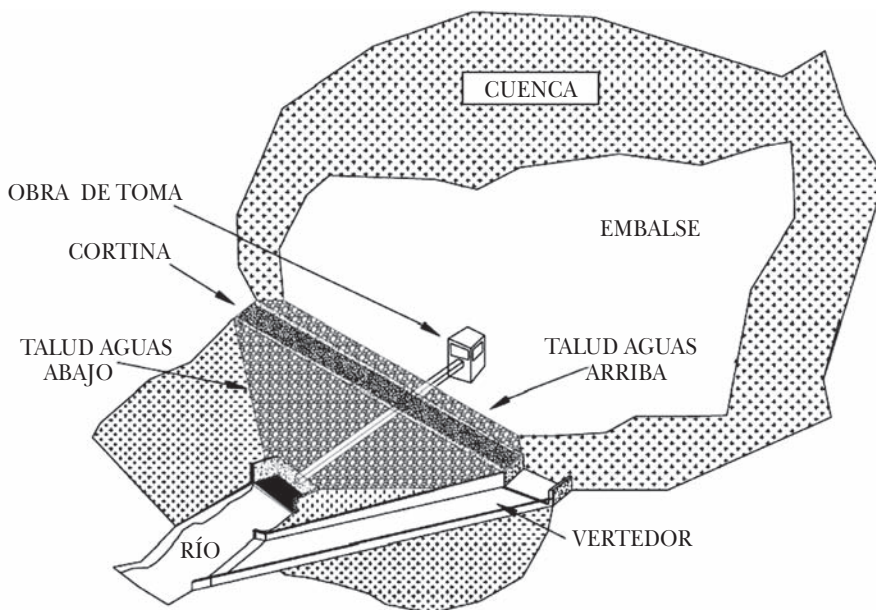
Por otra parte, en México, la Comisión Nacional del Agua (Conagua) afirma que “[...] las presas son barreras artificiales que están construidas generalmente con materiales térreos o a base de concreto con la finalidad de controlar o almacenar agua para una gran variedad de usos. Se pueden colocar a través de una corriente de agua o fuera del cauce” (1999).

<sup>2</sup> La primera hace uso de la estadística y la segunda de expertos.

Regularmente, una presa está constituida por distintos elementos que aseguran su estabilidad y buen funcionamiento. La forma y características de dichos elementos dependen del tipo de estructura, si bien en general las presas se encuentran integradas por cortina, vertedor, embalse, obra de toma y cuenca. La Figura I.1 muestra de manera esquemática las partes de una presa, así como su disposición más común.

Esencialmente, se reconocen tres tipos de cortinas: rígidas (como las de concreto, que pueden ser de arco o de contrafuertes así como las de las presas denominadas *de gravedad*, las cuales son presas que mantienen su estabilidad debido a su gran masa), flexibles (de materiales granulares como tierra y enrocamiento) y mixtas, que son la combinación de ambas (más detalles pueden encontrarse en Conagua, 1999).

**Figura I.1**  
**Elementos principales de una presa de tierra**

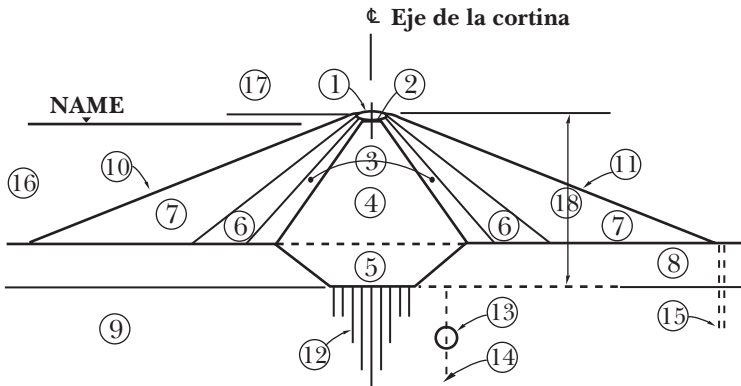


En cuanto a las presas construidas durante el siglo XX a nivel internacional, Marengo (2002) reportó la existencia de presas de concreto, materiales sueltos y otras que combinaban ambos. Asimismo, mencionó que de las 15,800 presas construidas mundialmente hasta 1975, 62.6% eran de tierra. En México se cons-

truyeron en el periodo de 1550 a 1998 un total de 1,017 presas, de las cuales casi 70% fueron de materiales sueltos, y el restante 30% lo constituyen conjuntamente las presas de gravedad, arco y de contrafuertes. En virtud de que las presas de tierra son las más abundantes en México, en seguida se describen sus principales características (Conagua, 1999).

Para que una presa se considere de tierra es necesario que más del 50% de su volumen sea de suelos compactados. La principal característica de sus componentes de construcción es que son muy permeables, por lo que es necesario añadirles elementos con características impermeables en la zona de la cortina y cimentación, para disminuir la pérdida de agua, y de este modo asegurar la estabilidad y funcionalidad de la estructura. Los materiales empleados para suscitar dicha condición de impermeabilidad pueden variar desde arcillas (en cuyo caso siempre se ubican en el corazón del relleno) hasta pantallas de concreto armado,

**Figura I.2**  
**Partes que integran una presa de tierra**



- |                                |                            |                         |
|--------------------------------|----------------------------|-------------------------|
| 1 Cresta o corona              | 7 Enrocamiento             | 13 Galería              |
| 2 Revestimiento de la corona   | 8 Depósito aluvial         | 14 Drenes               |
| 3 Filtros                      | 9 Roca basal               | 15 Pozos de aluvio      |
| 4 Corazón o núcleo impermeable | 10 Talud de aguas arriba   | 16 Embalse              |
| 5 Trinchera                    | 11 Talud de aguas abajo    | 17 Bordo libre          |
| 6 Transiciones                 | 12 Pantalla de inyecciones | 18 Altura de la cortina |

**NAME:** Nivel de Aguas Máximas Extraordinarias

las cuales se pueden construir ya sea en el centro del relleno o bien sobre el talud de la cortina “aguas arriba”.<sup>3</sup>

El mecanismo de resistencia de este tipo de estructuras es por gravedad, por lo que normalmente son obras con dimensiones considerables, con taludes muy extendidos, que requieren grandes volúmenes de material para su construcción. En la Figura I.2 se pueden observar las diferentes partes que integran una presa de tierra.

Estos componentes se resumen en cimentación, núcleo impermeable y taludes permeables cubiertos, en la zona aguas arriba, con una capa permeable. Su principal beneficio es la economía que representa construirlas con los materiales naturales de la zona, en comparación con sus contrapartes de concreto.

Respecto a las desventajas que implica la construcción de presas de tierra, se puede mencionar que son estructuras vulnerables (es decir, que pueden sufrir daños bajo la presencia de cierto tipo de peligros o amenazas) a la sobre-elevación de los niveles del embalse, originados por lluvias extraordinarias. Asimismo, son vulnerables a filtraciones y erosión interna en la cortina y cimentación. Precisamente la probabilidad de que se presenten este tipo de variables se cuantificará más adelante (desbordamiento y tubificación). Pero antes de ello, se describen ahora algunos conceptos relacionados con el tema de riesgos.

## II. CONCEPTOS Y DEFINICIONES: RIESGOS Y CONFIABILIDAD

En las etapas de diseño y construcción de obras de ingeniería civil debe tomarse en cuenta la posibilidad de que los proyectos pongan en riesgo la salud o seguridad pública, o provoquen pérdidas económicas a la comunidad (Merrit *et al.*, 1999). De acuerdo con el estándar británico BS 8800 (1996), un “riesgo” es la combinación de la probabilidad de que un evento peligroso ocurra y el efecto de sus consecuencias. La misma fuente define “análisis de riesgo” como el proceso para estimar la magnitud de un riesgo y decidir si es o no aceptable. En consecuencia, el término “administración de riesgo” se refiere al proceso mediante el cual la sociedad u organización reduce los riesgos a un nivel tolerable y se asegura de controlarlos, monitorearlos y comunicarlos a los posibles afectados (Renn, 1998).

En el contexto de la Ingeniería Civil, el concepto “riesgo” se define como las consecuencias esperadas asociadas a la ocurrencia de cierto evento; estas consecuencias son medidas generalmente en términos monetarios, pérdida de vidas humanas y consecuencias negativas (Faber, 2009).

<sup>3</sup> Se denomina “aguas arriba” a todas las obras construidas que se encuentran del lado del embalse; el talud aguas arriba está en contacto directo con el agua. En contraste, “aguas abajo” son las obras de toma y de la propia cortina que no se encuentran en contacto con el agua del embalse.

Así, la definición matemática de riesgo está dada por la expresión:

$$\text{Riesgo} = \text{Consecuencias adversas esperadas}$$

Para un solo evento adverso ( $E$ ), como por ejemplo la presencia de un sismo, el riesgo se calcula como el producto de la probabilidad de la ocurrencia del evento  $P_r(E)$  por las consecuencias de dicho evento  $C(E)$ :

$$R(E) = P_r(E) \times C(E)$$

Entonces el riesgo se cuantifica en las mismas dimensiones que las consecuencias, es decir, dinero, pérdidas humanas o consecuencias negativas (Straub, 2010).

El análisis de riesgo es, desde un punto de vista formal, una metodología empleada para determinar y evaluar la magnitud del riesgo. Cuando se combina con la planeación de diferentes acciones se puede definir como la toma de decisiones basadas en el riesgo (Straub, 2010).

El desarrollo y la protección de infraestructura impactan directamente al desarrollo del país. Por lo tanto, es necesario administrar los riesgos involucrados y conservar las edificaciones para garantizar que no causarán fatalidades ni pérdidas económicas a la comunidad como resultado de su eventual falla.

Por otro lado, uno de los parámetros comúnmente determinados en el proceso de evaluación de riesgo es el Índice de Confiabilidad, que se puede definir como el valor inverso de la función normal estándar acumulativa ( $\beta$ ) de la probabilidad de falla ( $P_f$ ), y se expresa como:

$$\beta = \Phi^{-1}(P_f)$$

La confiabilidad, para fines de ingeniería, relaciona fuerzas actuantes contra fuerzas resistentes tomando en consideración valores aleatorios que pudieran asumir las fuerzas; es decir, contempla la incertidumbre en el cálculo de acciones y respuestas.

El índice de confiabilidad, dentro del proceso de análisis de riesgo, es empleado para estimar la probabilidad de ocurrencia de las consecuencias. De nuevo, algunas de las principales técnicas empleadas para calcularlo son las Simulaciones de Montecarlo y los Métodos del Primero y Segundo Orden del Momento Estadístico (Silva y Heredia, 2007).

En materia de fallas, a mediados de la década antepasada la Comisión Internacional de Grandes Presas (ICOLD, 1995) reportó que en términos absolutos la mayor frecuencia de rupturas se presenta en estructuras de tierra (85% de los

casos estudiados), en contraste con otros tipos de presas (10% de gravedad, 3% de arco y 2% de contrafuertes).

Por ello, la evaluación del riesgo en presas de tierra en épocas actuales ha cobrado mayor importancia. Consecuentemente, entender los conceptos de administración del riesgo, confiabilidad y falla en términos de presas, es fundamental para mantenerlas seguras y cumpliendo sus funciones sin representar una amenaza para la sociedad.

Con estas ideas presentes se describirá ahora el Juicio Estructurado de Expertos (JEE), metodología que permitirá evaluar dos de las variables posteriormente analizadas.

### III. JUICIO ESTRUCTURADO DE EXPERTOS (JEE)

El JEE permite llegar a un consenso racional con base en el método científico, a través de la evaluación y validación de juicios emitidos por diferentes expertos (Cooke y Gossens, 2008). Esto partiendo de la premisa de que los especialistas cuentan con los conocimientos teóricos y prácticos sobre un tema específico, y que dominan las teorías y valores de ciertos parámetros de interés, mismos que no pueden ser experimentalmente medidos (como por ejemplo el desbordamiento de una presa y los daños ocasionados a las personas y sus bienes, al ambiente y las obras de infraestructura, pues sería muy costoso realizar un experimento que reflejara este tipo de acontecimientos). La incapacidad para cuantificar dichos parámetros responde a la presencia de restricciones de diferente índole, como la inaccesibilidad a datos.

La motivación del JEE es la insuficiencia o inexistencia de información en determinado campo del conocimiento. Por ello es necesario que el investigador obtenga datos directamente con los expertos en un tema, mediante una metodología sistemática, capaz de procesar las opiniones de los individuos, de manera que la información obtenida sea válida. Al realizar el ejercicio, los resultados deberán reflejar la incertidumbre de cada experto, y la información proporcionada por ellos debe ser obtenible a través de un experimento real que no se efectúa debido a las implicaciones o consecuencias que tendría; esta técnica no es nueva y tiene algunas variantes.

El llamado “modelo clásico”, que será utilizado en este trabajo, fue documentado de manera formal por Cooke (1991). Recientemente se llevó a cabo un recuento de las características principales de dicho modelo y se encontró que había sido aplicado en las industrias: nuclear, aeronáutica, química y de gas, construcción, vulcanología y salud (Cooke y Gossens, 2008), donde queda de manifiesto el desarrollo que este enfoque ha tenido durante casi dos décadas. Es importante

mencionar que, aunque parece una técnica netamente subjetiva, el método científico impone las siguientes condiciones para garantizar su validez como herramienta de cuantificación: (a) capacidad de escrutinio y confiabilidad; (b) control empírico; (c) neutralidad; y (d) equidad (Morales-Nápoles y Cooke, 2009).

El nombre “modelo clásico” tiene su origen en una analogía entre las pruebas de hipótesis de la estadística clásica y el cálculo de calibraciones en el JEE. Un concepto clave en dicho modelo es el de variables de calibración, que son aquellas cuyo valor verdadero, en el momento de la consulta, es conocido por el analista pero no por el experto. Por ejemplo, el analista puede averiguar la magnitud de un terremoto ocurrido en una región en una fecha específica y preguntarle al experto sobre dicha magnitud. Si el especialista responde correctamente, o se aproxima al valor verdadero de manera importante, entonces resultará bien calibrado.

Básicamente, los expertos son cuestionados acerca de variables que pueden tomar valores inciertos en un rango continuo, y generalmente se les preguntan los percentiles 5, 50 y 95 de su distribución subjetiva de incertidumbre. Cabe recordar que, una vez que se ordenan ascendentemente, el percentil 5 corresponde al valor que está por encima del 5% de los datos. De igual forma, el percentil 50 es aquel que está arriba del 50% de los datos (y por tanto por debajo del otro 50%), y así sucesivamente. Para extraer la información, los especialistas se enfrentan a “preguntas de interés” tales como: “Tome como referencia una eventual inundación debida a la falla de la presa bajo estudio. Ahora considere los costos públicos y privados totales, para un periodo de tiempo igual a la máxima esperanza de vida promedio restante de las personas afectadas, por todos los posibles daños y pérdidas en salud y vidas humanas. ¿Cuáles son el 5, 50 y 95 percentiles de su distribución de incertidumbre (medidos en pesos (\$) actuales)?”.

Así, el experto podría dar los siguientes valores: \$20,000, \$3'000,000 y \$19'000,000 respectivamente, lo cual indicaría que tiene una gran incertidumbre, como lo muestra la separación entre los percentiles, evidenciada por el rango tan amplio entre un valor y otro. La Figura III.1 presenta la distribución hipotética que se podría generar con estos datos.

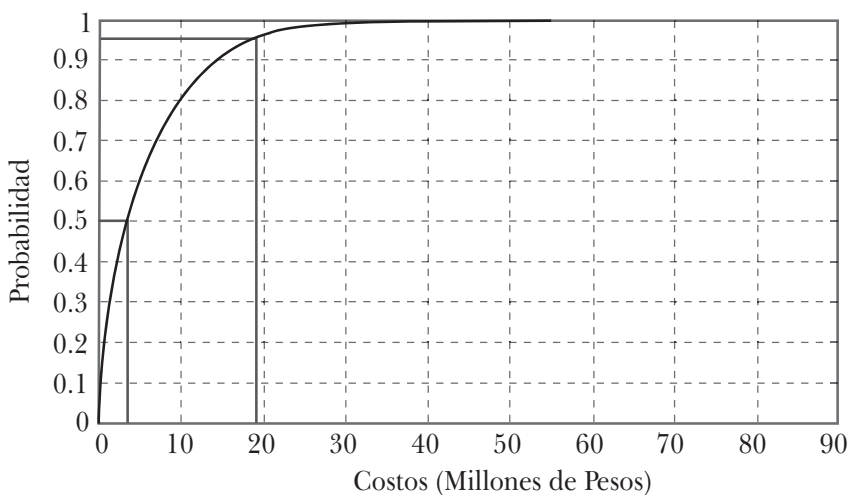
Las “preguntas de calibración”, de manera similar, presentan un formato idéntico al planteamiento descrito, aunque cuestionan aspectos de los que el analista cuenta con datos *a priori*. Intuitivamente se puede decir que las preguntas de calibración miden la probabilidad de que un conjunto de resultados experimentales correspondan, en un sentido estadístico, con las respuestas brindadas por los especialistas. Ésta alcanza su valor máximo en 1, y valores cercanos a cero indican que es poco probable que la información proporcionada por el experto sea adecuada en un sentido estadístico bajo el supuesto de que las preguntas de calibración representan la experiencia deseada en el grupo. En contraste con las

de calibración, las preguntas de interés miden el grado en el que una distribución está “concentrada” o “extendida” con relación a una distribución de probabilidad base, como la uniforme o la log-uniforme.

Una vez que las preguntas de interés y calibración son obtenidas, es necesario asignar pesos específicos a cada experto, con base en sus puntuaciones. Estos pesos son usados para calcular promedios ponderados (combinación lineal), los cuales representarán la opinión del grupo, la cual se denomina como “tomador de decisiones” (TD).

En el presente estudio el Tomador de Decisiones por Pesos Globales (TDPG) y el Tomador de Decisiones por Pesos Iguales (TDPI) han sido considerados. El primero

**Figura III.1**  
**Distribución de probabilidad acumulada hipotética de los costos económicos humanos generados por la falla de una presa**



se calcula con base en los pesos específicos de cada experto, dándole más importancia al especialista que obtuvo la mayor ponderación. En contraste, el segundo asigna pesos iguales a cada experto, por lo que se trata de la media aritmética. Todos aquellos expertos cuyo peso específico esté por encima de un nivel de significancia  $\alpha = 0.05$  serán considerados en el TDPG. Obsérvese que los pesos específicos se obtienen mediante el producto de las puntuaciones de información y calibración.

De esta forma, en la presente investigación se entrevistó a cuatro expertos con amplios conocimientos tanto de las presas en estudio como de la región donde se ubican. Los cuatro se habían desempeñado en el ámbito profesional en la administración de presas, y dos de ellos eran además profesores de ingeniería. Es impor-

tante mencionar que el proyecto<sup>4</sup> del que se deriva la información aquí mostrada es más amplio, y que el análisis presentado, para fines del presente libro, considera sólo una parte, la referente al análisis de las cuatro variables de interés.

Por ello, aquí sólo se requirió que cada uno de los cuatro especialistas proporcionaran información para cuantificar dos variables, y respondieran 20 preguntas de calibración, similares a las del ejemplo ya mencionado. Las otras dos variables que se presentarán se pudieron calcular de forma simplificada con datos estadísticos, por ejemplo, las tasas de precipitación y la frecuencia sísmica de las zonas donde se localizan las siete presas.

Así, se obtuvieron los resultados que se resumen en Tabla III.1. Los valores presentados se calcularon por medio del software especializado llamado EXCALIBUR (EXpert CALIBration), que fue desarrollado en la Universidad Tecnológica de Delft, Holanda (TU Delft).

Como se puede observar, en la primera columna se presentan las claves de los cuatro expertos participantes en el estudio (A, B, C y D), además de los dos expertos virtuales (TDPG y TDPI, que se reitera son calculados como promedios, a partir de los datos de los expertos “reales”). En la segunda, se aprecia la puntuación de calibración (PC) de los “seis” especialistas, donde el TDPI resultó ser mayor a 0.05 (0.07164). En la práctica, PCs menores a ese valor fallarían en otorgar el nivel de confianza requerido en el estudio. En cambio, puntuaciones de calibración superiores a 0.05 son suficientes para soportar la hipótesis del modelo clásico.

**Tabla III.1**  
**Resumen de indicadores del modelo clásico para los expertos**  
**de riesgos en presas de tierra del Estado de México**

<i>Expertos</i>	<i>Puntuación de Calibración (PC)</i>	<i>Puntuación de Información (PI) tomando en cuenta todas las variables</i>	<i>Puntuación de Información (PI) tomando en cuenta sólo las preguntas de calibración</i>
A	1.38E-04	0.92	0.83
B	3.95E-14	2.25	2.20
C	3.22E-09	1.51	1.58
D	3.57E-07	0.09	0.87
TDPG	9.21E-04	0.84	0.76
TDPI	7.16E-02	0.30	0.33

<sup>4</sup> Proyecto de Investigación: Delgado-Hernández, D. J. y De León, E. D. (2007), “Análisis de riesgo, confiabilidad estructural y mantenimiento de presas de tierra: un caso en el Estado de México”, Proyecto UAEM 2491-2007U, Universidad Autónoma del Estado de México, 14 nov. 07-14 nov 08.

De acuerdo con la misma tabla, la combinación de opiniones de los expertos es mejor que la de cada uno individualmente. Tanto el TDPG como el TDPI alcanzaron una calibración mayor que la de cada especialista por separado; el TDPI supera el nivel de confianza mínimo requerido en el enfoque (0.05), y es por tanto el que será usado en los cálculos posteriores.

La tercera y cuarta columnas presentan la Puntuación de Información (PI) de los expertos para todas las variables, y para cuando sólo se toman en cuenta las preguntas de calibración. El análisis de los datos muestra que el experto más informativo (B) es también el más bajo en la calibración. Éste ha sido un patrón recurrente en estudios anteriores (Morales-Nápoles y Cooke, 2009).

Pese a que el TDPI fue el mejor calibrado en el ejercicio, resultó ser también el menos informativo. No obstante, se reitera que es el único que excede el nivel de significancia requerido, por lo que el resto del análisis se realizará con base en los resultados de este tomador de decisiones. Las dos variables de interés (desbordamiento y tubificación) han sido tomadas en cuenta como distribuciones marginales de probabilidad. Habiendo presentado los conceptos detrás del JEE, en seguida se describe la muestra de presas utilizadas en la investigación.

#### IV. PRESAS EN EL ESTADO DE MÉXICO

En materia de presas, estudios estadísticos demuestran que las estructuras de tierra con más de 30 años de edad, y entre 15 y 30 m de altura, son las que presentan mayores índices de falla (Marengo, 2002). Por ello, se tuvo que investigar cuáles eran las presas en el Estado que cumplían simultáneamente con estos criterios de selección (edad, altura de cortina y material de construcción).

En la revisión de la literatura se encontró que, hasta 1976, la entidad contaba con 53 estructuras en general (SRH, 1976). De ellas, cinco (Huapango, San Antonio, Nádó, La Loma y El Molino) fueron construidas antes de 1900, y las otras 48 durante el siglo XX. A su vez, 19 se construyeron de tierra y enrocamiento entre 1942 y 1972. Así, se identificaron siete presas, localizadas en los municipios del Estado de México, particularmente en el valle de Toluca y en las proximidades del Distrito Federal, que cumplieron con los criterios de selección y que estadísticamente serían las más vulnerables (Foster *et al.*, 2000). Las estructuras seleccionadas fueron:

1. Embajomuy,
2. San Joaquín,
3. José Trinidad Fabela,

4. Dolores (La Gavia),
5. José Antonio Alzate (San Bernabé),
6. Ignacio Ramírez, y
7. El Guarda.

En la Figura IV.1 se muestra su localización geográfica. Se ha colocado un número sobre la presa correspondiente, a partir de la lista anterior, para apreciar su ubicación.

Los detalles constructivos y especificaciones de diseño de cada una de las siete estructuras se pueden consultar en SRH (1976), resaltando que además de los tres criterios mencionados, comparten otras particularidades como la escasa conservación, detectada por los investigadores durante la visita física a cada una de las obras. Por la cercanía que existe entre ellas y su ubicación geográfica, todas están sometidas a los dos peligros importantes expuestos en seguida.

**Figura IV.1**  
**Localización geográfica de las presas bajo estudio**



## V. PELIGROS

Dentro del alcance del presente trabajo, y en función de los fenómenos que pueden afectar la integridad estructural de las presas, los principales tipos de peligro o amenaza que se tomaron en consideración en la metodología aquí presentada fueron los naturales. Según lo establecido por Mileti, en el contexto de análisis y evaluación de riesgos, peligro o amenaza natural, se pueden definir como “fenómenos meteorológicos o geológicos extremos de baja probabilidad de ocurrencia que tienen el potencial de causar desastres, cuando se manifiestan sobre colectividades humanas” (1999: 376). En este caso los peligros tomados en consideración fueron la frecuencia sísmica y la precipitación, variables que se describen a continuación.

### V.1 Frecuencia sísmica

De acuerdo con McGuire (2008), el Análisis Probabilístico de Riesgo Sísmico (APRS) es la evaluación de la frecuencia anual de posibles excedencias de los niveles del movimiento del terreno en un sitio determinado. Este análisis da como resultado una curva de peligro sísmico (frecuencia anual de excedencia *vs.* amplitud del movimiento del terreno).

Uno de los primeros trabajos relacionados con estas estimaciones es el realizado por Cornell (1968); básicamente el cálculo de la frecuencia anual de excedencias se realiza conociendo algunas características particulares del terreno en el sitio de estudio, la distancia a la fuente (o fuentes) sísmica, así como la intensidad en escala Richter de los sismos producidos en las fuentes que mayoritariamente impactarían en el sitio de interés. Información detallada referente a la metodología APRS puede ser encontrada en Ellinwood (2001), Esteva y Ruiz (1989) y Ordaz (2004), entre otros.

Debido a que el APRS implica diferentes factores, como ya se mencionó, siendo uno de éstos la intensidad Richter del sismo, para los fines del presente capítulo, y de forma simplificada, se ejemplifica la cuantificación de dicho fenómeno en términos de la generación de una variable aleatoria como sigue: el número de sismos registrados en una región y en un periodo de tiempo determinado se denomina aquí como “frecuencia sísmica”. La información recabada en el presente estudio, obtenida del Servicio Sismológico Nacional (SSN), ha permitido la construcción de las distribuciones de frecuencia necesarias. En su cuantificación se han considerado los eventos sísmicos mayores a 5.5 grados de magnitud en la escala de Richter, comprendidos entre el año 2000 y el 2008. Sólo se tomó ese periodo, en virtud de que son los registros completos a los que se pudo tener acceso. No obstante, es recomendable ampliar la base de información para incrementar la precisión del cálculo.

Los datos empleados corresponden a los registrados en la zona cercana a las estructuras de interés (Valle de Toluca). En la Tabla V.1 se muestra un resumen de la frecuencia sísmica por año, de los eventos registrados de 2000 a 2008. Asimismo, se presenta la cantidad de movimientos clasificados por su magnitud, que va de aquéllos menores a 3 grados, hasta los que alcanzaron los 8 grados (que en este periodo en particular son inexistentes).

Con base en estos datos, se construye la distribución de probabilidad de la “frecuencia sísmica”; la manera de caracterizar esta variable será mediante el número de sismos mayores o iguales a 5.5 grados en la escala de Richter, registrados por año, con algún efecto en la región de interés. Se ha elegido esa magnitud, ya que son los sismos que la superan los que pueden provocar la eventual falla de la cortina de una presa de tierra (Ramírez, 2001), y su magnitud está catalogada en dicha escala como una que deriva en daños moderados en las edificaciones.

**Tabla V.1**  
**Sismos reportados por el SSN, del periodo comprendido entre 2000 y 2008**

AÑO	TOTAL DE SISMOS	SISMO DE MAGNITUD						
		< 3	3	4	5	6	7	8
2000	1052	37	463	531	18	2	1	0
2001	1344	17	704	585	32	6	0	0
2002	1688	4	879	761	40	4	0	0
2003	1324	5	729	568	18	3	1	0
2004	945	1	429	491	24	0	0	0
2005	847	1	459	373	12	2	0	0
2006	1077	0	589	464	23	1	0	0
2007	1234	0	533	670	27	4	0	0
2008	1772	4	1037	709	18	4	0	0

Como resultado, se obtuvo una distribución marginal que es esencialmente discreta; sin embargo, se ha propuesto modificarla a una distribución continua. Para ajustarla, se encontraron los parámetros de un número de distribuciones que minimizan la suma de las diferencias de cuadrados, con respecto a una distribución empírica. Entonces, la que obtenga la menor suma de diferencias de cuadrados es elegida. De esta forma, en la Figura V.1 se muestra la distribución acumulada de probabilidad correspondiente a la variable frecuencia sísmica.

Los valores de la media y desviación estándar son 6,67 y 2,10 respectivamente, lo cual indica que es común que se tengan cuando menos entre 6 y 7 movimientos telúricos por año, con magnitudes superiores a la establecida.

**Figura V.1**  
**Distribución acumulada de probabilidad para la variable “frecuencia sísmica” (número de sismos por año mayores a 5.5 grados Richter en la región de interés)**



## V.2 Precipitación

La precipitación es cualquier forma de hidrometeoro que cae del cielo y llega a la superficie terrestre. Las dimensiones de una cuenca hidrográfica son muy variadas, y para determinar qué tanto llueve en ellas es frecuente que en una región se sitúen varias estaciones pluviométricas. Así, en lo que se refiere a la variable considerada en este capítulo, también de forma simplificada, se ha hecho uso de las mediciones registradas en las estaciones pluviométricas cercanas a las siete presas de interés.

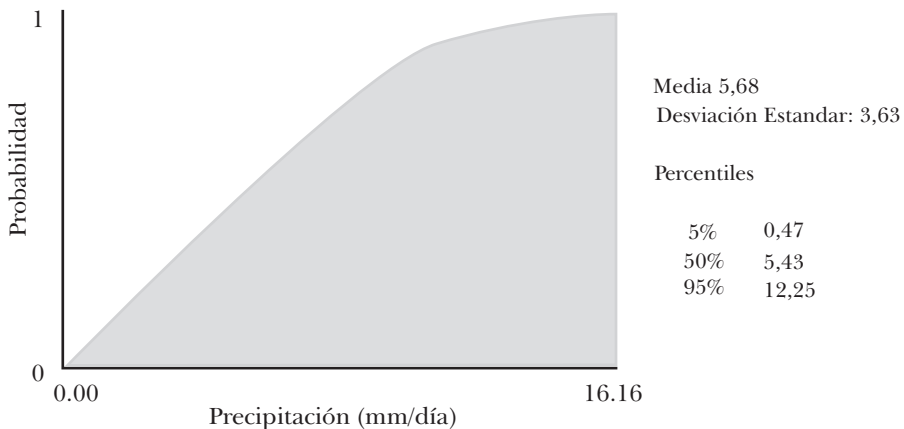
La mayoría de las fallas en presas han sido consecuencia de un evento de este tipo, es decir, lluvias extraordinarias que rebasan sus capacidades de diseño, poniendo en peligro tanto vidas humanas como bienes materiales y recursos ambientales. Consecuentemente, el hombre ha requerido desarrollar distintos métodos que le permitan predecir este tipo de fenómenos, y así poder tener mayor certidumbre en cuanto a su ocurrencia.

La intensidad de la precipitación se mide en milímetros por hora, o unidades de longitud por unidad de tiempo. En general, para determinar la cantidad de precipitación que se presenta en una cuenca, en un periodo determinado, se pueden utilizar varios métodos, siendo los más comunes los aritméticos, polígonos de Thiessen, el método de las isoyetas (Aparicio, 1994), y el de promedios móviles. Así, para el caso de estudio se utilizó el último enfoque, obteniendo primero los datos de las estaciones pluviométricas cercanas a las presas de interés, con base en el programa llamado ERIC (Extractor Rápido de Información Climatológica), el cual contaba con registros en México que comprenden del periodo 1920 a 1998.

En segundo lugar, con estos datos de precipitación se crearon promedios móviles de siete días para suavizar la distribución y eliminar los picos en semanas donde sólo había llovido uno o dos días, los cuales permitieron generar las distribuciones requeridas para sustentar el cálculo. Por ejemplo, en una semana en la que sólo había llovido 70 mm en un día, se calculó un promedio de 10 mm/día para cada día de la semana.

En la Figura V.2 se muestra la gráfica obtenida de distribución acumulada de probabilidad correspondiente a la precipitación. Dicha variable fue definida por los milímetros de lluvia promedio por día que caen en las cuencas de las presas bajo estudio. La gráfica se lee de la siguiente manera: en el eje de las abscisas se representa la precipitación promedio diaria medida en mm/día, mientras que en el eje de las ordenadas se tiene la probabilidad de ocurrencia del fenómeno.

**Figura V.2**  
**Distribución acumulada de probabilidad para la variable**  
**“precipitación” (mm/día)**



Una vez descritas las dos variables asociadas con los peligros a los que están expuestas las presas de tierra de la muestra, se procede ahora a la presentación de los conceptos de vulnerabilidad, y después a la exposición de los principales daños esperados en las estructuras analizadas.

## VI. VULNERABILIDAD

Como se ha discutido, las presas son vulnerables ya sea ante la presencia de un sismo o de una precipitación extraordinaria, es decir, pueden resultar dañadas por sus efectos. De igual forma, la infraestructura que se encuentra aguas abajo de las estructuras podría verse afectada en la eventualidad de la ruptura de una presa. En este sentido se señala que las obras existentes en las zonas cercanas a las presas de interés comparten las siguientes características: asentamientos humanos con poblaciones de unos cuantos cientos de habitantes hasta de 20,000 personas, campos de cultivo, carreteras libres de peaje y torres de transmisión de energía eléctrica.

Para calcular los daños que se podrían generar como resultado de la falla de una de ellas, Delgado *et al.* (2008) analizaron la evaluación de costos por métodos directos, indirectos y mixtos. Como resultado, concluyeron que los costos esperados para el caso de las siete presas mexiquenses, en caso de que colapsaran, podían alcanzar los \$150'000,000.00 por año. Aunque existen otras metodologías para evaluar dichos costos, como la reportada por Baró *et al.* (2007), en la que se acude a los censos del INEGI para determinar la población que habita la región, el nivel socioeconómico de los residentes, las áreas de siembra y demás infraestructura existente, en el presente capítulo no se aborda esta cuantificación con detalle. Sin embargo, sí se desarrollan dos variables relacionadas con los principales daños que pueden sufrir las presas de tierra.

## VII. PRINCIPALES DAÑOS EN LAS PRESAS DE TIERRA

Con base en la revisión de la literatura realizada (Delgado *et al.*, 2012 y Morales-Nápoles *et al.*, 2013), se concluyó que dos de los daños más relevantes para las presas de tierra eran la tubificación y el desbordamiento. Aunque existen métodos analíticos y experimentales para cuantificarlos, la falta de recursos (económicos, de tiempo y de personal) impidió su aplicación en esta parte de la investigación. Sin embargo, como se verá, el JEE es una alternativa factible para estimar las distribuciones de probabilidad de estas variables. En seguida se presentan los detalles de cuantificación para cada una.

## VII.1 Tubificación

El agua, al fluir a través de la tierra, genera fuerzas erosivas que tienden a empujar las partículas del suelo, arrastrándolas en la dirección del flujo. Al momento en que se produce este arrastre, se dice que ha comenzado el fenómeno de tubificación, el cual forma canalículos que aumentan de diámetro conforme avanza el proceso, y donde a su vez el agua fluye con mayor velocidad. Para el caso de una presa, la tubificación normalmente comienza en el talud de aguas abajo y progresa hacia atrás, es decir, en el interior de la cortina, hasta que se presenta la inminente falla de la estructura (Juárez y Rico, 1996).

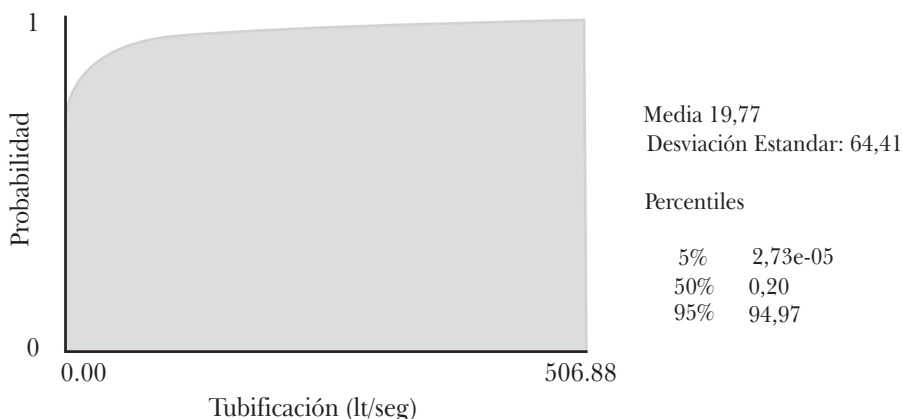
Una de las características del fenómeno es el tiempo de desarrollo, ya que puede haberse iniciado al momento de poner en operación el embalse, y mostrar indicios de su presencia muchos años después sin que ya se pueda remediar. En otras palabras, sin el monitoreo adecuado, los síntomas de una tubificación pueden ser imperceptibles.

Un ejemplo de tubificación en México se presentó en la presa de la Laguna, ubicada en el estado de Puebla, la cual forma parte del sistema hidroeléctrico de Necaxa (Marsal, 1983). En ella se registró la ruptura del empotramiento izquierdo a causa de diversas tubificaciones, provocando tanto orificios de gran diámetro como la generación de diversas brechas en la cortina. Estos daños, a su vez, causaron la destrucción de caminos y el deslizamiento masivo de taludes en la región de aguas abajo.

De este modo, estadísticamente hablando, la tubificación ha representado la segunda causa de falla en presas de tierra en el mundo después del desbordamiento, con un 25% del total de éstas entre 1964 y 1983 (Marengo, 1994). Pese a su importancia, las investigaciones en el área son relativamente escasas y se tuvieron dificultades para encontrar datos nacionales que fueran de utilidad para generar la distribución de probabilidad buscada. Por ello se tuvo que recurrir al JEE, herramienta que permitió generar la información adecuada.

Para medir la tubificación presente en el cuerpo de la cortina de la presa, se empleó el gasto que transita a través de ella, medido en lt/seg. Cabe mencionar que este fenómeno también se puede presentar en la cimentación, pero la literatura reveló que es la cortina el elemento más susceptible de verse afectado por esta razón (Foster *et al.*, 2000). Así, en la Figura VII.1 se muestra la gráfica de la distribución acumulada correspondiente a la tubificación, obtenida con la ayuda del programa UniNet, desarrollado también en TU Delft. El volumen de agua, caracterizado como la tubificación, se grafica en el eje x; y en el eje y, la probabilidad de que el fenómeno se presente, estimada con base en la opinión de los expertos, como se describió en la sección precedente del JEE.

**Figura VII.1**  
**Distribución acumulada de probabilidad para la variable**  
**“tubificación” (lt/seg)**



## VII.2 Desbordamiento

Este tipo de evento se presenta, generalmente, por una deficiente estimación del gasto de diseño del vertedor, al considerar una avenida máxima menor a la que se debe desalojar. Otras causas comunes se relacionan con el mal funcionamiento de la obra de toma o que esté fuera de servicio, el deslizamiento de laderas aguas arriba de la cortina, sismos que inducen oleaje, oleajes producto del viento, y la liberación repentina de grandes volúmenes de agua en sistemas de presas en serie. Estas fuentes provocan el aumento del nivel del agua en el embalse, que a su vez se traduce en un excedente de agua que se vierte sobre la cortina, erosionándola y dañando el talud de aguas abajo (Juárez y Rico, 1996).

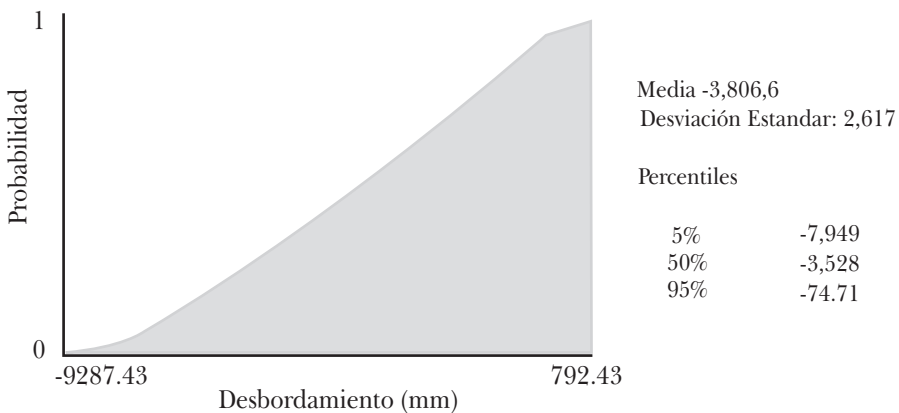
Pese a las razones naturales que pueden provocar una falla, son más bien las omisiones humanas las que más impacto pueden llegar a tener. Específicamente, en el diseño de las obras de una presa, la razón principal que genera un mal funcionamiento de sus elementos es la extrapolación de estudios previos que no presentaron fallo en sus proyectos originales. Esta práctica se debe ya sea a la insuficiencia de información para diseñar una nueva estructura, o a la simplicidad que representa hacer adaptaciones de proyectos previos similares. Como resultado, se insiste en que las estadísticas mundiales sobre las causas de falla más comunes en presas de tierra colocan al desbordamiento como la razón más

importante de sus colapsos, representando más de la mitad de las causas –55.5% del total– (Marengo, 1994).

En la investigación aquí presentada el desbordamiento tuvo que ser cuantificado a partir del JEE, ya que no se encontraron registros sobre eventos similares en la zona de interés. Cabe mencionar que, durante las entrevistas a los expertos, dos de ellos revelaron que las lluvias del verano de 2008 habían sido muy intensas y que por lo menos una presa en la entidad (ninguna de las siete estudiadas) podría presentar desbordamientos. Hasta donde se conoce, no se presentó incidente alguno en ese momento, por lo que el caso no fue investigado con mayor profundidad.

En la Figura VII.2 se muestran los valores obtenidos después de procesar la información solicitada a los expertos, para la variable correspondiente al desbordamiento. Se cuantificó en milímetros, medidos a partir de la corona de la cortina (el nivel del agua del embalse localizada en la cota de la corona significa que se tiene una altura de lámina de agua igual a cero), con el fin de medir la elevación promedio de la lámina, que se presentaría al ocurrir un eventual desbordamiento sobre la cortina.

**Figura VII.2**  
**Distribución acumulada de probabilidad para la variable**  
**“desbordamiento” (mm)**



Los valores mostrados en la Figura VII.2 corresponden a los percentiles solicitados a los expertos, es decir, 5, 50 y 95. Dichos percentiles corresponden a las cuantías  $-7,949$ ,  $-3,528$  y  $-74.71$ . Como puede observarse, los valores calculados son negativos, lo que significa que el nivel del agua estará normalmente por debajo de la cota de la corona.

En lo que se refiere a la interpretación, se puede afirmar que los expertos consideraron que, en promedio, para el 5% de los eventos presentados se tendrán valores menores a  $-7,949$  mm, es decir, cantidades que crecen en sentido negativo (ejemplo:  $-8,000$  mm). En el mismo orden de ideas, para el 50% de los casos se presentarán valores menores a  $-3,528$  mm, y para el 95% de los eventos se presentarán valores menores a los  $-74,71$  mm. De hecho, de acuerdo con la opinión de los peritos, sería sorprendente que se presentaran niveles menores a  $-9287,43$ , y mayores a  $792,43$  (ver Figura VII.2). La media y la desviación estándar corresponden a  $-3,806.06$  y  $2,617$ , respectivamente.

Habiendo presentado las cuatro variables y su cuantificación, se manifiesta que todas fueron integradas en un modelo de evaluación de riesgos en presas de tierra, cuyos detalles se encuentran en Delgado *et al.* (2012) y Morales-Nápoles *et al.* (2013). Así, se procede ahora a la discusión de resultados y al establecimiento de las conclusiones del estudio.

## VIII. DISCUSIÓN Y CONCLUSIONES

La modelación de procesos no lineales, como los cuatro analizados en este capítulo, revela que se han desarrollado metodologías cuantitativas y cualitativas que facilitan su comprensión. En el caso particular de las variables frecuencia sísmica y precipitación, se tuvo acceso a información registrada en dos bases de datos que permitieron su cuantificación simplificada directa mediante el uso de la estadística.

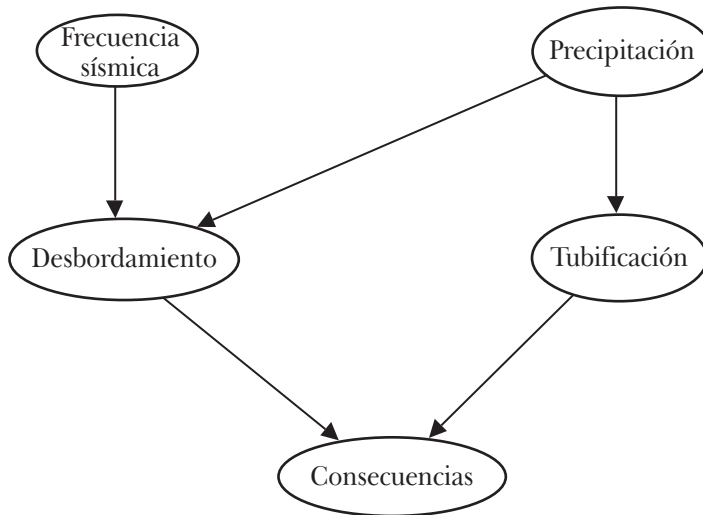
Sin embargo, las variables tubificación y desbordamiento no contaban con información relevante registrada para la zona de interés. En consecuencia, se acudió a cuatro expertos para que contribuyeran con sus conocimientos y experiencia en su cuantificación. Para ello se siguió la metodología del JEE, en la cual se emplean dos puntuaciones: calibración e información.

A través de la ponderación de los expertos, se obtienen los valores óptimos que representan su opinión agregada. Acto seguido, se construyen las distribuciones de probabilidad de los eventos de interés, que se pueden utilizar de forma similar a las generadas mediante datos duros. Como en cualquier otro ejercicio de modelación, es importante subrayar que la calidad de los datos de entrada determina la calidad de los de salida, por lo que se deben explorar diversas fuentes de

información antes de generar las distribuciones ya sea por métodos cuantitativos o por el JEE.

Una vez obtenidas las distribuciones, independientemente del método que les dio origen, se pueden utilizar en la construcción de modelos matemáticos para evaluar los riesgos. En este sentido, los autores han desarrollado uno con base en Redes Bayesianas (RB), que utiliza la información hasta aquí presentada (Delgado *et al.*, 2012 y Morales-Nápoles *et al.*, 2013), y la integra en una herramienta capaz de cuantificar los costos esperados por la eventual falla de una de las siete presas de interés (Delgado *et al.*, 2009), lo cual constituye un modelo que representa el sistema complejo no lineal estudiado (ver Figura VIII.1).

**Figura VIII.1**  
**Red Bayesiana parcial para la evaluación de riesgos en presas de tierra**



Con este tipo de herramientas se pueden priorizar los presupuestos destinados a distintas actividades de mantenimiento, y generar planes de evacuación en casos extremos donde se compruebe que una presa tiene el potencial de fallar. Finalmente, se espera que los resultados aquí reportados sean de utilidad para

soportar futuras investigaciones tanto en materia de seguridad de presas, como en lo referente a la cuantificación de las variables que producen sus fallas vistos desde el punto de vista de sistemas complejos no lineales.

## REFERENCIAS

- Aparicio, F. (1994), *Fundamentos de hidrología de superficies*, Limusa, México.
- Baró, S. J. E., Díaz, D. C., Esteller, A. M. V. y Calderón, G. (2007), "Curvas de daños económicos provocados por inundaciones en zonas habitacionales y agropecuarias de México. Parte I: Propuesta metodológica", *Ingeniería Hidráulica en México*, vol. 22, núm. 1, pp. 91-102.
- BS 8800 (1996), *Guide to Occupational Health and Safety Management Systems*, British Standard Institutions (BSI), Reino Unido.
- Canto Contreras, J. T. y Alamilla López, J. L. (2000), "Análisis de confiabilidad de estructuras de varios grados de libertad", *Memorias del XII Congreso Nacional de Ingeniería Estructural*, 1-4 de noviembre, León, Guanajuato.
- Conagua (1999), *La importancia de la seguridad de presas. Manual para capacitación en seguridad de presas*, Comisión Nacional del Agua, México.
- Cooke, R. M. (1991), *Experts in Uncertainty: Opinion and Subjective Probability in Science*, Oxford University Press, EU.
- Cooke, R. M. y Gossens, L. H. J. (2008), "TU Delft Expert Judgment Data Base", *Reliability Engineering & System Safety*, vol. 93 núm. 5, pp. 657-674.
- Cornell, A. (1968), "Engineering Seismic Risk Analysis", *Bulletin of the Seismological Society of America*, 58, 1583-1606.
- Delgado, H. D., Rivero, S. J., Pérez, P. B., Pérez, F. D., Morales, N. O. y De León, E. D. (2008), "Evaluación de daños provocados por la falla de presas de tierra: un caso en el Estado de México", *XVI Congreso Nacional de Ingeniería Estructural*, Sociedad Mexicana de Ingeniería Estructural, Veracruz, Veracruz, México, nov. 6-8.
- Delgado, H. D., Pérez, P. B., Morales, N. O., Pérez, F. D., Rivero, S. J. y De León, E. D. (2009), "Falla de presas de tierra por deslizamiento de taludes y formación de brechas originadas por sismo: un caso en el Estado de México", *XVII Congreso Nacional de Ingeniería Sísmica*, Sociedad Mexicana de Ingeniería Sísmica, Puebla, Puebla, México, nov. 11-14.
- Delgado Hernández, D. J., Morales-Nápoles, O., De León Escobedo, D. y Arteaga Arcos, J. C. (2012), "A Continuous Bayesian Network for Earth Dams' Risk Assessment: An Application", *Structure and Infrastructure Engineering*. In press. DOI:10.1080/15732479.2012.731416.

- Ellinwood, B. R. (2001), "Earthquake Risk Assessment of Building Structures", *Reliability Engineering and Systems Safety*, 74, 251-262.
- Esteva, L. y Ruiz, S. (1989), "Seismic Failure Rates of Multistory Frames", *Journal of Structural Engineering, ASCE*, vol. 115, núm. 2, pp. 268-284.
- Faber, M. H. (2009), *Risk and Safety in Engineering, Lecture Notes*. Swiss Federal Institute of Technology, Zurich. Recuperado de <[http://www.ibk.ethz.ch/emeritus/fa/education/ws\\_safety/Safety09/Non\\_printable\\_script.pdf](http://www.ibk.ethz.ch/emeritus/fa/education/ws_safety/Safety09/Non_printable_script.pdf)>.
- Foster, M., Fell, R. y Spannagle, M. (2000), "The Statistics of Embankment Dam Failures and Accidents", *Canadian Geotechnical Journal*, vol. 37, núm. 5, pp. 1000-1024.
- ICOLD (1995), "Dam Failure Statistical Analysis", International Commission on Large Dams (ICOLD), Bulletin 99.
- ICOLD (2008), "International Commission of Large Dams". Recuperado de <<http://www.icoldcigb.net/>>, (consultado en mayo 19 de 2008).
- Juárez, B. E. y Rico, R. A. (1996), *Mecánica de suelos. Tomo II. Teoría y aplicaciones de la mecánica de suelos*, Limusa, México.
- Marengo, H. (1994), "Análisis de riesgo de falla en presas de tierra y enrocamiento ante avenidas", tesis doctoral, Facultad de Ingeniería de la UNAM, México.
- Marengo, H. (2002), "Consideraciones de riesgo de falla en presas mexicanas", Seminario de Seguridad de Presas, Comisión Nacional del Agua, México, D.F., septiembre 5 y 6.
- Marsal, J. (1983), *Presas de tierra y enrocamiento*, Limusa, México.
- McGuire, R. K. (2008), "Probabilistic Seismic Hazard Analysis: Early History", *Earthquake Engineering and Structural Dynamics*, 37, 329-338.
- Merritt, F. S., Kent Loftin, M. y Ricketts, J. T. (1999), *Manual del Ingeniero Civil*, McGraw-Hill, México.
- Mileti, D. (1999), *Disasters by Design: A Reassessment of Natural Hazards in the United States*. Joseph Henry Press. ISBN: 0-309-51849-0.
- Morales-Nápoles, O. y Cooke, R. (2009), "Introducción al modelo clásico de juicio estructurado de expertos: breve recuento del pasado y una aplicación reciente", *Ciencia Ergo Sum*, vol. 16, núm. 3, pp. 309-318.
- Morales-Nápoles, O., Delgado-Hernández, D. J., De León Escobedo, D. y Arteaga-Arcos, J.C. (2013), "Continuous Bayesian Network for Earth Dams' Risk Assessment: Methodology & Quantification", *Structure and Infrastructure Engineering*. In press. DOI: 10.1080/15732479.2012.757789.
- Novak, P., Moffat, A. I. B., Nalluri, C. y Narayan, R. (2001), *Estructuras hidráulicas*, Segunda Edición, McGraw-Hill Interamericana, Bogotá.
- Ordaz, M. (2004), "Some Integrals Useful in Probabilistic Seismic Hazard Analysis", *Bulletin of the Seismological Society of America*, 94 (4), 1510-1516.

- Ramírez, L. (2001), “Modelado de la respuesta sísmica de presas de tierra y enrocamiento”, tesis de licenciatura en Ingeniería Civil, Facultad de Ingeniería, Universidad Nacional Autónoma de México, México.
- Renn, O. (1998), “The Role of Risk Perception for Risk Management”, *Reliability Engineering & System Safety*, vol. 59, núm. 1, pp. 49-62.
- Silva-González, F. L. y Heredia-Zavoni, E. (2007), “Confiabilidad sísmica de plataformas marinas con daño por fatiga para la elaboración de planes de inspección basados en riesgo”, *Revista de Ingeniería Sísmica*, núm. 77, pp. 1-21.
- SRH (1976), “Presas construidas en México”, Secretaría de Recursos Hidráulicos, México.
- Straub, D. (2010), *Engineering Risk Analysis. Lecture Notes*, Technological University of Munich.
- Waldrop, M. (1993), *Complexity: The Emerging Science at the Edge of Order and Chaos*, Simon & Schuster, Nueva York.

# ESTUDIO DE LOS ORBITALES DEL ELECTRÓN EN UNA TAPA SEMIESFÉRICA DE UN NANOTUBO DE CARBONO USANDO FUNCIONALES DE LA DENSIDAD

*Patricia Rosendo Andrés*  
*Porfirio Rosendo Francisco\**

## Resumen

En la actualidad, la nanociencia es un área de la investigación básica cuyas aplicaciones están en la medicina, la astronomía, los circuitos electrónicos y los semiconductores, por mencionar algunas. A finales de los ochenta el estudio se profundizó a los nanotubos de carbono principalmente sobre parámetros energéticos, geométricos y estructurales (Gulseren *et al.*, 2002), estudiados en el cuerpo del cilindro sin tomar en cuenta las tapas semiesféricas, debido a su gran capacidad reactiva; sin embargo, aún no se conocen en su totalidad las características físicas y químicas que presentan los materiales nanoestructurados.

En este trabajo se analizan y estudian las principales características de la tapa semiesférica para nanotubos zig-zag, sobre parámetros energéticos y geométricos, lo que permite entender su comportamiento y llegar a un control coherente. El estudio se hace con la herramienta de Funcionales de la Densidad (para la parte energética), se usa la Teoría de Hückel en el análisis de los orbitales Moleculares (OM) mas relevantes y el cálculo de las integrales Alfa y Beta (de Coulomb y Enlace, respectivamente). Con base en la Teoría  $\pi$ -electrónica semiempírica se considera que los OM se pueden ver como una Combinación Lineal de Orbitales Atómicos (CLOA) y se obtienen sus magnitudes, se analizan y discuten las principales características, entre ellas destaca la forma geométrica, pues en la tapa semiesférica el orbital  $\pi$  es el que más contribuye y proporciona estabilidad, simetría y polarizabilidad.

**Palabras clave:** Nanotubos de carbono, orbitales moleculares, funcionales de densidad, simetría, densidad electrónica.

\* Ambos autores pertenecen a la Facultad de Ciencias, Universidad Autónoma del Estado de México.

## INTRODUCCIÓN

En los últimos años ha habido un gran avance en el diseño y fabricación de nanotubos con diámetros del orden de nanómetros. En particular, se han hecho estudios en diversos tipos de materiales; por ejemplo, en semiconductores como arseniuro de galio, zinc, aluminio y cobalto, entre otros. Un caso especial es el de la molécula  $C_{60}$  que fue descubierta en 1985 por los norteamericanos Robert Curl y Richard Smalley de la Universidad Rice, en Estados Unidos, y por el inglés Harold Kroto de la Universidad de Sussex, en Gran Bretaña. Por este descubrimiento fueron galardonados con el premio Nobel de Química en 1996.

Pero fue a finales de los ochenta cuando el desarrollo del estudio se empezó a profundizar con respecto a los nanotubos y es entonces cuando se empieza a hablar de la curvatura en la estructura atómica del cuerpo del cilindro del nanotubo, considerando incluso torsión en el cuerpo, y se observa la contribución de la hibridación de los orbitales  $\sigma$ - $\pi$  principalmente (Kleiner y Eggert, 2001), incluso mediante simulación se observan y calculan los efectos de estiramiento debidos a la reactividad química en este tipo de estructuras y se discuten las condiciones para aumentar su reactividad local mediante deformaciones controladas (Rochefort *et al.*, 1999), además se estudia la envolvente de energía, estructuras de bandas y densidad de estados, estructura electrónica en nanotubos de carbono con pequeños diámetros por mecanismos *ab-initio* y método de orbitales; con el fin de tratar de explicar algunas de sus propiedades e incluso para encontrar nuevos materiales.

Los nanotubos de carbono son reconocidos como la última fibra de carbón sintetizada, con gran resistencia y enorme conductividad térmica debida a la simetría que presentan, mostrando además procesos de emisión de campo (Srivastava *et al.*, 1999). Transportan con alta eficiencia la corriente eléctrica, además de que no disipan calor. También pueden funcionar como un semiconductor activo en nanoescala, debido a su topología única controlada por propiedades electrónicas que permiten al índice de quiralidad dar la forma específica en la configuración atómica y en el arreglo estructural.

Estudios recientes muestran que el análisis de los nanotubos se lleva cabo en el cuerpo del cilindro, los nanotubos de una sola pared han sido caracterizados por espectroscopia Raman (Terrones *et al.*, 2000). Acorde con la gran información existente en la literatura, se ha demostrado que los extremos del nanotubo, es decir, las tapas, son mucho más reactivos que las paredes (Li *et al.*, 2003). Los nanotubos de una y varias paredes han sido caracterizados con diferentes técnicas experimentales; por ejemplo, espectroscopia en el infrarrojo y análisis termogravimétrico (Li *et al.*, 2003), entre otras.

En algunos estudios teóricos se han considerado los efectos de interacción Coulombiana sobre nanotubos de una sola pared y se ha determinado que cerca de la banda prohibida el nanotubo tiene una resistividad intrínseca que permite el proceso de dispersión a cierta temperatura con comportamiento lineal. Sin embargo, en la actualidad no se tiene información sobre el estudio teórico del efecto de la curvatura en una tapa semiesférica de un nanotubo de carbono y, en gran medida, esto se debe a que es un problema no lineal y complejo de resolver. Usando la herramienta de funcionales de la densidad (DFT) y considerando la curvatura en las tapas, en este pequeño trabajo se estudian y plantean algunas condiciones para la unión de nanotubos por sus extremos, permitiendo construir un alambre nanométrico sin rupturas en los extremos o deformaciones, para ello se hace uso de la tapa como centro de atracción y la configuración electrónica de su estructura; esto nos ayuda a poder comprender algunas otras propiedades como conductividad eléctrica, resistencia y flexibilidad, y esto eventualmente nos daría información de cómo poder manipular la tapa del nanotubo permitiendo el paso de información o algún tipo de energía, por ejemplo.

Sabemos que si se logra el control coherente sobre la tapa semiesférica de un nanotubo, se podrá manipular sin mayor problema la estructura desde el punto de vista geométrico (Gulseren *et al.*, 2002), de modo que se puedan unir varios nanotubos por sus extremos hasta formar, al menos teóricamente, alambres largos de diámetro del orden de nanómetros, sin fracturas entre las uniones, obteniendo con esto alambres con varias ventajas, entre ellas alta resistencia y flexibilidad, además de las propiedades de semiconductor, metal o superconductor, características que presentan los nanotubos de carbono, debido a la orientación del ángulo de quiralidad.

Muchas de las aplicaciones están encaminadas a la tecnología de transporte y confinamiento de energía. En la medicina se especula de muchas posibles aplicaciones; por ejemplo, en la unión de nervios ópticos que conectan al ojo con el cerebro, cuando los nervios se rompen se presenta pérdida de visión, diversos estudios demuestran que estos nervios tienen una terminación muy regular y, según se sabe, en forma de nanotubo. Si se logra entender el comportamiento de la tapa se obtendría información para el control de los nervios ópticos lo que permitiría la unión de estos con el cerebro logrando un incremento en la visión. Otra aplicación importante es en la parte de nuevos materiales, pues debido al avance en la manipulación de las propiedades, los nanotubos pueden ser utilizados para transportar densidades muy altas de corriente formando materiales superconductores, permitiendo con esto el manejo de grandes cantidades de información, entre otras muchas de sus aplicaciones.

En 1964, Pierre Hohenberg y Walter Kohn demostraron que todo observable de un sistema electrónico puede ser calculado a partir de la densidad electrónica, esto se aplica a sistemas en estado base para obtener la energía molecular y la función de onda. En 1965, Khon y Sham presentaron una forma de aproximar la energía; esta teoría surgió como una forma alternativa de resolver sistemas de muchas partículas; la ventaja que se tiene sobre la mecánica cuántica (MC) es que no es necesario resolver ecuaciones diferenciales de segundo orden asociadas a cada uno de los electrones del sistema, otra ventaja que posee DFT sobre MC es que el primero sólo depende de la densidad del sistema y, por tanto, de suponer la posición  $(x, y, z)$  en MC si se quiere resolver un sistema de  $N$ -átomos, se tendría que resolver un sistema de  $3N$  ecuaciones (cada una depende de la posición) lo que complica ampliamente la parte analítica.

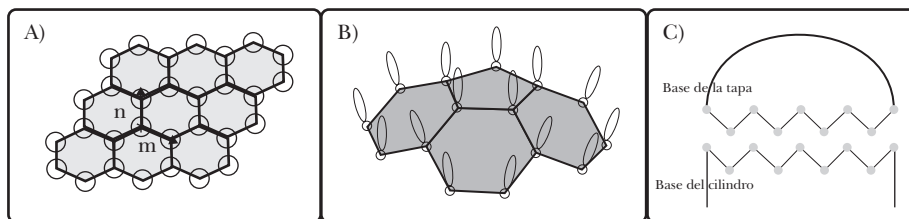
En este trabajo se realiza un análisis teórico de la curvatura de la tapa de nanotubos de carbono considerando el tipo de enlace, arreglo geométrico, y usando la teoría de orbitales moleculares de Hückel se analiza la energía entre vecinos más cercanos, se discuten los primeros resultados en el sistema bajo la influencia de la curvatura y se comparan con los obtenidos en el cuerpo del nanotubo.

## I. NANOTUBOS DE CARBONO

El diamante, el grafito y los fullerenos son formas elementales en las que la naturaleza nos presenta al carbono, conocidos como alótropos. Una forma en la que se presentan dichos alótropos es en la nanoestructura de carbono (del orden de nanómetros) de la cual sobresalen los nanotubos por su amplia gama de propiedades, éstos se pueden ver como láminas de grafito enrolladas sobre sí mismas (su vector de onda es caracterizado por dos enteros positivos  $[n, m]$ ) como se observa en la Figura I.1.

La forma más simple en la que los fullerenos se presentan en la naturaleza es a la manera de un balón de fútbol (domo geodésico o bucky-ball) compuesto por 12 pentágonos y 20 hexágonos; los hexágonos mantienen la planaridad y cada pentágono introduce un cierto ángulo de curvatura, de forma que son necesarios 12 pentágonos para cerrar una superficie sobre sí misma.

Un nanotubo se puede caracterizar conociendo su diámetro, helicidad o índice de quiralidad (el ángulo con el que se enrolla la placa de grafito), esto último le proporciona ciertas particularidades; por ejemplo, puede tener propiedades eléctricas típicas de un metal o semiconductor dependiendo del grado de helicidad que presente o se induzca; sin embargo, por el tamaño de la estructura que presentan, también poseen muchas otras propiedades que pueden ser explicadas



**Figura 1.1.** A) Arreglo estructural de la placa de grafito; B) vista transversal de la tapa semiesférica de un nanotubo, se muestra un pentágono rodeado de hexágonos; C) unión de la base de la tapa y el cuerpo del cilindro

o predichas con ayuda de la mecánica cuántica. El diámetro de los nanotubos es muy pequeño y pueden ser usados como “alambres cuánticos” y transportar densidades muy altas de corriente de varios órdenes de magnitud a diferencia de la transportada por los conocidos alambres de cobre, además de que son materiales muy resistentes y flexibles en comparación con los convencionales.

Cabe mencionar que se conocen hasta el momento dos tipos importantes de nanotubos: “sillón” y “zig-zag”, esto se debe a que en el enrollamiento de las hojas de grafito para localizar al átomo  $(n, m)$  sobre la superficie del tubo, respecto al átomo  $(0, 0)$  se presentan dos posibilidades físicamente aceptables: a) cuando  $(n, m)$  o  $n - m$  es un múltiplo de tres, el nanotubo siempre tendrá un comportamiento metálico y será del tipo sillón; b) para los nanotubos zig-zag, que son de la forma  $(n, 0)$ , por lo general son semiconductores y, por tanto, presentan un *gap* bien definido, es decir, una banda prohibida donde el electrón, físicamente, no tiene permitido moverse ya que su frecuencia de movimiento se atenúa por completo; debido a su característica de semiconductor, este tipo de estructuras son inertes y se pueden dopar con diferentes impurezas, e inclusive al tubo se le puede introducir cierto tipo de elementos. Otra clasificación importante se da a partir de la cantidad de nanotubos concéntricos que poseen: los “monocapa”, o sea un solo nanotubo, y los “multicapa”, con varios tubos concéntricos.

El estudio de los nanotubos cerrados de carbono ha tenido un gran avance en los últimos años, sobre todo en el análisis de los parámetros energéticos y geométricos del cuerpo del cilindro (Gulseren *et al.*, 2002; Rochefort *et al.*, 1999); los efectos de curvatura han sido estudiados analíticamente como una función de la quiralidad; sin embargo, no se han reportado, hasta el momento, resultados de estudios similares en las tapas semiesféricas (mitad de un fullereno) de dichas estructuras, lo cual nos daría información relevante para poder tener mayor control en la síntesis de dichas estructuras.

Uno de los objetivos principales de este capítulo es el estudio teórico de la mitad de un fullereno tomando en cuenta la curvatura característica del arreglo de átomos que constituyen la tapa semiesférica, en función de la cual se obtiene la energía del sistema en estudio. Para esto se planteó el problema, como primera opción, con base en las herramientas teóricas de la mecánica cuántica pero en el camino nos dimos cuenta de que el plantear las ecuaciones que describe dicho sistema implicaba resolver más de 20 ecuaciones diferenciales de segundo orden, esto debido a que cada electrón tiene asociada una función de onda, y como primera aproximación se consideran los átomos que constituyen la celda unitaria; es decir, se deben tomar en cuenta 4 núcleos, cada uno rodeado de seis electrones, lo que convierte al sistema en no lineal y complicado de resolver; es por eso que se consideró un camino alternativo para el tratamiento de dicho problema, sin perder la esencia del objetivo que es encontrar la energía, revisando la literatura y los diferentes métodos de aproximación tales como *ab-initio*, semiempíricos y de la mecánica atómica y molecular, llegamos a la conclusión de que el método de funcionales de la densidad (DFT, por sus siglas en inglés) es el método que nos permite encontrar la energía para el sistema en estudio de una manera más simple, pues calcula la energía y otras propiedades moleculares a partir de la densidad electrónica del sistema en el estado base; se aplica a sistemas pequeños, estudia la parte electrónica de estructuras de átomos y moléculas, y se generaliza a sólidos, una ventaja que tiene es que la posición sólo depende de tres variables ( $x, y, z$ ), y a diferencia de la mecánica cuántica no se conocen de manera exacta las ecuaciones para describir al sistema; por tanto, es un método de aproximación.

Como ya se ha mencionado anteriormente, la parte fundamental de DFT es la densidad electrónica, para esto se debe plantear la densidad propia del sistema; una característica importante que posee el nanotubo de carbono es que los elementos portadores de carga que lo conforman cumplen con el principio de exclusión de Pauli; por tanto, se tiene un conjunto de fermiones, partículas con *spin* semientero, entonces podemos basarnos en la teoría del gas de Fermi de electrones libres para el planteamiento de la densidad considerando la geometría de los puntos reticulares:

$$D(\epsilon) = \frac{mV^{\frac{2}{3}}}{\pi^2 \hbar^2} (3\pi^2 N)^{1/3}$$

donde  $\epsilon$  es la energía de Fermi,  $V$  el volumen de la celda unitaria,  $N$  el número de electrones,  $\hbar$  la constante de Planck y  $m$  la masa.

## II. TEORÍA DE FUNCIONALES DE LA DENSIDAD (DFT)

Para una molécula de  $n$ -electrones de  $3n$  coordenadas espaciales y  $n$  coordenadas de spin se acostumbra definir al Hamiltoniano como:

$$R = -\frac{1}{2} \sum_i V_i^2 - \sum_i \sum_{\alpha} \frac{Z\alpha}{r_{i\alpha}} + \sum_i \sum_{j>i} \frac{1}{r_{ij}}$$

donde  $V_i$  es la energía potencial.

Puesto que el operador Hamiltoniano contiene solamente términos espaciales de uno y dos electrones, la energía molecular se puede escribir en términos de integrales que implican únicamente seis coordenadas espaciales. En este sentido, la función de onda de una molécula poli-electrónica contiene más información de la que es necesaria para describir al sistema físico y lo difícil de este tipo de problemas es interpretar correctamente el significado físico de manera directa. Esto ha obligado a la búsqueda de funciones que impliquen menos variables que la función de onda y que se puedan utilizar para calcular la energía y otras propiedades.

En 1964, Pierre Hohenberg y Walter Kohn probaron (Teorema de Hohenberg-Kohn) que para moléculas con estado fundamental no degenerado (la energía molecular del estado fundamental), la función de onda y todas las demás propiedades electrónicas están determinadas unívocamente por la densidad de probabilidad electrónica del estado fundamental  $\rho_0(x, y, z)$ , una función de solamente tres variables (el subíndice cero indica el estado fundamental). La teoría del funcional de la densidad (DFT) intenta calcular  $E_0$  y otras propiedades moleculares del estado fundamental a partir de su densidad electrónica,  $\rho_0$ , ya que  $E_0 = E_0[\rho_0]$ .

No olvidemos que la función de onda puramente electrónica del estado fundamental,  $\psi_0$ , de una molécula de  $n$ -electrones es una función propia del Hamiltoniano electrónico, y que en unidades atómicas, este se reescribe como:

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^n V_i^2 + \sum_{i=1}^n V(r_i) \sum_j \sum_{i>j} \frac{1}{r_{ij}}$$

La cantidad  $V(r_i)$ , es la energía potencial de la interacción entre el  $i$ -ésimo electrón y los núcleos, esto significa que depende de las coordenadas ( $X_i, Y_i, Z_i$ ) del electrón y de las coordenadas nucleares. Debido a que la ecuación de Schrödinger electrónica se resuelve para posiciones fijas de los núcleos, las coordenadas nucleares no son variables; así,  $V(r_i)$  en la ecuación de Schrödinger es una función sólo de ( $X_i, Y_i, Z_i$ ). En DFT,  $V(r_i)$  se denomina *potencial externo*, que

actúa sobre el  $i$ -ésimo electrón, ya que está producido por las cargas externas al sistema de electrones.

Una vez que se ha especificado el potencial externo,  $V(r_i)$ , y el número de electrones, se determinan las funciones de ondas electrónicas y las energías permitidas de la molécula como soluciones de la ecuación de Schrödinger electrónica.

De acuerdo con Hohenberg y Kohn, la función de onda del estado fundamental y la energía (todas las funciones de onda y energías de los estados excitados) están determinados por la densidad electrónica del estado fundamental.

$$E = \bar{T} + \bar{V}_{Ne} + \bar{V}_{ee}$$

Cada uno de los valores promedio en esta ecuación es una propiedad molecular determinada por la función de onda electrónica del estado fundamental, y a su vez, está determinada por  $\rho_0(r)$ . Por tanto, cada uno de estos promedios es un funcional de  $\rho_0$ :

$$E = E_v[\rho_0] = \bar{T}[\rho_0] + \bar{V}_{Ne}[\rho_0] + \bar{V}_{ee}[\rho_0]$$

donde

$$\bar{V}_{Ne}[\rho_0] = \sum_{i=1}^n V(r_i)$$
$$v(r_i) = - \sum_{\alpha} \frac{Z_{\alpha}}{r_{i\alpha}}$$

Lo que implica que  $V_{Ne}[\rho_0]$  se conoce, pero los funcionales  $T[\rho_0]$  y  $V_{ee}[\rho_0]$  son desconocidos.

Si suponemos que el funcional  $F[\rho_0]$  queda definido como:

$$F[\rho_0] = \bar{T}[\rho_0] + \bar{V}_{ee}[\rho_0]$$

entonces, éste es independiente del potencial externo.

## II.1 Método de combinación lineal de orbitales atómicos (LCAO)

Este método se utiliza para obtener orbitales moleculares a partir de los orbitales atómicos. Esto es, para cualquier molécula diatómica el método da funciones de onda de la forma:

$$\psi = C_1\chi_1 + C_2\chi_2$$

$\chi_1$  y  $\chi_2$  son orbitales atómicos de funciones de onda para los átomos 1 y 2, y  $C_1$  y  $C_2$  son constantes de reflexión que constituyen los orbitales moleculares. Para una molécula diatómica homonuclear se tiene:

$$\psi = N (\chi_1 + \chi_2)$$

La región en la cual  $\psi = 0$  y la densidad electrónica es cero es llamada plano nodal, y el orbital molecular es anti-unido (es decir, no traslapado). Hay dos reglas importantes que deben cumplirse en el método:

- 1) El orbital molecular al combinarse debe tener la misma simetría con respecto al eje internuclear.
- 2) Los orbitales atómicos al combinarse deben tener la misma energía.

La ecuación de Schrödinger con LCAO son funciones de onda del tipo dado en la ecuación  $\psi$  con la cual se obtiene la energía,  $E_{\pm}$  de la forma:

$$E_{\pm} = (E_A \pm \beta)/(1 \pm S)$$

donde  $E_A$  es la energía asociada con el orbital atómico sobre al átomo  $A$ , la cantidad  $\beta$  es la energía de resonancia y es una cantidad negativa. Los orbitales moleculares son separados en energía por aproximadamente  $2\beta$ . La magnitud de la energía de resonancia es un indicador de la diferencia entre el orbital atómico y el orbital molecular que se forma de ellos. La cantidad  $S$  es la integral de traslape, esto es, la medida de hasta qué punto los dos orbitales se traslapan.

## II.2 Teoría de orbitales moleculares (OM)

Las principales teorías que se han utilizado para estudiar los OM son: del electrón libre, la de Pariser-Parr-Pople y la de Hückel (Levine, 2001); particularmente para nuestro problema se utilizará la de Hückel, debido a que la primera sólo nos da

información de la interacción que existe entre electrones libres y no entre átomos vecinos dentro de la red, y la segunda se basa en la interacción entre espines y orbitales, la cual todavía no será tratada en este trabajo.

Las dificultades para aplicar teorías como *ab-initio* a moléculas grandes y medianas provocó la utilización de métodos semi-empíricos tales como los que se utilizan para moléculas conjugadas donde predominan los orbitales  $\Pi$ .

Los orbitales moleculares (OM) canónicos de una molécula se dividen en dos, los  $OM-\sigma$  y  $OM-\pi$ , la única diferencia entre ellos es el valor propio de la reflexión sobre el plano molecular, lo que implica diferente magnitud y forma. Coulson estableció la justificación para la separabilidad de los orbitales canónicos (Levine, 2001), esto se debe a que los orbitales  $\sigma$  poseen mayor simetría que los orbitales  $\pi$ , debido a que éstos poseen mayor polarizabilidad que aquéllos, esto permite la susceptibilidad a perturbaciones tales como las que ocurren en las reacciones químicas.

Orbital Molecular de Hückel (HMO por sus siglas en inglés): Esta teoría electrónica semi-empírica apareció en 1930, según la cual el Hamiltoniano se aproxima en una forma simple:

$$\hat{H} = \sum_{i=1}^n \hat{H}(i)$$

donde  $H_i$  incorpora los efectos de las repulsiones electrónicas de una forma promedio.

El tratamiento por HMO no especifica una forma explícita del Hamiltoniano por lo que éste se debe aproximar. Puesto que el Hamiltoniano electrónico es la suma de los Hamiltonianos unielectrónicos, es posible realizar una separación de variables como en OM del electrón libre. Los OM de Hückel satisfacen la siguiente relación:

$$\hat{H}(i)\phi_i = e_i\phi_i$$

La ecuación anterior no se puede resolver directamente ya que se complica considerablemente, por lo que se acostumbra usar el método variacional. El método de HMO aproxima los OM como una combinación lineal de orbitales atómicos (CLOA). Por ejemplo, en un cálculo con base mínima de un hidrocarburo conjugado plano, los únicos OA de simetría  $\Pi$  son los orbitales  $2p\Pi$  del carbono, donde por  $2p\Pi$  entendemos los OA  $2p$  reales que se encuentran perpendiculares al plano molecular.

Los valores óptimos de los coeficientes para los  $n_c$  (número de átomos de carbono) de los OM más bajos satisfacen

$$\sum_{S=1}^{nc} [(H_{rs} - S_{rs}e_i)C_{si}] = 0$$

donde los  $e_i$  son las raíces de la ecuación secular

$$\det (H_{rs} - S_{rs}e_i) = 0$$

con

$$H_{rr} = \int f_r^*(i) \hat{H}(i) f_r(i) dv(i) \equiv \alpha$$

$$H_{rs} = \int f_r^*(i) \hat{H}(i) f_s(i) dv(i) \equiv \beta$$

$$H_{rs} = 0$$

la integral  $H_{rr}$  tiene el mismo valor para cualquier átomo de carbono de la molécula; sin embargo,  $H_{rs}$  tiene el mismo valor para dos átomos de carbono cualesquiera ligados entre sí; y se anula para dos átomos de carbono no enlazados entre sí. La integral  $S_{rr}$  es igual a 1, ya que los orbitales atómicos están normalizados. La integral de solapamiento  $S_{rs}$  se supone que se anula para  $r \neq s$ .

Cuando  $C_r$  y  $C_s$  no están enlazados entre sí:

$$S_{rs} = \int f_r^*(i) f_s(i) dv(i) = \delta_{rs}$$

$$f_r = Cr2p\pi$$

$\alpha$  = integral de Coulomb,

$\beta$  = integral de enlace o resonancia.

Cuando los carbonos no se encuentran enlazados, se considera entonces que están lo suficientemente separados en el espacio, de tal forma que es razonable suponer  $H_{rs}=0$ .

Para nuestro caso particular necesitamos encontrar el valor de las integrales  $\alpha$  y  $\beta$ ; sin embargo, puesto que nuestro sistema es no lineal, dichas soluciones son complicadas de resolver. Para ello se partirá de funciones de onda simples y conocidas para formar la combinación lineal adecuada; por ejemplo, en algunos casos es posible considerar a  $\beta$  como cero en función de la separación entre elementos de la red, esto facilita considerablemente el cálculo de los orbitales de Slater. Específicamente, de acuerdo con las características físicas, es factible con-

siderar a las funciones de onda de átomos hidrogenoides como punto de partida para nuestro sistema:

$$1s = \frac{1}{\pi^{1/2}} \left( \frac{z}{a} \right)^{3/2} e^{-Zr/a}$$
$$2s = \frac{1}{4(2\pi)^{1/2}} \left( \frac{z}{a} \right)^{3/2} \left( 2 - \frac{Zr}{a} \right) e^{-Zr/a}$$

lo que implicaría que la primera forma de  $\phi_1$  sería la siguiente:

$$\phi_1 = C_1 1s + C_2 2s$$

Además, observemos que una base mínima para la construcción de orbitales atómicos se puede construir a partir de la combinación lineal de los elementos de la primera forma de  $\phi_1$  con  $2px$ ,  $2py$ ,  $2pz$ , esto para cada átomo de carbono de la red, donde:

$$2px = \frac{1}{4(2\pi)^{1/2}} \left( \frac{z}{a} \right)^{5/2} r e^{-Zr/2a} \sin\theta \cos\phi$$
$$2py = \frac{1}{4(2\pi)^{1/2}} \left( \frac{z}{a} \right)^{5/2} r e^{-Zr/2a} \sin\theta \sin\phi$$
$$2pz = \frac{1}{4(2\pi)^{1/2}} \left( \frac{z}{a} \right)^{5/2} r e^{-Zr/2a} \cos\theta$$

por tanto, una base mínima está compuesta de:

$$\phi_1 = C_1 1s + C_2 2s + C_3 2px + C_4 2py + C_5 2pz$$

### III. RESULTADOS

Considerando a un nanotubo zig-zag, monocapa y semiconductor, se toma en cuenta la mitad del casquete semiesférico y, debido a la geometría que presenta, se considera a un pentágono rodeado de hexágonos en la parte más alta; además de suponer a los átomos unidos mediante resortes ideales, se considera a uno de los puntos reticulares de la red como centro de donde parten los vectores base que definen la celda unitaria. Es importante aclarar que para fines prácticos de

nuestro problema y por la geometría característica de la tapa semiesférica, los vectores base están localizados en diferentes planos debido a la curvatura del sistema. Se supone también que el sistema físico es idealizado como un sistema que no está bajo la influencia de fuerzas externas que interactúen sobre él. Otra consideración digna de mencionarse es que el sistema es dinámico localmente y no hay rotación alguna.

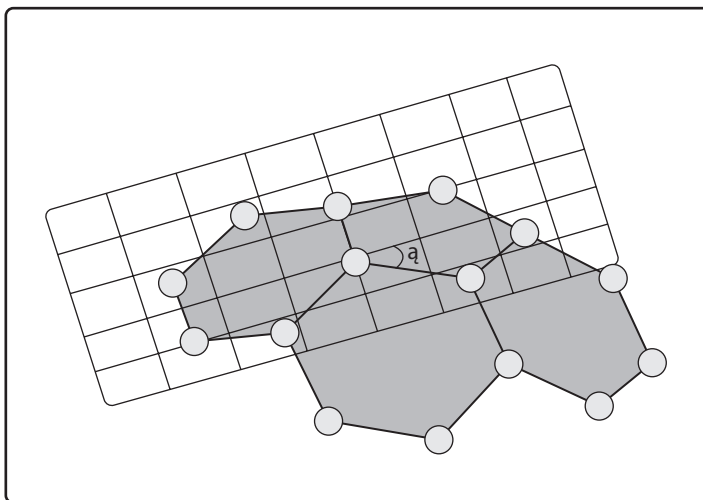
En la Figura III.1 se observa el arreglo estructural que tiene la tapa semiesférica del nanotubo y los vectores que constituyen la celda unitaria.



**Figura III.1.** Arreglo estructural de la tapa del nanotubo, se observa un pentágono rodeado de hexágonos y los vectores de la celda unitaria.

De acuerdo con esta figura, si se considera un diagrama de cuerpo libre donde se observa un punto reticular (en DFT es llamado comúnmente núcleo) rodeado de sus vecinos más cercanos, observaríamos un punto reticular, que constituye el pentágono, como centro; los vecinos más cercanos a este núcleo son tres, la magnitud de la distancia interatómica que los separa es de 1.42 Å, los vectores linealmente independientes se consideran a lo largo de la línea que une cada vecino cercano con el punto reticular del centro.

En la Figura III.2 se observa la tapa semiesférica de un nanotubo de carbono con un plano de referencia perpendicular a su eje, esta consideración es importante porque de aquí se parte para poder tener idea de la curvatura del sistema. En cada punto reticular se encuentra un átomo de carbono, dando forma a la celda unitaria.



**Figura III.2.** Se muestra un plano que pasa tangente a uno de los puntos reticulares del pentágono como punto de referencia a la curvatura.

Para nuestro problema en particular partimos de la configuración electrónica del carbono, usando la configuración de Möller:

$$C^6 = 1s^2 2s^2 2p^2$$

y tomando en cuenta el principio de exclusión de Pauli, el cual nos dice que no hay dos partículas con los mismos números cuánticos, consideramos la simetría del sistema. Este aspecto es de suma importancia ya que, como bien sabemos, de aquí se puede obtener información cualitativa sobre la función de onda y diferentes propiedades moleculares. Con base en esto y de acuerdo con nuestro objeto de estudio se tiene:

- a) Eje de simetría de orden  $n$ ; si existe una rotación  $360/n$ , donde  $n$  es un elemento de los enteros a dicho eje y se da una configuración físicamente indistinguible de la posición original. Nuestro sistema posee un eje de simetría de orden  $n=1$ , es decir, la rotación debe ser sobre  $360^\circ$ .

- b) Plano de simetría: existe si la reflexión de todos los núcleos respecto a ese plano da una configuración físicamente indistinguible a la original. El sistema no tiene plano de simetría ya que la mitad del cristal no se ve reflejada en un plano que pase por un punto de la red.
- c) Centro de simetría: existe si la operación de invertir todos los núcleos con respecto al centro da una configuración indistinguible de la original. El sistema no posee un centro de simetría porque si se invierten los núcleos no da una configuración igual.
- d) Eje alternante de simetría de orden  $n$ : si la rotación de orden  $n$  seguida de una reflexión en el plano normal al mismo lleva al cuerpo a una posición indistinguible, el sistema no posee este cuarto elemento de simetría debido a que no cumple con la reflexión.

En vista de que no cumple con todas las condiciones de simetría, el sistema, aparte de que se complica para su estudio, deja de ser lineal y entonces no puede ser estudiado bajo las teorías convencionales de la física lineal y debe ser tratado como uno no lineal.

Además de las propiedades de simetría es necesario calcular el volumen de la celda unitaria. Lo cual no es fácil, ya que por efecto de la curvatura se complica un poco; para poder hacerlo observemos al sistema desde arriba y notemos cómo se comportan los vectores unitarios con base en los ejes de referencia, esto es, el eje Z estaría hacia afuera de la página, como se observa en la figura siguiente (Figura III.3):

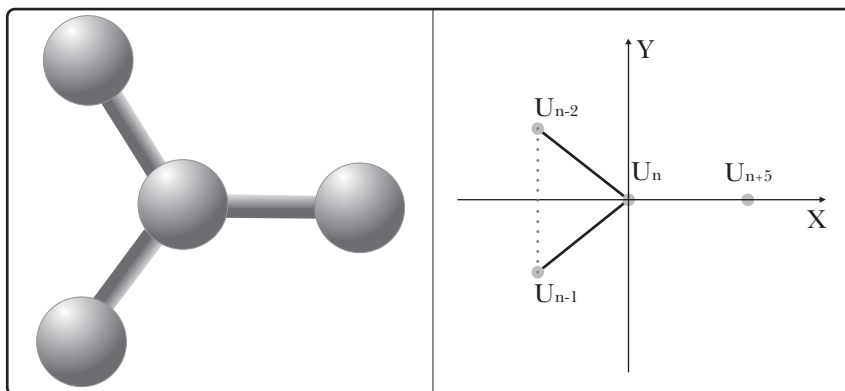


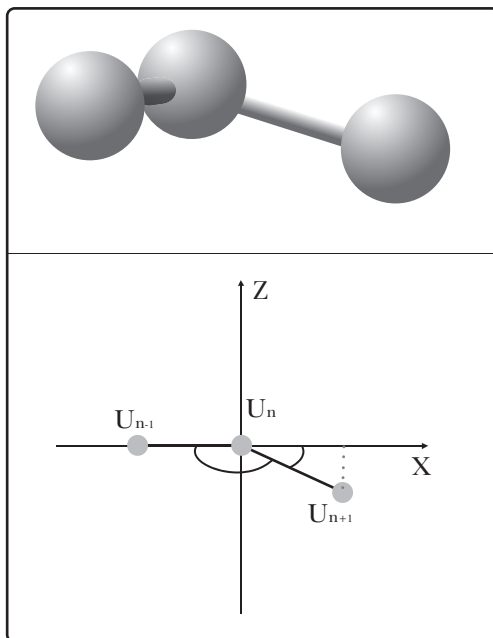
Figura III.3. Vista superior de los átomos que constituyen la celda unitaria.

Considerando ahora una vista transversal del sistema, como se observa en la Figura III.4, se obtienen los siguientes vectores:

$$\bar{a} = (\chi_1, 0, Z_1)$$

$$\bar{b} = (\chi_2, y_2, 0)$$

$$\bar{c} = (\chi_3, y_3, 0)$$

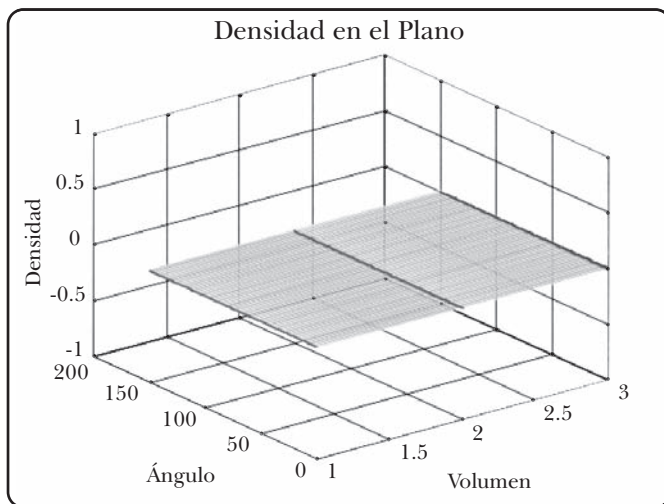


**Figura III.4.** Vista transversal de los átomos que constituyen la celda unitaria. El átomo  $U_{n-2}$  estaría justo detrás de  $U_{n-1}$ . El plano XY es el plano de referencia.

Como se observa en la última figura, la dirección en la que el elemento  $U_{n+1}$  se puede mover es en la que varía el ángulo  $a$  respecto del plano de referencia XY, es importante mencionar que por el hecho de que varíe  $a$  no necesariamente

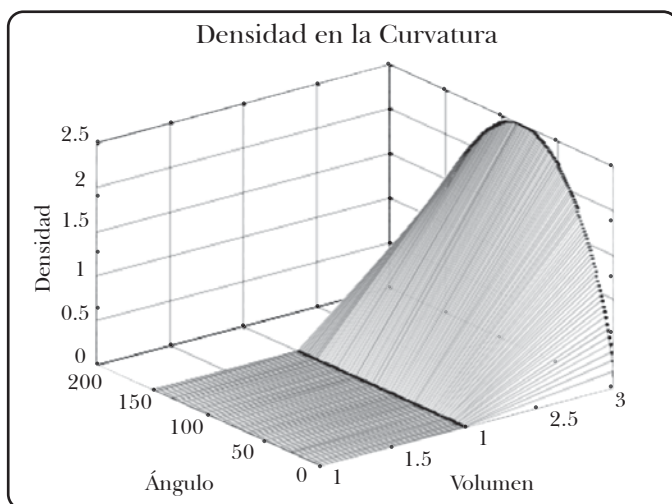
$b$  debe cambiar. Una vez visualizada la posible dirección de movimiento de los elementos de la celda unitaria y en consecuencia la de los vectores unitarios, se calcula el volumen de la celda, el cual presenta su valor máximo cuando  $\alpha=44^\circ$ ; físicamente lo que nos dice este punto máximo es que el vector  $\bar{a}$ , tiene permitido cambiar de dirección desde  $0^\circ$  hasta  $44^\circ$ , si se sobrepasa este límite el sistema se vuelve totalmente inestable y tiende a colapsar. Esto significa que si la tapa del nanotubo está, por ejemplo, bajo la presencia de una presión, podrá resistir una magnitud tal que desplace de la posición de equilibrio a los elementos de la red hasta una posición dada por el valor máximo posible de  $a$  sin que la estructura de la tapa se deforme permanentemente.

No olvidemos que la densidad electrónica de todo sistema físico está en función del volumen, y de acuerdo con lo anterior ya estamos en posibilidades de poder definir el volumen que nos delimitan los vectores unitarios. Para poder entender de forma más simple cuál es el comportamiento de la densidad electrónica en la tapa del nanotubo es indispensable analizar el caso más simple; esto es, cuando estamos en el plano; esto significa que  $\alpha = 0$ ; en otras palabras, el volumen que delimitan los vectores de la celda unitaria es igual con cero, por tanto, la densidad electrónica en el plano  $XY$  es constante y genera un plano como se observa en la Figura III.5.



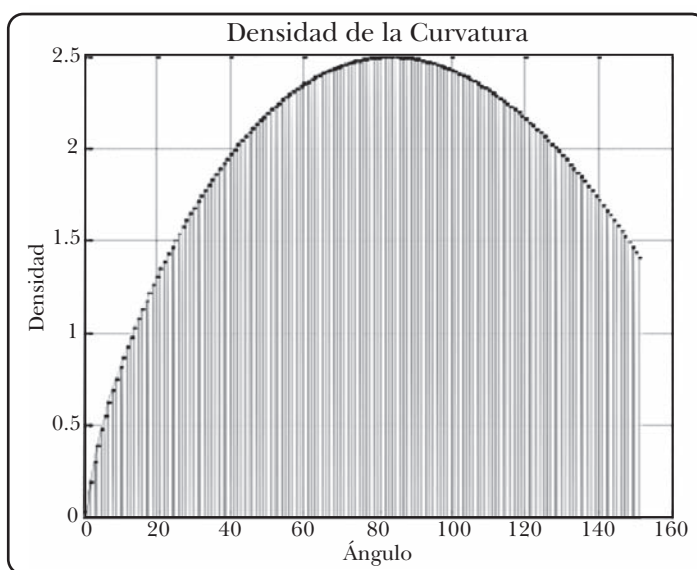
**Figura III.5.** Comportamiento de la densidad electrónica en el plano de referencia  $XY$ .

A diferencia del caso sencillo en el plano, en la curvatura de la tapa del nanotubo (que es el sistema en estudio) se genera un volumen producido por los vectores de la celda unitaria, dando un volumen máximo en  $\alpha=44^\circ$ , lo que ocasiona que el comportamiento de la densidad electrónica deje de ser constante y nos defina una curva en función del volumen y  $\alpha$ , como se observa en la Figura III.6.



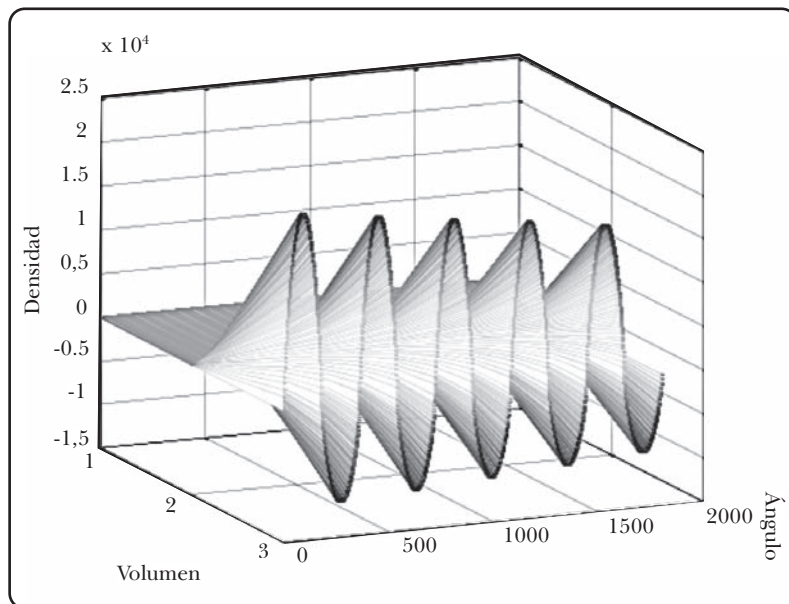
**Figura III.6.** Comportamiento de la densidad electrónica en la curvatura de la tapa semiesférica de un nanotubo de carbono. Alfa se presenta en radianes.

Para poder visualizar mejor el comportamiento de la densidad electrónica se hace una proyección sobre el plano  $Y-Z$  y se tiene una gráfica que nos define un comportamiento de tipo parabólico de la densidad en función de  $\alpha$ , como se observa en la Figura III.7, donde el punto máximo es en  $44^\circ$ , este ángulo es el permitido para la geometría del sistema, es decir, después de este ángulo el sistema colapsaría al centro de la tapa.



**Figura III.7.** Proyección en el plano  $Y-Z$  para la densidad electrónica de la tapa semiesférica de un nanotubo de carbono.

En la siguiente figura (III.8) se muestra cómo evoluciona la densidad según lo describe la teoría; sin embargo, físicamente, como ya se dijo anteriormente, el valor de  $\alpha$  está acotado.



**Figura III.8.** Comportamiento periódico de la densidad electrónica descrito por la teoría.

En la figura anterior se observan los máximos y mínimos de la densidad con comportamiento periódico, se observa claramente el aumento en la densidad para valores positivos de  $\alpha$ .

Puesto que ya contamos con la densidad electrónica, esto nos da la posibilidad de poder calcular la energía del sistema, la cual se encuentra constituida por tres términos que corresponden a la energía cinética electrónica, la energía potencial y la energía de interacción electrón-electrón, descritas matemáticamente por las siguientes ecuaciones:

$$E[n] = T[n] + V[n] + U[n]$$

$$E[n] = \frac{-\hbar^2}{2m} \int d^3n^*(r) \nabla^2 n(r) + \int d^3m(r)v(r) + \frac{1}{2} \int \int \frac{n(r)n(r')}{|r-r'|} drdr'$$

el primer término de la ecuación da la contribución de la energía del electrón alrededor del núcleo, el segundo término corresponde a la parte de la energía que mantiene unidos entre sí a los puntos reticulares y el tercer término proporciona la parte de la energía de interacción de los electrones.

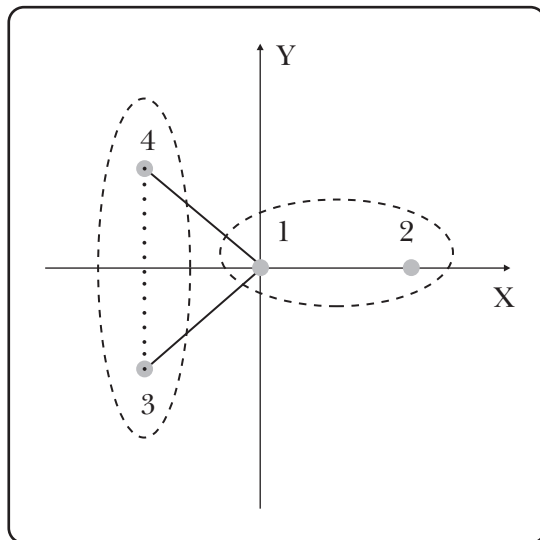
Por otro lado, es importante estudiar y analizar las propiedades de los orbitales moleculares, para ello se van a considerar tres casos importantes, de los cuales sólo uno se tomará en cuenta debido a la contribución física a nuestro problema en particular; el primer caso es considerando la celda unitaria y las interacciones del átomo central con los tres vecinos mas cercanos, pero al establecer la situación de esta manera se repite el átomo del centro dos veces, por lo tanto esa aproximación no es del todo válida aunque se tienen cuatro integrales de Coulomb y tres integrales de Enlace, de ahí que la ecuación secular no es difícil de visualizar, esto es, la matriz característica se presenta de la siguiente forma:

$$\begin{pmatrix} \alpha & \beta & \beta & \beta \\ \beta & \alpha & 0 & 0 \\ \beta & 0 & \alpha & 0 \\ \beta & 0 & 0 & \alpha \end{pmatrix}$$

simplificando, se obtiene el siguiente determinante que define la ecuación secular a resolver:

$$\begin{vmatrix} X & 1 & 1 & 1 \\ 1 & X & 0 & 0 \\ 1 & 0 & X & 0 \\ 1 & 0 & 0 & X \end{vmatrix}$$

Para el segundo caso, se toman en cuenta las interacciones por pares, es decir, el átomo del centro  $U_n$  con el vecino más cercano  $U_{n+1}$ , como se muestra en la figura III.9, y la unión de los átomos  $U_{n-1}$  con el  $U_{n-2}$ , esto para evitar que se duplique el del centro como ocurrió en el caso anterior. Además se asume que no tiene interacción el átomo  $U_{n-1}$  con el  $U_{n+1}$ , tampoco el átomo  $U_{n-2}$  con el  $U_{n+1}$ ; de igual forma, no hay interacción para el átomo del centro con los de la izquierda, es decir, no hay relación entre el  $U_n$  con el  $U_{n-1}$  ni con el  $U_{n-2}$ .



**Figura III.9.** Interacciones por pares de los puntos reticulares.

Después de hacer las consideraciones anteriores se obtiene que varios de los componentes de la matriz característica del sistema tengan entrada cero, esto es:

$$\begin{pmatrix} \alpha & \beta & 0 & 0 \\ \beta & \alpha & 0 & 0 \\ 0 & 0 & \alpha & \beta \\ 0 & 0 & \beta & \alpha \end{pmatrix}$$

Simplificando, se obtiene el determinante que define a la ecuación secular, el cual presenta la siguiente estructura:

$$\begin{vmatrix} X & 1 & 0 & 0 \\ 1 & X & 0 & 0 \\ 0 & 0 & X & 1 \\ 0 & 0 & 1 & X \end{vmatrix}$$

Cuyas soluciones desde un punto de vista general son de la forma:

$$e_k = \alpha - \beta X$$

$$X = \pm 1$$

esto implica que las soluciones serán:

$$e_1 = \alpha + \beta$$

$$e_2 = \alpha + \beta$$

$$e_3 = \alpha - \beta$$

$$e_4 = \alpha - \beta$$

Se comprueba una vez más que los orbitales más relevantes son los de magnitud uno, correspondientes a los orbitales de geometría elíptica denominados con el nombre de  $\pi$ .

En el tercer caso se considera un pentágono y el hexágono vecino mas cercano, es decir, como un anillo con número impar de carbonos, donde la numeración es cíclica; para este sistema se obtiene una matriz de 9x9 y es difícil de tratar el determinante ya que se obtienen nueve integrales de Coulomb y diez integrales de enlace, lo que se observa claramente en la siguiente matriz:

$$\begin{pmatrix} \alpha & \beta & 0 & 0 & 0 & 0 & 0 & \beta & 0 \\ \beta & \alpha & \beta & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \beta & \alpha & 0 & 0 & 0 & 0 & 0 & \beta \\ 0 & 0 & 0 & \alpha & \beta & 0 & 0 & 0 & \beta \\ 0 & 0 & 0 & \beta & \alpha & \beta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta & \alpha & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \alpha & \beta & 0 \\ \beta & 0 & 0 & 0 & 0 & 0 & \beta & \alpha & \beta \\ 0 & 0 & \beta & \beta & 0 & 0 & 0 & \beta & \alpha \end{pmatrix}$$

Mediante métodos algebraicos se puede simplificar el resolver el determinante correspondiente, ya que se obtiene una matriz por bloques, un bloque es de 3x3 y el otro es de 6x6, este último se conoce como determinante circulante, por la forma que lo caracteriza.

$$\begin{pmatrix} \alpha & \beta & 0 \\ \beta & \alpha & \beta \\ 0 & \beta & \alpha \end{pmatrix}$$
$$\begin{pmatrix} \alpha & \beta & 0 & 0 & 0 & \beta \\ \beta & \alpha & \beta & 0 & 0 & 0 \\ 0 & \beta & \alpha & \beta & 0 & 0 \\ 0 & 0 & \beta & \alpha & \beta & 0 \\ 0 & 0 & 0 & \beta & \alpha & \beta \\ \beta & 0 & 0 & 0 & \beta & \alpha \end{pmatrix}$$

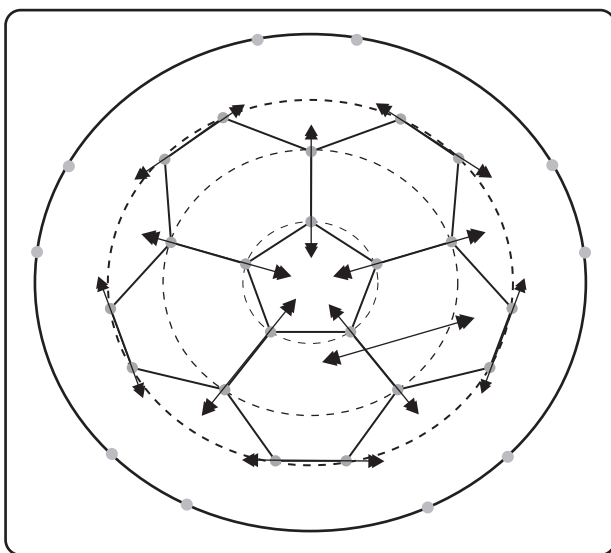
Al resolver la integral de Coulomb, se obtiene que  $\alpha = 0.456$  eV, para cada uno de los átomos que constituyen la celda unitaria. Cuando se generaliza a segundos y terceros vecinos el valor de  $\alpha$  no cambia, se mantiene constante. Para el caso de segundos vecinos la celda unitaria se considera un pseudo-átomo y toma las mismas consideraciones que el átomo original, además se tiene una matriz de 7x7 entradas, para terceros vecinos la matriz es de 9x9.

Al resolver la integral de enlace mediante métodos numéricos considerando a primeros vecinos se tiene que  $\beta = 0.014$  eV, para segundos vecinos  $\beta = 0.0014$  eV y para terceros  $\beta = 0.00014$  eV, lo que nos indica este comportamiento es que cuando se generaliza a segundos vecinos el átomo central siente un apantallamiento por los primeros vecinos, por eso el valor numérico disminuye en un orden de magnitud, para terceros vecinos el apantallamiento es producido por los primeros y segundos vecinos y el orden de magnitud es considerable.

El valor numérico que se obtiene para la integral de Coulomb y para la integral de enlace a primeros vecinos, nos lleva a observar que el potencial coulombiano es mucho mayor que el de interacción entre dos átomos o puntos reticulares.

Debido a que el sistema tiene uniones carbono-carbono y enlaces covalentes, se tiene que los orbitales mas relevantes son los  $\Pi$ , la forma elíptica que poseen mantiene la curvatura en la tapa; además, la orientación de cada orbital es radial a la superficie de la tapa, lo que le da gran estabilidad al sistema. Cuantitativamente, sabemos que la orientación de cada orbital permite una proyección con respecto al plano de referencia, lo que implica un momento angular diferente de cero. Cuando se calcula la energía del orbital se obtiene una energía mayor comparada con la energía del orbital de enlace, por tanto, se tiene que los orbitales generados por dos átomos son enlazantes, los cuales poseen gran estabilidad y simetría con respecto al plano nodal, son perpendiculares al eje de enlace y radiales a la tapa semiesférica del nanotubo, además son de atracción en la región inter-nuclear.

Debido a que los núcleos de una molécula vibran en torno a su posición de equilibrio, es importante incluir la energía vibracional molecular. Respecto al movimiento vibracional de los átomos se tiene que, para moléculas lineales está compuesta por  $3N-5$  modos normales de vibración; para moléculas no lineales se tienen  $3N-6$  modos normales de vibración, donde  $N$  es el número de átomos en la molécula. Lo que significa que si cada átomo tiene tres grados de libertad ( $X, Y, Z$ ) y si el sistema contiene  $N$  átomos, entonces éste tiene  $3N$  grados de libertad; para una molécula no lineal, se tiene movimiento translacional y rotacional en ( $X, Y, Z$ ). El sistema en estudio contiene cuatro átomos de carbono que constituyen la celda unitaria, originando seis modos normales de vibración; pero si generalizamos el sistema y se considera un pentágono rodeado de hexágonos incluidos los pentágonos alternos, se tiene un total de 30 átomos; suponiendo el movimiento hacia el centro de todos los pentágonos, el análisis vibracional tiene un comportamiento de convergencia, esto para el pentágono del centro; los átomos siguientes tienen una divergencia radial, el tercer pseudo-átomo tiene trayectoria tangencial, donde los átomos cercanos se oponen en la trayectoria y los lejanos se atraen; finalmente para el cuarto pseudo-átomo se tiene una convergencia radial, esto se nota al observar la Figura III.10:



**Figura III.10.** Modos normales de vibración físicamente aceptables acorde con la geometría de la tapa semiesférica del nanotubo de carbono.

Dentro de los resultados mas relevantes es importante mencionar que se tienen dos tipos de orbitales: enlazantes (OM-E) y antienlazantes (OM-A), los primeros mantienen mayor simetría, son de baja energía con respecto al orbital del núcleo original, la región que abarcan (entre cada núcleo) es de atracción, además de ser un orbital altamente estable. A diferencia de los OM-A son menos simétricos, tienen menor energía que el núcleo original, la región inter-nuclear es de repulsión y son inestables.

El sistema en estudio revela que los orbitales TT correspondientes a la celda unitaria son enlazantes; por tanto, presentan alta simetría, además tienen baja energía, la región donde se encuentra el orbital enlazante es de atracción inter-nuclear y, finalmente, es más estable que el orbital propio del núcleo.

Cabe mencionar una diferencia entre estable y reactivo. En diferentes referencias se menciona que la tapa del nanotubo es mas reactiva que el cuerpo, por eso es mas fácil unir dos nanotubos por las tapas, pero eso es debido a que se tiene una tapa reactiva, es decir, la tapa (y el cuerpo del nanotubo) está sujeta a cambios químicos, presión, temperatura, cantidad molar, entre otros, por lo que se dice que el sistema es altamente reactivo, a diferencia de un sistema estable el cual está sujeto a un rango de energía bien específico, eso difícilmente se logra controlar; por tanto, se dice que cuando se controla la energía que se aplica al sistema se tiene un sistema bajo un *control coherente*.

Para un estudio completo de la tapa semiesférica del nanotubo, se requiere un análisis sobre el momento angular orbital total y el momento angular de *spin* total. Como primera aproximación se analizan los puntos reticulares individualmente y se obtiene que el momento angular orbital electrónico es de magnitud 1, lo que equivale a tener un orbital P; este análisis se realiza para los cuatro átomos que constituyen la celda unitaria. Siguiendo con el análisis del momento angular para la molécula en estudio, usando el criterio de la teoría de Hückel por pares, se consideran dos pares como en la Figura III.10.

Generalizando, para calcular el momento angular para los segundos vecinos, se considera a los átomos que constituyen la celda unitaria como un pseudo-átomo y se generaliza a terceros vecinos. Se hace un análisis detallado por pares de los átomos y se obtiene que el momento angular es cero, esto debido a que el conjunto de puntos reticulares se encuentra en equilibrio y no están cambiando de posición o rotando.

Cabe mencionar que el trabajo estaría incompleto si no se menciona qué pasa, físicamente hablando, cuando se realiza un análisis en la base de la tapa, es decir, donde se unen la base de la tapa y el cuerpo del cilindro, uno se pregunta: ¿qué

tipo de orbitales hay?, ¿cuál predomina?, ¿cómo es la geometría de la trayectoria de los electrones?, para ello observamos que cuando se estudia un nanotubo semiconductor se tiene una estructura en zig-zag, y para el análisis de un nanotubo metálico la estructura del cuerpo es sillón, lo que implica gran diferencia en la unión de la base de la tapa con el cuerpo del cilindro. Debido a que la tapa del nanotubo posee momento angular se tienen orbitales elípticos, pero el cuerpo del cilindro posee momento angular de magnitud cero, lo que implica orbitales esféricos.

## CONCLUSIONES

Usando la teoría de funcionales de la densidad se obtiene que la densidad electrónica para el caso del plano es cero, y para el sistema con curvatura tiene un comportamiento parabólico, con punto máximo en  $\alpha=44^\circ$ , que es el ángulo máximo en el que el sistema tiene permitido moverse; después de este ángulo el sistema colapsa.

Al calcular el comportamiento de la densidad en función del ángulo se obtiene un comportamiento periódico, lo que garantiza que el sistema sea cristalino y tenga periodicidad, de donde el sistema en estudio revela que los orbitales más relevantes son enlazantes; por tanto, presentan alta simetría con baja energía, la región donde se encuentra el orbital enlazante es de atracción inter-nuclear y, finalmente, es altamente estable.

El sistema en estudio revela que los orbitales más relevantes son enlazantes; por tanto, presentan alta simetría, además tienen baja energía, la región donde se encuentra el orbital enlazante es de atracción inter-nuclear y, finalmente, es más estable que el orbital propio del núcleo.

Después de un estudio semiclásico se observó que los orbitales que más contribuyen en la curvatura de la tapa son los orbitales  $\Pi$ , cuya forma geométrica es una elipse orientada de forma radial, lo que le da mucha estabilidad al sistema en estudio. Además se logró establecer las direcciones preferenciales en los modos de vibración característicos.

Después de resolver la integral de Coulomb y la de enlace, el valor numérico demuestra que el potencial coulombiano es mucho mayor que el de interacción entre dos átomos o puntos reticulares para cualesquiera en la tapa semiesférica del nanotubo de carbono.

Por otro lado, se logró identificar las direcciones de vibración físicamente aceptables para las elementos de la red en la tapa semiesférica.

## REFERENCIAS

- Ding, J. W.; X. H. Yang; J. X. Cao; D. L. Wang; Y. Tang; Q. B. Yang (2003). *Phys. Condens. Matter*, 15, 439.
- Gulseren, O.; T. Yildirim; S. Ciraci (2002). *Phys. Rev.*, B65, 153405.
- Levine, Ira N. (2001). *Química cuántica*, 5a. edición, Prentice Hall.
- Kleiner, A.; S. Eggert (2001). *Phys. Rev.*, B65, 113402.
- Li, W. Z.; C. H. Liang; W. J. Zhou; J. S. Qiu; Z. H. Zhou; G. Q. Sun; Q. J. Xin (2003). *Phys. Chem. B* 107 6292.
- Miyamoto, Y.; A. Rubio; M. L. Cohen; S. G. Louie (1994). *Phys. Rev.*, 50, 7.
- Rochefort, P.; A. Avouris; F. Lesage; D. R. Salahub (1999). *Phys. Rev.*, B60, 13824.
- Srivastava, D.; W. Brenner; J. D. Schall; K. D. Ausman; M. Yu; R. S. Ruoff (1999). *Phys. Chem.*, 103, 4330.
- Terrones, M.; H. Terrones; E. Hernández; N. Grobert; J-C. Charlier; P. M. Ajayan (2000). *Phys. Rev.*, 84, 8.

**Complementarias**

- Baughman, R. H.; A. A. Zakhidov; W. A. de Heer (2002). *Science*, 297, 787.
- He, Z. B.; J. H. Chen; D. Y. Liu; H. Tang; W. Deng; W. F. Kuang (2004). *Mater. Chem. Phys.*, 85, 396.
- Imholt, T. J.; C. A. Dyke; B. Hasslacher; J. M. Perez; D. W. Price; J. A. Roberts; J. B. Scott; A. Wadhawan; Z. Ye; J. M. Tour (2003). *Chem. Mater.*, 15, 3969.
- Kuznetsova, A.; J. T. Jr. Yates; J. Liu; R. E. Smalley (2000). *Chem. Phys.*, 122, 9590.
- Li, W. Z.; C. H. Liang; W. J. Zhou; J. S. Qiu; H. Q. Li; G. Q. Sun; Q. Xin (2004). *Carbon*, 42, 436.
- Liu, Z. L.; X. H. Lin; J. Y. Lee; W. Zhang; M. Han; L. M. Gan (2002). *Langmuir*, 18, 4054.
- Pearce, J. V.; M. A. Adams; O. E. Vilches; M. R. Johnson; H. R. Glyde (2005). *Phys. Rev. Lett.*, 95, 185302.
- Rajesh, B.; K. R. Thampi; J. M. Bonard; X. Xanthopoulos; H. J. Mathieu; B. J. Viswanathan (2003). *Phys. Chem. B.*, 107, 2701.
- Rajesh, B.; K. R. Thampi; J. M. Bonard; H. J. Mathieu; N. Xanthopoulos; B. Viswanathan (2003). *Chem. Commun.* 2022.
- Serp, P.; M. Corrias; P. Kalck (2003). *Appl. Catal. A.*, 253, 337.
- Simonyan, V. V.; J. K. Johnson; A. Kuznetsova; J. T. Yates Jr. (2001). *Chem. Phys.*, 114, 180.

- Stan, G.; M. J. Bojan; S. Curtarolo; M. Gatica y M. W. Cole (2000). *Phys. Rev. B*, 62, 2173.
- Stan, G.; M. W. Cole (1998). *Surf. Sci.*, 395, 280.
- Steigerwalt, E. S.; G. A. Deluga; C. M. Lukehart J. (2002), *Phys. Chem. B*, 106, 760.
- Wang, C. Y.; T. H. Chen; S. C. Chang; T. S. Chin; S. Y. Cheng (2007). *Applied Physics Letters*, 90, 103111.
- Wang, J.; R. P. Deo; P. Poulin; M. J. Mangey (2003). *Am. Chem. Soc.*, 125 14706.
- Zhang, M.; S. L. Fang; A. A. Zakhidov; S. B. Lee; A. E. Aliev; C. D. Williams; K. R. Atkinson y R. H. Baughman (2005). *Science*, 309, 1215.



# ALGORITMOS GENÉTICOS APLICADOS A LA OPTIMIZACIÓN DE CENTRALES DE CICLO COMBINADO Y CALDERAS DE RECUPERACIÓN DE CALOR

*María Dolores Durán\**

## **Resumen**

En el presente trabajo se explica una metodología para la optimización termoeconómica de centrales de ciclo combinado, enfocándose en los parámetros termodinámicos y geométricos de diseño de la caldera de recuperación de calor.

Esta metodología está dividida en 2 módulos: el primero de ellos parte de las condiciones de diseño de la turbina de gas, realiza la simulación de la misma y con estos resultados aplica la herramienta de algoritmos genéticos para obtener los parámetros termodinámicos óptimos de la caldera de recuperación de calor, así como el producto del coeficiente global de transferencia de calor y el área (UA) para cada sección de la misma. Por su parte, el segundo módulo considera los parámetros termodinámicos de diseño obtenidos en el primer módulo y, a partir de ello, aplica la misma herramienta de algoritmo genético (AG) para determinar los parámetros geométricos que den como resultado el valor del producto UA obtenido para cada sección de la caldera.

Para dar una visión más clara de esta metodología, en este capítulo se incluye primeramente una descripción general de los ciclos combinados y de las calderas de recuperación de calor. Posteriormente se explica el modelo termoeconómico empleado y, finalmente, la metodología de optimización, y se presentan los resultados que se obtuvieron para las diferentes configuraciones de plantas analizadas.

**Palabras clave:** Algoritmos genéticos, optimización de centrales de ciclo combinado, calderas de recuperación de calor, análisis termoeconómico.

\* Profesora de Tiempo Completo. Facultad de Ingeniería, Universidad Autónoma del Estado de México, <mduran@fi.uaemex.mx>.

## INTRODUCCIÓN

El rendimiento de las centrales de ciclo combinado (CC) depende directamente del diseño específico de cada uno de sus componentes; entre éstos la caldera de recuperación de calor (CRC) es especialmente importante, ya que es a la unión entre el ciclo de gas y el de vapor. La profundización en el estudio de estos equipos resulta justificable porque su diseño depende de varios parámetros tanto térmicos como geométricos que deben seleccionarse cuidadosamente para lograr un mejor aprovechamiento de la energía térmica sin descuidar el costo del sistema.

Debido a la importancia de la caldera de recuperación de calor, varios trabajos (Attala *et al.*, 2001; Dechamps, 1995 y Tsatsaronis, 1993) se han enfocado a la optimización tanto térmica como termoeconómica de este elemento. Algunas investigaciones en este campo (como la de Valdés *et al.*, 2003) realizan la optimización termoeconómica usando como variables dependientes los parámetros termodinámicos que determinan el diseño de la CRC.

Teniendo como base algunos de los trabajos mencionados, en el presente capítulo se muestra la aplicación de la técnica de algoritmos genéticos para la optimización termoeconómica de centrales de ciclo combinado, dividida en dos partes: la primera que consiste en la selección de los parámetros óptimos de la CRC y la segunda que se enfoca en la optimización de los parámetros geométricos de la caldera, tratando de obtener la menor superficie de intercambio y pérdidas de carga. La metodología propuesta resulta de mucho interés, ya que generalmente cuando se diseña de una central térmica de este tipo, ésta no se optimiza, sino que se toman algunas propuestas del valor de los parámetros de diseño y, con base en ello, se simula y predice su desempeño.

Habitualmente, cuando una empresa generadora de energía adquiere una central térmica (de ciclo combinado o simple) la recibe “llave en mano” y no tiene manera de predecir su comportamiento en condiciones de diseño o a cargas parciales. Con la técnica que aquí se presenta se puede proyectar una central de ciclo combinado, optimizarla y predecir su comportamiento; con base en ello se puede solicitar una configuración más adecuada a las necesidades de cada caso.

Por otro lado, en la fase de diseño de la caldera de recuperación de calor, se observa que existe un sinnúmero de parámetros geométricos que permiten cumplir con las variables térmicas determinadas; sin embargo, es preciso encontrar la combinación que permita tener la menor superficie de intercambio de calor y pérdidas de carga sin sacrificar eficiencia. Existen muy pocos trabajos dedicados a ello, entre los cuales particularmente se encuentra el de Rapún (1999) quien optimiza el reparto de áreas en las diferentes secciones de la caldera o el de Franco y Giannini (2006), quienes lo que optimizan es el volumen ocupado por la CRC proponiendo

una geometría compacta. Sin embargo, no existe una metodología que permita optimizar la caldera a partir de los parámetros térmicos y geométricos de la misma.

En el presente documento se muestra una metodología que permite subsanar la problemática descrita y optimizar el diseño de una CRC. Para ello, primeramente se realiza una descripción de los componentes de un ciclo combinado y particularmente de la CRC; posteriormente se presenta la descripción del algoritmo genético que se aplica y el modelo de optimización termoeconómica y de reparto de áreas de la caldera. Finalmente, se presentan algunas tendencias que pueden apoyar en el diseño de calderas de recuperación de calor y ciclos combinados.

Es importante mencionar que la metodología que se describe puede ser aplicada también a la optimización de otros sistemas térmicos como sistemas de cogeneración, calderas con combustión o intercambiadores de calor.

*Nomenclatura*

$A$	Área de la caldera de recuperación de calor (CRC) ( $m^2$ )
$C_a$	Costo de amortización por año (€/año).
$C_c$	Costo de combustible por kWh (€/kWh)
$C_{Tc}$	Costo total de combustible (€/año)
$C_{TG}$	Costo fijo de la turbina de gas (€)
$C_{TV}$	Costo fijo de la turbina de vapor (€)
$C_{CRC}$	Costo fijo de la caldera de recuperación de calor (€)
$C_{kWh}$	Costo de generación (€/kWh)
$C_{om}$	Costo de operación y mantenimiento (€/año)
$C_{Tot}$	Costo total por año (€/año)
$C_{f-CRC}^{UA}$	Costo fijo por el incremento del área de la caldera de recuperación de calor (€)
$h$	Horas al año de trabajo de la planta (h/año)
$k$	Costo de una sección de la CRC por el producto $UA$ (€/K/kW)
$K$	Constante de penalización
$m_a$	Flujo másico de aire (kg/s)
$m_g$	Flujo másico de los gases de escape (kg/s)
$m_c$	Flujo másico de combustible (kg/s)
$N$	Vida útil de la planta (años)
$Pen$	Función de penalización
$S$	Precio de venta del kWh (€/kWh)
$PP$	Pinch point (K)
$AP$	Approach point (K)
$U$	Coefficiente global de transferencia de calor ( $Wm^{-2}K^{-1}$ )
$W$	Potencia total (kW)
$\bar{W}$	Potencia media anual de la planta (kW)

$X$	Título de vapor (%)
$Dp$	Pérdida de carga (mbar)
$DT$	Diferencia terminal de temperaturas (K)
$\eta$	Rendimiento térmico
$UA_{term}$	Producto UA obtenido a partir de los parámetros térmicos
$UA_{geom}$	Producto UA obtenido a partir de los parámetros geométricos
$v_v$	Velocidad del vapor

<i>Subíndices</i>	
AP	Alta presión
IP	Presión Intermedia
BP	Baja presión
inl	Condiciones de entrada
esc	Condiciones de escape
ec	Economizador
ev	Evaporador
Sc	Sobrecalentador
1	Entrada a la CRC
2	Salida del gas del sobrecalentador
3	Entrada del gas al economizador
4	Salida del gas de la caldera de recuperación de calor
a	Entrada del vapor a la turbina
b	Entrada del vapor al sobrecalentador
d	Salida del agua del economizador
e	Entrada del agua a la CRC

## I. GENERALIDADES SOBRE LOS CICLOS COMBINADOS DE TURBINAS DE GAS Y TURBINAS DE VAPOR

En el presente apartado, primeramente se describirán de manera muy general los ciclos combinados, con el fin de identificar cuáles son los elementos que los componen y los parámetros de diseño que influyen mayormente en su desempeño y costo.

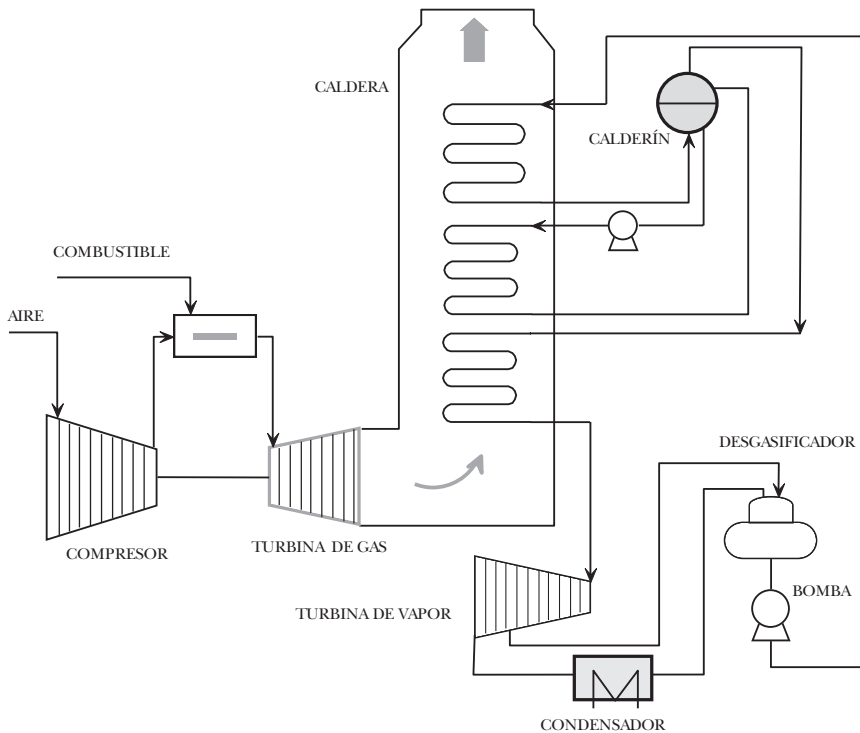
### I.1 Generalidades de una planta de ciclo combinado

Un ciclo combinado es un sistema térmico que aprovecha la energía de la combustión en dos etapas: un ciclo de alta temperatura (ciclo de gas) y un ciclo de baja temperatura (ciclo de vapor). Esta tecnología se ha ido desarrollando ya desde los años sesenta, cuando comenzaron a instalarse comercialmente ciclos combinados

que utilizaban ciclos de gas y vapor, pero anteriormente ya se había hablado incluso de ciclos mercurio-vapor (Horlock, 1991).

La configuración más simple de este tipo de plantas es la de un nivel de presión. Este tipo de configuración prácticamente no se instala en la actualidad; sin embargo, su descripción es muy útil para entender el funcionamiento de estos sistemas: consta de una turbina de gas, una caldera de recuperación de calor y una turbina de vapor (véase Figura I.1).

**Figura I.1**  
**Ciclo combinado de un nivel de presión**



En forma general, el funcionamiento de este tipo de plantas se puede describir de la siguiente manera: el ciclo de gas consta de un compresor para alimentar con aire a la cámara de combustión, en la que se realiza la combustión con un dosado muy pobre y cuyos productos se expanden en la turbina de gas, en la que se genera energía mecánica para, por una parte, impulsar al compresor y, por otra, impulsar la carga (alternador). Posteriormente, los gases calientes se introducen

en la caldera de recuperación de calor, en la cual se genera el vapor sobrecalentado que alimentará la turbina de vapor.

Con la unión de ambas plantas de potencia se obtiene un rendimiento mayor que el que se obtendría con cada una de ellas por separado (mayor a 50%). Esto es porque se aprovecha la energía térmica residual de los productos del escape de la turbina de gas, consiguiendo, además, que disminuya la temperatura de los gases que escapan a la atmósfera. En el siguiente subapartado se describe con más detalle la caldera de recuperación de calor (CRC) por ser el elemento que compete a este análisis.

### ***1.1.1 Caldera de recuperación de calor (CRC)***

La función de una CRC es recuperar la energía calorífica proveniente de los gases de escape de la turbina de gas y emplearla para generar vapor en las condiciones deseables para alimentar la turbina de vapor. La diferencia principal entre éstas y las calderas convencionales es el mecanismo de transferencia de calor, ya que estas últimas, al tener combustión, presentan focos con temperaturas muy altas en los cuales el intercambio de calor por radiación es predominante. En cambio, en las CRC el mecanismo de transferencia de calor es en general de carácter convectivo.

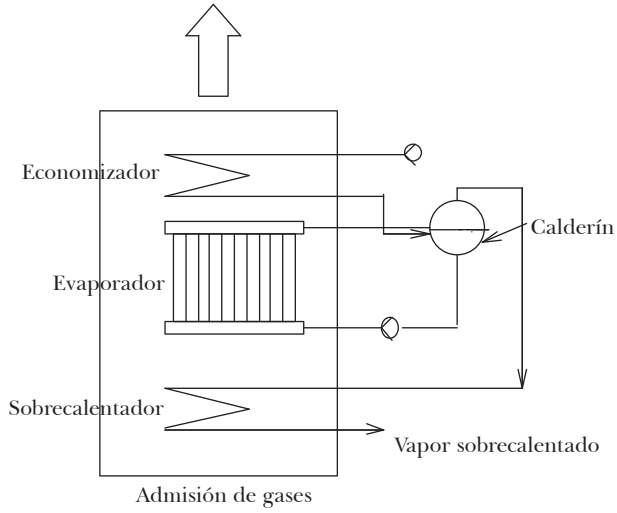
Sea cual fuere el tipo de la caldera de recuperación de calor, ésta consiste en un intercambiador de calor de flujos perpendiculares en el que el agua, como líquido subenfriado, entra por la zona de escape de los humos, es decir, la última zona de la CRC conocida como *economizador* (véase la Figura I.2). En esta sección el agua se lleva hasta una temperatura un poco menor que la de saturación, correspondiente a la presión de trabajo.<sup>1</sup>

Posteriormente el agua, prácticamente en su estado de saturación, se envía a un depósito conocido como *colector* en el que se produce la separación de las fases. La fase líquida del agua del colector se hace circular por otra sección de la CRC, que está colocada inmediatamente después del economizador y que se conoce con el nombre de *evaporador*, en la que se produce la evaporación del agua y el vapor resultante es llevado nuevamente al colector.

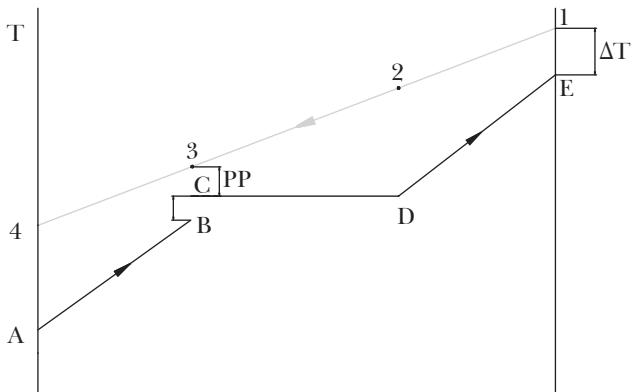
Por su parte, el vapor que se ha separado en el colector se hace circular por la primera sección de la CRC, que está justo a la entrada de los gases provenientes de la turbina de gas. Esta sección es conocida como *sobrecalentador* y en ella se lleva al vapor a las condiciones de entrada de la turbina de vapor (sobrecalentado).

<sup>1</sup> El hecho de que en esta sección no se alcance la temperatura de saturación del vapor es muy importante puesto que debe evitarse que se presente evaporación en este elemento, por ello debe existir un margen de seguridad en cuanto a la temperatura.

**Figura I.2**  
**Caldera de recuperación de calor de un nivel de presión**



**Figura I.3**  
**Diagrama energía-temperatura de una CRC de un nivel de presión**



El diagrama energía-temperatura de esta configuración de CRC se presenta en la Figura I.3. Se observa la línea recta continua (de derecha a izquierda) que representa el enfriamiento de los gases de la turbina de gas y, de izquierda a derecha, la línea que representa el calentamiento del agua. En el diagrama pueden distinguirse claramente las tres secciones de la CRC; la primera, de A a B, corresponde al economizador; la segunda, de C a D, al evaporador; y la tercera, de D a E, al sobrecalentador. Puede observarse que la región de C a D es plana debido a que en esta zona se produce la evaporación del agua y, por tanto, no hay incremento de temperatura de la misma. Por otra parte, en el mismo diagrama se observan también tres diferencias de temperaturas importantes:

- a) *Pinch Point (PP)*. Corresponde a la diferencia entre la temperatura de los gases en la salida del evaporador y la del agua a la entrada del mismo (diferencia entre el punto C y 3 en el diagrama). Esta diferencia es un parámetro muy importante en el diseño del CRC, ya que cuanto menor sea ésta mayor será el rendimiento energético del ciclo, pero también mayor será el área de intercambio de calor y, por tanto, el costo.
- b) *Approach point (AP)*. Es la diferencia entre la temperatura del agua que abandona el economizador y la temperatura de saturación a la presión de trabajo (diferencia entre los puntos B y C del diagrama). Esta diferencia de temperaturas es un margen de seguridad para evitar evaporación en el economizador a cargas parciales. Algunos autores (Horlock, 1991) sugieren que su valor no debe ser menor a 3K, aunque depende del nivel de presión del que se trate.
- c) *Diferencia terminal de temperaturas ( $\Delta T$ )*. Corresponde a la diferencia entre la temperatura del vapor sobrecalentado que abandona la caldera y la temperatura de entrada de los gases a la misma. Su valor tiene una influencia significativa tanto en el rendimiento como en el costo; por ese motivo, este parámetro, al igual que el *approach point* y el *pinch point*, se considerará como variable de diseño dentro de la optimización termoeconómica.

## II. DESCRIPCIÓN DE LA METODOLOGÍA EMPLEADA

### II.1 Generalidades de los algoritmos genéticos

Como ya se ha mencionado, se empleará un algoritmo genético (AG) para la solución del problema de optimización. Este tipo de algoritmos se ha empleado con éxito en otro tipo de aplicaciones donde se involucran ecuaciones no lineales que resultan sumamente difíciles de resolver por un método de gradiente.

De manera general, un AG funciona imitando el comportamiento de la naturaleza, evaluando en la función objetivo una serie de individuos que contienen el valor de las variables independientes (que para este caso se llamará función salud). De entre ellos se seleccionarán aquellos que presenten el mejor valor de la función objetivo para que sean los padres de las siguientes generaciones, aplicando operadores como cruce y mutación (entre otros). El algoritmo continúa hasta que se obtiene el valor más óptimo. Los fundamentos de los algoritmos genéticos y sus aplicaciones se encuentran descritos en diversas investigaciones (De Jong, 1975; Goldberg, 1989; Herrera y Verdegay, 1996; Martorell *et al.*, 2002).

El campo de aplicación de los algoritmos genéticos es amplio, incluye ingeniería nuclear (Martín del Campo *et al.*, 2001) y la ingeniería estructural (Coello y Chistiansen, 2000). Asimismo ya ha sido aplicada con éxito para la optimización termoeconómica de las centrales de ciclo combinado (Toffolo y Lazzaretto, 2002), mientras que Mohagheghi y Shayegan (2009) también la aplican para optimizar parámetros de diseño de una caldera de recuperación de calor.

Trabajos que sustentan el que aquí se presenta (Valdés *et al.*, 2003; Durán, 2004) muestran la descripción de la optimización de diferentes configuraciones de ciclo combinado aplicando algoritmos genéticos.

## **II.2 Descripción del algoritmo de optimización empleado**

Como se ha explicado, el objetivo principal es desarrollar una metodología para la optimización termoeconómica de una central de ciclo combinado, para lo cual era necesario seleccionar una metodología adecuada y se decidió aplicar la de algoritmos genéticos debido a que es fácil de programar y ha dado buenos resultados en problemas complejos.

Se ha observado que resulta muy viable la aplicación de este tipo de algoritmos para la optimización de sistemas térmicos como el que aquí se presenta, puesto que estos problemas incluyen ecuaciones trascendentes que impiden la aplicación de un método de gradiente tradicional y que deben resolverse de manera iterativa.

Así pues, en esta aplicación primeramente se realizó el programa en Visual Basic de un algoritmo genético básico (únicamente incluye los operadores cruce y mutación); sin embargo, al realizar otras pruebas se observó que para las configuraciones más complejas (dos y tres niveles de presión), en algunas ocasiones el algoritmo se saturaba en un máximo local. Esto se debe a que cuando el problema presenta muchas variables independientes el cromosoma se hace muy largo, lo que ocasiona que haya poca diversidad en la población. Por tal motivo se consi-

deró necesario incluir otros operadores que introdujeran diversidad y aseguraran una mejor convergencia del algoritmo.

Con el fin de mantener la diversidad en la población se decidió, primeramente, modificar la mutación en el algoritmo, incluyendo la mutación adaptativa; no obstante, se observó que este operador no era suficiente para diversificar la población y se optó por aplicar también la herramienta de reinicialización. Mayor información sobre el algoritmo empleado se presenta en Durán (2004). Los parámetros que se emplearon para el algoritmo de optimización se muestran en la Tabla II.1.

**Tabla II.1**  
**Parámetros del algoritmo genético modificado**

<i>Parámetros del algoritmo</i>	<i>Valor</i>
Porcentaje de cruce	65%
Porcentaje de mutación inicial	17%
Población	120
Número de generaciones	100
Número de reinicializaciones	4
% que permanece en cada reinicialización	50%

### III. DESCRIPCIÓN DEL MODELO DE OPTIMIZACIÓN TERMOECONÓMICO

En la metodología que aquí se describe, se parte de la hipótesis de que es posible optimizar el ciclo combinado modificando solamente algunos parámetros de diseño de la CRC, considerando una cierta turbina de gas.

Se plantea un modelo termoeconómico que define el costo del ciclo en función de sus componentes y, particularmente, el costo de la CRC en función de la superficie de intercambio. Con lo anterior se plantea la función objetivo para dos criterios de optimización, como se describe a continuación.

#### III.1 Modelo de costo y modelo de flujo anual de caja

Actualmente existen diferentes modelos de análisis termoeconómico de plantas de potencia. Algunos de ellos se enfocan principalmente a la valoración del “kilowatt instalado”, que simplemente es el cociente entre la inversión realizada y la potencia de la planta. Este criterio no conduce necesariamente a buenos resul-

tados económicos, ya que una inversión pequeña no garantiza ni bajos costos de producción ni grandes beneficios.

Otros modelos se basan en analizar y minimizar el costo energético, el cual impacta principalmente en el costo de operación de la planta (Tsatsaronis, 1993). Este tipo de modelos es de gran utilidad para sistemas en los que se tiene más de un producto (calor y trabajo) como los de cogeneración.

Por su parte, el modelo termoeconómico que se aplica en este trabajo considera dos criterios de optimización: el primero de ellos pretende minimizar el costo de generación, mientras que el segundo pretende maximizar el flujo anual de caja. No se trata de una optimización multiobjetivo, pues los problemas se resuelven por separado y se comparan los resultados obtenidos. Como se observará más adelante, los resultados son diferentes dependiendo del criterio que se elija y es de esperar que si se decide minimizar el costo de generación, el área de la caldera de recuperación de calor será menor que si se elige maximizar el flujo de caja. Los modelos termoeconómicos aplicados se describen a continuación:

a) *Costo de generación*. Corresponde al costo total por año, dividido por la producción media anual:

$$C_{kWh} = \frac{C_{Tot}}{\bar{W} \cdot h} \quad (3)$$

Donde  $\bar{W}$  es la potencia media anual y  $h$  corresponde al número de horas de operación de la planta al año, que en este caso se consideraron 7,000 h.

b) *Flujo anual de caja*. Se define como el ingreso total en un año menos el costo total anual:

$$B = I_{Tot} - C_{Tot} \quad (4)$$

En esta ecuación el ingreso anual ( $I_{tot}$ ) está definido a partir de la potencia neta generada y el precio de venta de la energía generada ( $S$ ), es decir:

$$I_{Tot} = S \cdot \bar{W} \cdot h \quad (5)$$

### III.2 Comparación de los modelos de flujo de caja y costo de generación

Se plantea que al realizar la optimización utilizando los criterios propuestos por separado los resultados serían distintos; sin embargo, con ambos se obtendrá un

diseño con un rendimiento alto. La razón por la cual se selecciona un criterio de optimización u otro depende de las condiciones del mercado y del objetivo de la optimización. Si existe una fuerte competencia dentro del mercado eléctrico y es difícil vender la energía generada, se debe minimizar el costo de generación para obtener un diseño más factible. Por el contrario, en un mercado regulado (como en México), con los precios fijos y en el cual es fácil vender la energía, la optimización debe conducir a un diseño que proporcione mayores beneficios, por lo que se debe intentar obtener un flujo de caja máximo. En este último caso es de esperar que aumente el costo de generación pero que, al mismo tiempo, se presente una mejora en el rendimiento y en la potencia generada, con lo cual puede aumentar el ingreso obtenido.

Según se ha explicado, para una turbina de gas dada, el rendimiento del ciclo aumenta con el área de intercambio de la CRC; sin embargo, este incremento en el área tendrá como consecuencia un incremento en el costo, por lo que deben existir óptimos para cada criterio. Los valores de estos óptimos se obtendrán cuando las funciones correspondientes a cada modelo con respecto a la superficie de intercambio sea nula, es decir:

$$\frac{dB}{dA} = 0 \quad (6)$$

$$\frac{dC_{kWh}}{dA} = 0 \quad (7)$$

Al sustituir las ecuaciones (3) y (5) en la ecuación (4), se puede escribir:

$$B = (S - C_{kWh}) \cdot \bar{W} \cdot h \quad (8)$$

Aplicando la derivada con respecto a la superficie de intercambio de la CRC se obtiene:

$$\frac{dB}{dA} = -\frac{dC_{kWh}}{dA} \cdot \bar{W} \cdot h + (S - C_{kWh}) \cdot h \cdot \frac{d\bar{W}}{dA} \quad (9)$$

Ahora, evaluando para el punto donde el costo de generación es mínimo:

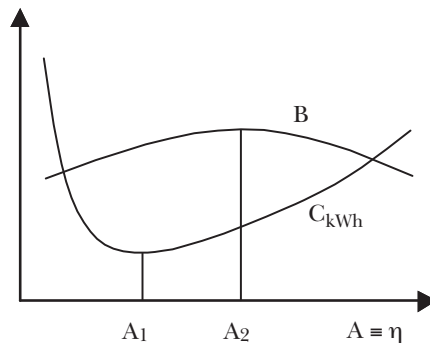
$$\frac{dB}{dA} = (S - C_{kWh}) \cdot \frac{d\bar{W}}{dA} \cdot h \quad (10)$$

La expresión (10) siempre es positiva, ya que el precio de venta es mayor que el costo de generación ( $S > C_{kWh}$ ) y  $\frac{d\bar{W}}{dA} > 0$ . Por tanto, es de esperar que la curva

que define la variación flujo de caja con respecto al área será creciente en el punto en el que el costo de generación alcance su valor óptimo.

Considerando lo anterior, se espera que las curvas de la variación del flujo de caja y del costo de generación frente al área de intercambio de calor sean como se ilustra en la Figura III.4. En ella puede observarse que el máximo flujo de caja tiene una mayor área de intercambio que aquella correspondiente al mínimo costo de generación ( $A_2 > A_1$ ) y, por tanto, un mayor rendimiento. Por lo anterior, el diseño de máximo beneficio permite mayores costos de generación porque se ve recompensado por los mayores ingresos.

**Figura III. 4**  
**Flujo de caja máximo y costo de generación mínimo**



### III.3 Modelo de atribución de costos

El costo total anual de la ecuación 3 incluye el costo del combustible, el costo de operación y mantenimiento y el costo de amortización:

$$C_{Tot} = C_{Tc} + C_{om} + C_a \quad (11)$$

Cada uno de estos costos se define como sigue:

#### III.3.1 Costo total de combustible

Éste es un costo variable, ya que es en función de la energía producida y está definido por la siguiente ecuación:

$$C_{TC} = C_c \cdot \frac{\overline{W}}{\eta} \cdot h \quad (12)$$

Donde  $C_c$  corresponde al precio del combustible,  $h$  es el número de horas de operación de la planta y  $\eta$  es el rendimiento del ciclo combinado.

### **III.3.2 Costo de amortización**

El costo de amortización corresponde a un costo fijo cuyo valor debe recuperarse año con año para que, al cabo de cierto número de años, se haya recuperado la inversión actualizada a ese año:

$$C_a = C_F \cdot \beta \quad (13)$$

Donde  $\beta$  es el factor de anualidad y es función de la tasa de interés y del número de años en los que se pretende recuperar la inversión:

$$\beta = \frac{i(1+i)^N}{(1+i)^N - 1} \quad (14)$$

El costo fijo incluye el costo de los componentes del ciclo (turbina de gas, turbina de vapor, etc.), y el costo asociado a la CRC, es decir:

$$C_{Fl} = C_{TG} + C_{CV} + C_{CRC} + C_{resto} \quad (15)$$

En la expresión anterior,  $C_{resto}$  incluye costos que no están involucrados directamente con el equipo, como el del terreno, el de la instalación, transporte, etcétera. Su valor depende de muchos factores y es muy difícil estimarlo, pero algunos autores (Attala *et al.*, 2001; Kelhofer, 1999) lo consideran 50% de la inversión total.

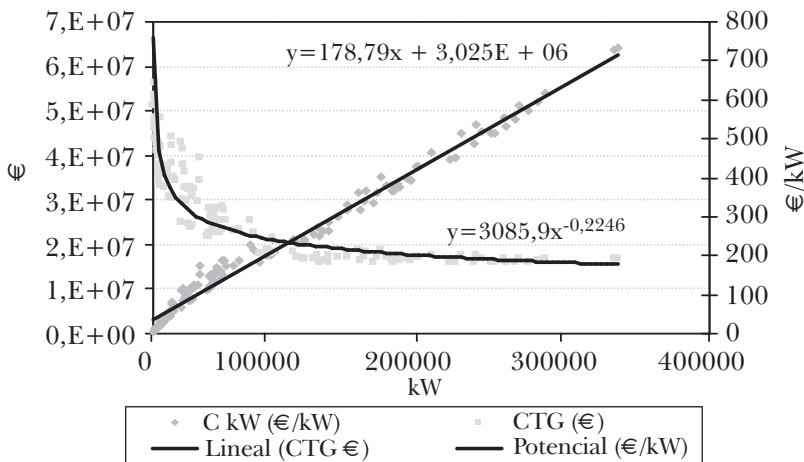
El costo de las turbinas de gas y de vapor es función de la potencia nominal, y el costo de la CRC es función de la superficie de intercambio, como se describe a continuación:

#### *a) Costo de la turbina de gas*

Este componente no es parte del diseño del ciclo combinado, sino que es un elemento estandarizado y durante la fase de diseño sólo se selecciona el más adecuado en función de la potencia que se requiere. El costo de estos elementos es proporcionado por el fabricante y depende de su potencia nominal.

En la Figura III.5 se presenta la gráfica del costo por unidad de potencia y de costo total de algunas turbinas comerciales (costos del año 2002). Estos datos son precios reales del mercado y fueron obtenidos a partir de una base de datos de una versión de demostración del programa GTPPro. Puede observarse que, independientemente de las economías de escala y del fabricante, el costo de las turbinas de gas sigue una tendencia lineal en función de la potencia del equipo, mientras que el costo específico presenta una tendencia decreciente.

**Figura III.5**  
**Costo de las turbinas de gas**



Los datos presentados en la figura anterior se ajustaron mediante regresión lineal, obteniendo para el costo de la turbina de gas la siguiente ecuación:

$$C_{TG} = 0.1788 \cdot W + 3.0253 \quad (16)$$

Donde el costo se da en millones de euros y la potencia en MW.

La ecuación anterior define al costo de la turbina de gas en función de la potencia nominal, sin tener en cuenta los parámetros de diseño de dicha turbina.

#### b) Costo de la turbina de vapor

Las turbinas de vapor no son equipos tan comerciales como las turbinas de gas y se cuenta con muy poca información respecto a su costo. Sin embargo, de acuer-

do con algunas evaluaciones realizadas en trabajos previos (Rovira, 2004; Durán, 2004), se estima el costo de una forma similar al de la turbina de gas como:

$$C_{TV} = 2.75 + 0.115 \cdot W_{TV} \quad (17)$$

*c) Costo de la caldera de recuperación de calor*

El costo de la CRC incluye varios componentes como el costo del diseño, de la estructura e instalación y de los materiales de la superficie de intercambio de calor. Sería deseable conocer todos estos costos; sin embargo, se cuenta con muy poca información al respecto, por ello en el presente trabajo se consideró el costo de la caldera únicamente como el correspondiente a las superficies de intercambio y los demás se incluyeron como parte de  $C_{resto}$  en la ecuación (15).

Es necesario un modelo que sea capaz de predecir el costo de dicha caldera en función de sus parámetros termodinámicos, sin conocer el diseño geométrico de la misma. En la bibliografía existen varios modelos que desarrollan correlaciones para predecir este costo. Algunos de ellos (Franco y Russo, 2002) calculan el valor de la superficie de intercambio a partir del producto  $UA$  (producto del coeficiente global de transferencia de calor por el área de cada sección de intercambio de calor de la caldera) considerando al coeficiente de transferencia de calor constante para cada sección. Y definen el costo de la caldera en función de la superficie de intercambio, de la forma:

$$C_{f-CRC}^A = \sum_{sec} k_{sec} A_{sec} \quad (18)$$

Donde el subíndice *sec* se refiere a cada una de las secciones de la caldera de recuperación de calor y la constante  $k_{sec}$  expresa el costo del material correspondiente a la misma.

Otros autores (Attala *et al.*, 2001; Valero *et al.*, 1994) en lugar de considerar el coeficiente global de transferencia de calor constante definen el costo de cada sección de la caldera en función del producto  $UA$  de la forma:

$$C_{f-CRC,sec}^{UA} = \sum_{sec} k_{sec} (UA)_{sec}^x \quad (19)$$

Donde el exponente ( $x$ ) al que se encuentra elevado el producto  $UA$  se incluye debido a la economía de escala; es decir que cuanto mayor sea el área de la caldera más barata será la unidad de  $UA$ . Algunos autores (Valero *et al.*, 1994;

Tsatsaronis y Pisa, 1994) consideran este exponente como 0.8 y proponen el costo de la caldera de la siguiente forma:

$$C_{f-CRC,sec}^{UA} = \sum_{sec}^{n_{sec}} k_{sec} (UA)_{sec}^{0.8} \quad (20)$$

Se aplicó este modelo durante la optimización termoeconómica. Posteriormente, en el modelo de optimización geométrica se determinó el valor del coeficiente global de transferencia de calor ( $U$ ) y la superficie de intercambio de calor para cada sección de la caldera.

Sería deseable obtener el área de intercambio y el coeficiente de transferencia y con ello obtener el costo de la CRC de la ecuación (20); sin embargo, hacer esto sería un procedimiento muy largo, por ese motivo se decidió considerar, a manera de aproximación en el modelo termoeconómico, el costo de la CRC como lo proponen Valero *et al.* (1994) (ecuación 21) y posteriormente, una vez que se ha realizado la optimización, calcular el área de la caldera optimizada.

$$C_{f-CRC}^{UA} = \sum_{ec} k_{ec} (UA)_{ec}^{0.8} + \sum_{ev} k_{ev} (UA)_{ev}^{0.8} + \sum_{sh} k_{sc} (UA)_{so}^{0.8} \quad (21)$$

En la ecuación anterior, los valores de los coeficientes  $k$  de cada sección de la caldera dependen del material del que está hecha y son diferentes para cada una de ellas. Se cuenta con poca información acerca de los costos de los materiales de la caldera de recuperación de calor, por lo que resulta difícil definir este valor. Además, la poca información que se tiene es útil para determinar un costo por unidad de área pero no por unidad de producto  $UA$  (€/kW/K). Así pues, para establecer este valor se consideraron algunos reportes (Staff report, 2000) y trabajos realizados por otros autores (Attala *et al.*, 2001).<sup>2</sup>

### III.3.3 Costo de operación y mantenimiento

Incluye todos los costos de mantenimiento del equipo; se divide en dos partes: una correspondiente al costo fijo, es decir que no depende de la energía generada, y otra que incluye los gastos de operación y mantenimiento y que depende de la producción. Su valor es difícil de estimar; sin embargo, en muchos casos éste se considera igual a 10% del costo total, es decir:

$$C_{om} = 0.10 \cdot (C_{Tot}) \quad (22)$$

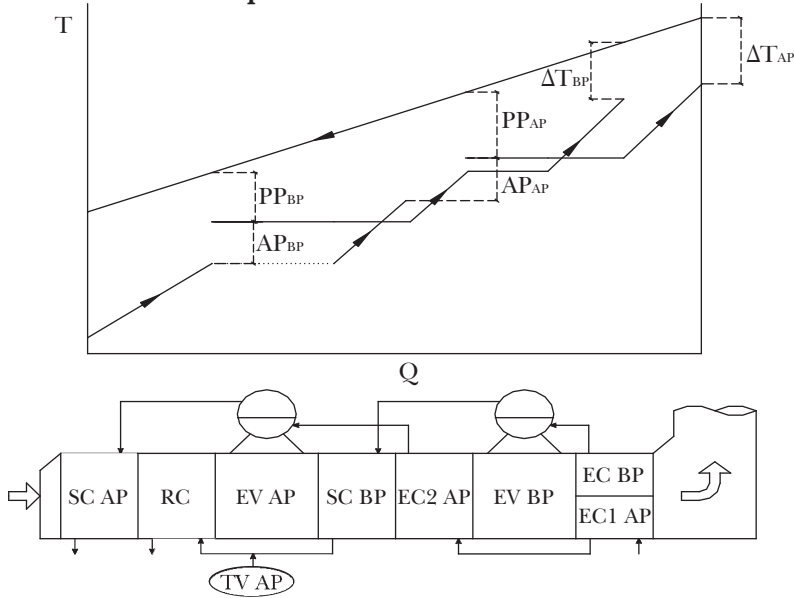
<sup>2</sup> El valor de las constantes que se consideraron fueron  $k_{ec} = 250$  USD/K,  $k_{ev} = 400$  USD/K y  $k_{sc} = 800$  USD/K.

Una vez que ya se tiene establecido el modelo termoeconómico que se va a emplear se puede llevar a cabo el proceso de optimización del sistema. Este proceso se realizó mediante Visual Basic, el cual se describe en la sección siguiente.

### III.4 Optimización del ciclo combinado de dos niveles de presión con recalentamiento

Para mostrar la aplicación del modelo, en este apartado se muestran los resultados de la optimización de una configuración de un ciclo combinado. La configuración que se seleccionó es una de dos niveles de presión con recalentamiento (2PR), se eligió ésta porque es una de las más aplicadas en la práctica y porque presenta un rendimiento alto. Las variables de diseño y el intervalo de búsqueda correspondiente se definen en la tabla III.2, mientras que el esquema de la caldera y el diagrama energía temperatura correspondiente se presentan en la Figura III.6.

**Figura III.6**  
**Esquema y diagrama energía temperatura de un ciclo de dos niveles de presión con recalentamiento**



- |  |  |
|--|--|
| ECBP: Economizador de baja presión.    | SCBP: Sobrecalentador de baja presión. |
| EC1AP: Economizador 1 de alta presión. | EVAP: Evaporador de alta presión.      |
| EVBP: Evaporador de baja presión       | RC: Recalentador.                      |
| EC2AP: Economizador 2 de alta presión. | SCAP: Sobrecalentador de alta presión. |

**Tabla III.2**  
**Variables de diseño y límites del intervalo de búsqueda para la configuración de dos niveles de presión con recalentamiento**

<i>Variables de diseño</i>	<i>Límite inferior</i>	<i>Límite superior</i>
Presión de baja (bar)*	3	8
Pinch point de baja (K)	3	10
Approach point de baja (K)	3	8
Diferencia terminal de temperaturas en el SC de baja (K)	5	30
Presión de alta* (bar)	20	180
Pinch point de baja (K)	3	15
Approach point de baja (K)	3	8
Diferencia terminal de temperaturas en el SC de alta (K)	15	70

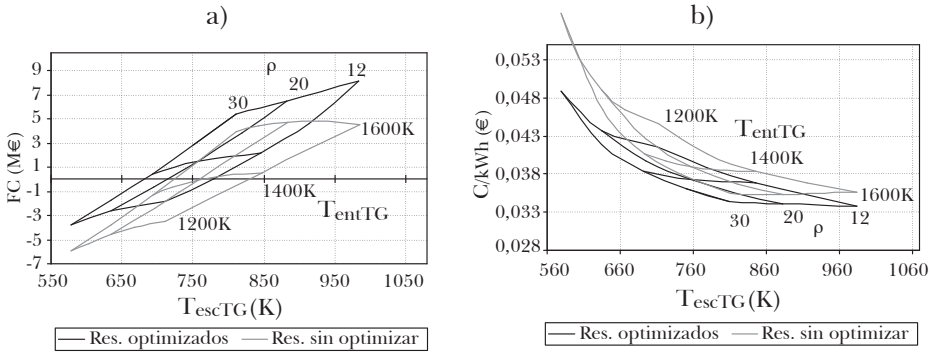
\* Los límites inferior y superior de la presión en el colector se ajustaron en algunos casos debido a que cuando la temperatura de entrada a la CRC es muy elevada, la presión en el nivel de alta alcanza valores muy grandes.

Utilizando estos límites de las variables de diseño se realizó la optimización termoeconómica de esta configuración, considerando diferentes valores de la relación de compresión ( $\rho$ ) y temperatura de entrada a la turbina de gas ( $T_{entTG}$ ), en función de la temperatura de escape de la misma ( $T_{escTG}$ ), comparando una configuración optimizada termoeconómicamente (línea oscura en las Figuras III.7 y III.8) con una no optimizada en la cual los parámetros de diseño son seleccionados de acuerdo con los valores planteados por algunos fabricantes (línea gris en las Figuras III.7 y III.8).

En dichas Figuras se representa la variación de los resultados de la optimización de esta configuración (flujo de caja, costo y rendimiento) para diferentes valores de los parámetros del ciclo de gas. Se puede observar que existen diferencias entre los resultados optimizados y aquéllos sin optimizar; particularmente para el caso del rendimiento, se observa que mediante la optimización se puede incrementar el rendimiento global del ciclo en cerca de cuatro unidades porcentuales.

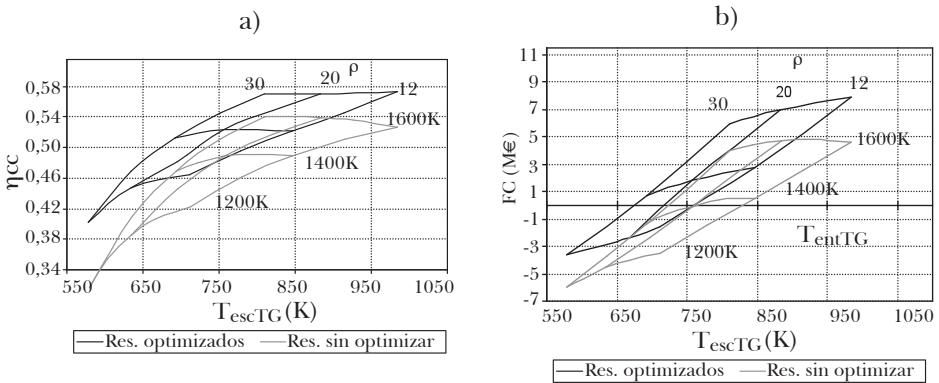
Por otro lado, si se comparan las gráficas que representan el flujo de caja y el costo obtenido con una configuración de dos niveles de presión sin recalentamiento (Figuras III.7a) con aquella con recalentamiento (Figura III.8b), se observa que, aun cuando el costo de generación es muy similar para ambas configuraciones, el flujo de caja es mayor para la de dos niveles de presión con recalentamiento, para las características seleccionadas de la turbina de gas.

**Figura III.7**



- a) Variación del flujo de caja optimizado con respecto a los parámetros de diseño de la turbina de gas.
- b) Variación del coste de generación optimizado con respecto a los parámetros de diseño de la turbina de gas.

**Figura III.8**



- a) Variación del rendimiento, hallada mediante la optimización del flujo de caja, con respecto a los parámetros de diseño de la turbina de gas.
- b) Variación del flujo de caja optimizado con respecto a los parámetros de diseño de la turbina de gas de una configuración de 2 niveles de presión sin recalentamiento.

### III.4.1 Parámetros de diseño del ciclo de dos niveles de presión con recalentamiento

Los resultados de la optimización termoeconómica (para diferentes valores de relación de compresión y temperatura de entrada a la turbina de gas) se presentan en la Tabla III.3. En ella puede observarse el efecto de las condiciones de diseño del ciclo de gas sobre el valor hallado por el AG al optimizar la configuración. Este tipo de resultados pueden ser útiles para la toma de decisiones en cuanto al diseño de una planta de ciclo combinado. Los tomadores de decisiones pueden considerar un intervalo de valores de los parámetros de diseño cercano al que aquí se presenta o aplicar el modelo que aquí se propone para evaluar la configuración del ciclo que pretenden instalar. De acuerdo a esto, en el apartado final del presente capítulo se mostrarán algunas tendencias que pueden apoyar al diseño de este tipo de sistemas.

**Tabla III.3**  
**Resultados de la optimización con respecto al flujo de caja de una configuración de dos niveles de presión con recalentamiento para diferentes condiciones de entrada de la turbina de gas**

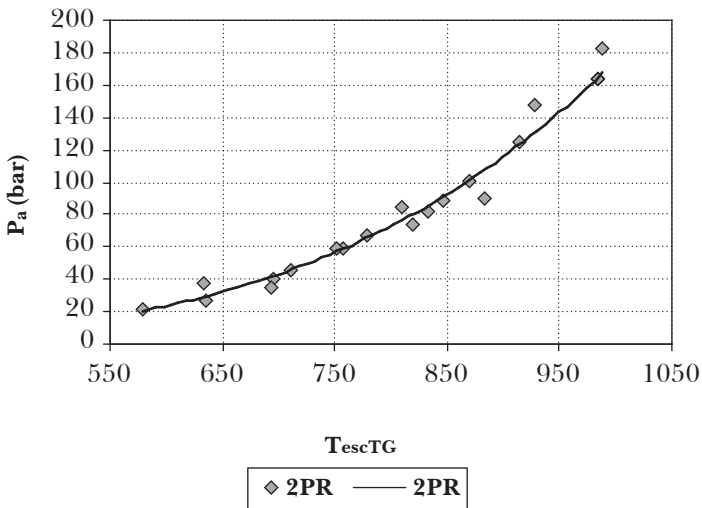
Parámetros del ciclo de gas			Parámetros de diseño del ciclo combinado								Resultados de la optimización		
$rc$	$T_{entTG}$ (K)	$T_{escTG}$ (K)	$P_b$ (bar)	$PP_b$ (K)	$AP_b$ (K)	$\Delta T_b$ (K)	$P_a$ (bar)	$PP_a$ (K)	$AP_a$ (K)	$\Delta T_a$ (K)	Rend- cc	FC (M€)	$C_{kwh}$ (€)
12	1200	711,8	3	5	3	6,5	45,05	9	3,0	70	0,466	-1,75	0,0410
12	1400	846,7	4,05	5,3	3,3	20,5	88,7	10,2	3	20,2	0,522	2,150	0,0369
12	1600	984,5	6,3	5,8	3,5	21,6	163,9	11,6	3,3	26	0,573	8,116	0,0338
20	1200	634,4	3	4,8	3	6,9	28,1	10,8	3	70	0,447	-2,540	0,0437
20	1400	757,5	3,2	5,1	3	6,5	58,5	11,2	3	70	0,523	1,521	0,0372
20	1600	884,0	4,88	5,3	3	17,1	110,1	11,4	3	20,7	0,571	6,528	0,0341
30	1200	578,9	3	4,7	3	7,5	21,0	11,4	3	70	0,403	-3,732	0,0488
30	1400	693,1	3	4,8	3	7,3	40,8	11	3	70	0,513	0,462	0,0383
30	1600	810,9	4	4,9	3,07	7,7	82,8	12,6	3	70	0,569	5,462	0,0344

### III.4.2 Tendencias de los valores de la presión en los niveles de alta y de baja

Se realizó también un análisis de la influencia que tiene la temperatura de entrada de los gases a la CRC sobre algunos parámetros de diseño. Se observa que el

parámetro que se ve mayormente influenciado por el valor de la temperatura de escape de la turbina de gas es la presión de los colectores de alta y de baja. En la Figura III.9 se muestra la variación de la presión en los niveles de alta y de baja con la temperatura de los gases.

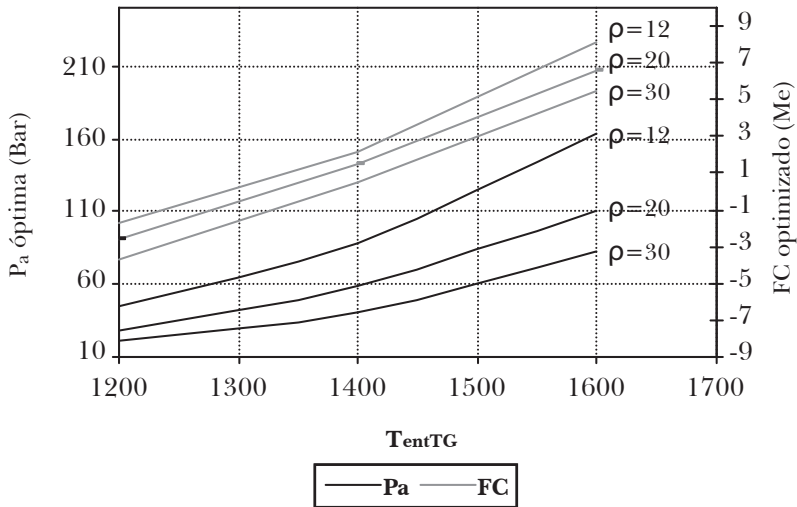
**Figura III.9**  
**Variación de la presión de alta en función de la temperatura de escape de la turbina de gas**



Por otro lado, en la Figura III.10 se muestra la gráfica de la variación de la presión en el nivel de alta y del flujo de caja optimizado en función de la temperatura de entrada a la turbina de gas y la relación de compresión. Mediante esta gráfica, a partir de los parámetros de diseño del ciclo de gas, se puede obtener el flujo de caja esperado y el valor que debe tener la presión en el nivel de alta para alcanzar este valor.

El análisis que se realizó en el presente apartado para la configuración de dos niveles de presión con recalentamiento se puede realizar para cualquier configuración, aplicando la metodología propuesta, de esta manera se podrá determinar el intervalo dentro del cual es recomendable que se encuentren los valores de los parámetros de diseño para obtener un diseño termoeconómicamente optimizado.

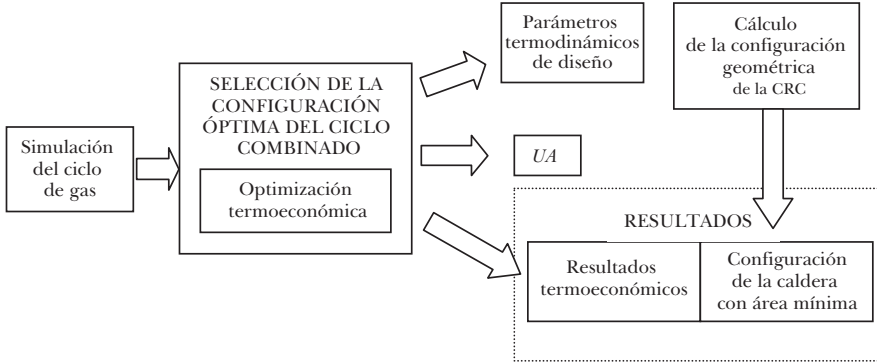
**Figura III.10**  
**Variación del flujo de caja optimizado y la correspondiente presión de alta para diferentes valores de  $\rho$  y  $T_{entTG}$**



#### IV. DESCRIPCIÓN DE LA METODOLOGÍA PARA EL CÁLCULO GEOMÉTRICO DE LA CALDERA DE RECUPERACIÓN DE CALOR

Una vez realizada la optimización termoeconómica de una caldera de recuperación de calor, se cuenta con el valor de los parámetros termodinámicos de diseño de la misma; sin embargo, ahora se presenta el problema de realizar el diseño geométrico de la caldera, ya que existe un sinnúmero de combinaciones de parámetros geométricos (número de tubos, longitud, diámetro, etc.) que podrían ajustarse para obtener el valor de los parámetros termodinámicos determinados. Por tanto, es necesario desarrollar una metodología que además de determinar los parámetros geométricos permita mantener la superficie de intercambio en un valor razonable para minimizar el costo. La metodología completa de optimización tanto termoeconómica como geométrica se representa en la Figura IV.11. En el presente apartado se describe la segunda parte de la metodología, correspondiente al diseño geométrico de la CRC; para una mejor comprensión de la misma es necesario describir primero parámetros geométricos de estos sistemas.

**Figura IV.11**  
**Estrategia de optimización de las centrales de ciclo combinado**

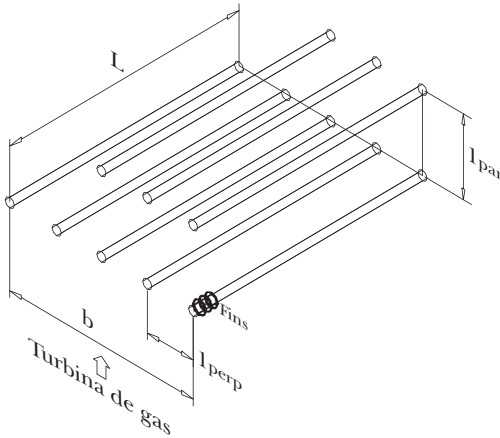


#### IV.1 Descripción de los parámetros geométricos de la CRC

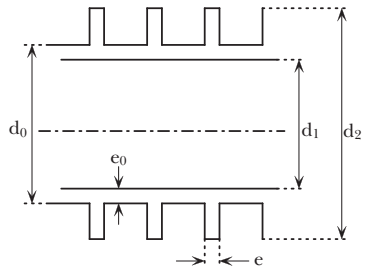
La configuración más simple de una caldera de recuperación de calor acoplada al ciclo combinado se muestra en la Figura I.1; como se ha mencionado, está compuesta por diferentes secciones de transferencia de calor. Cada una de ellas puede considerarse como un intercambiador de calor de flujos cruzados. Su sección transversal, así como algunos parámetros geométricos (separación entre los tubos, el diámetro de los mismo, etcétera) se muestran en la Figura IV.12. En muchos diseños de CRC se emplean tubos aleteados para mejorar la transferencia de calor, los parámetros geométricos que determinan el diseño de éstos también son importantes, en la Figura IV.13 se muestra la sección transversal del tubo aleteado.

A manera de resumen, en la Tabla IV.4 se muestran las variables geométricas de diseño de una CRC. Puede observarse que los valores de algunos de ellos son comunes a todas las secciones de la caldera. Asimismo, en el presente trabajo se está considerando el ancho de la caldera común a todas las secciones, esto se hace para facilitar la fabricación de la misma.

**Figura IV.12**  
**Banco de tubos**



**Figura IV.13**  
**Tubo aleteado**



Los parámetros que se describen en la tabla serán las variables que se considerarán para el diseño de la CRC.

**Tabla IV.4**  
**Parámetros geométricos de la caldera de recuperación de calor**

<i>Parámetro geométrico</i>	<i>Nomenclatura</i>	<i>Valor</i>	<i>Sección</i>
Número de tubos por hilera	N		Econ Evap. Sobre
Número de hileras de tubos	$N_{tprof}$		
Número de aletas por unidad de longitud	z		
Diámetro exterior de tubos	$d_0$		Común
Diámetro de las aletas	$d_2$		
Longitud de la hilera de tubos	L		
Ancho total de la hilera	b		
Separación entre hileras de tubos en sentido transversal al gas	$l_{perp}$		
Separación en paralelo	$l_{par}$	0.1	
Espesor de los tubos	$e_t$		
Espesor de las aletas	e	0.0008	
Diámetro interior de los tubos	$d_i$		

### IV.1.1 Restricciones para el diseño de la CRC

Algunos valores de los parámetros geométricos de diseño están restringidos debido a razones operativas y de facilidad de ensamble principalmente. Algunos límites operativos de estos parámetros fueron sugeridos por Deschamps (1995) y Rapún (1999), estos valores se muestran en la Tabla IV.5.

**Tabla IV.5**  
**Límites inferiores y superiores de algunos parámetros geométricos de la caldera**

<i>Parámetro geométrico</i>	<i>Límite superior</i>	<i>Límite inferior</i>
Número de hileras de tubos ( $Nt$ )	1	22
Número de aletas por unidad de longitud ( $Z$ ) ( $fms/m$ )	0	280
Diámetro de las aletas ( $d_2$ ) ( $m$ )	0.50	0.70
Longitud de la hilera de tubos ( $L$ ) ( $m$ )	10	44
Espesor de la aleta ( $e$ ) ( $m$ )	0.0008	0.0008

El valor de la separación de los tubos en dirección transversal depende del ancho de la caldera (véase la ecuación 23) y debe controlarse de manera que no sea menor que el diámetro de las aletas, puesto que en tal caso el diseño no sería válido.

$$l_{perp} = \frac{b - (Nt - 1)d_o}{Nt - 1} \quad (23)$$

Asimismo, el coeficiente de película exterior no debe ser muy grande, algunos autores (Rapún, 1999) lo limitan entre 30 y 100 W/m<sup>2</sup> K. Su valor viene determinado principalmente por el ancho de la caldera.

Debido a lo anterior, se observó que el ancho de la caldera no es un parámetro libre (su valor no se puede generar aleatoriamente durante el procedimiento de algoritmos genéticos) sino que depende de otras variables termodinámicas y de las restricciones descritas anteriormente. Por ello se planteó un procedimiento iterativo para calcular este valor de tal manera que se cumplan las restricciones establecidas.

## IV.2 Modelo matemático para el diseño de la CRC

### IV.2.1 Definición del modelo

En este apartado se describe el modelo matemático empleado para la determinación de los parámetros geométricos de una CRC de una central de CC de un nivel de presión, el desarrollo para una central de ciclo combinado de dos o tres niveles de presión es análogo.

El coeficiente UA de cada sección se obtiene a partir de la diferencia logarítmica media de temperaturas y sus balances térmicos, es decir:

$$U_{ec} A_{ec} \frac{t_3 - t_d - (t_4 - t_e)}{\ln \frac{t_3 - t_d}{t_4 - t_e}} = \dot{m}_g (h_3 - h_4) = \dot{m}_v (h_d - h_e) \quad (24)$$

$$U_{ev} A_{ev} \frac{(t_2 - t_3)}{\ln \frac{t_1 - t_b}{t_3 - t_b}} = \dot{m}_g (h_2 - h_3) = \dot{m}_v (h_b - h_d) \quad (25)$$

$$U_{sc} A_{sc} \frac{t_1 - t_a - (t_2 - t_b)}{\ln \frac{t_1 - t_a}{t_2 - t_b}} = \dot{m}_g (h_1 - h_2) = \dot{m}_v (h_a - h_b) \quad (26)$$

Este coeficiente, por otro lado, es función de la geometría de las superficies de intercambio (diámetro y longitud de tubos, número de aletas, etcétera); sin embargo, no existe una metodología específica para definir la configuración geométrica a partir de este valor. Por lo anterior, se plantea la necesidad de una herramienta con la que sea posible determinar los parámetros geométricos que proporcionen el valor del producto UA deseado. En Rapún (1999) se presenta una herramienta que consiste en un procedimiento iterativo mediante el cual se van modificando los parámetros geométricos de la caldera hasta obtener el producto UA correspondiente. No obstante, con este procedimiento se observa que en algunos casos el área de intercambio resultante es demasiado grande, lo que incrementa mucho el costo de la caldera.

Teniendo como base lo anterior, se desarrolla una metodología que permita hallar el valor de los parámetros geométricos (y por consiguiente del área) que proporcionen un valor del producto  $UA_{geom}$  muy próximo al obtenido con las ecuaciones 24 a 26 ( $UA_{term}$ ) para cada sección de la caldera. El modelo matemático empleado propone entonces minimizar la diferencia entre el valor calculado

a partir de los parámetros geométricos y el calculado a partir de las ecuaciones 24 a 26, es decir:

$$Error = \frac{UA_{geom} - UA_{term}}{UA_{term}} \quad (27)$$

La herramienta que se empleará para la optimización será nuevamente la de algoritmos genéticos (AG), que funciona estrictamente para la obtención de valores máximos, por lo que debe modificarse la ecuación que define el error, de manera que el resultado que se encuentre sea un máximo. Así pues, la ecuación del error se define de la siguiente manera:

$$f_e = \frac{1}{1 - Error} = \frac{1}{1 - \frac{UA_{geom} - UA_{term}}{UA_{term}}} \quad (28)$$

En esta ecuación puede observarse que en la medida en la que el error disminuye, el valor de  $f_e$  aumenta hasta un valor máximo conocido de 1.

#### ***IV.2.2 Modelo termodinámico para obtener el coeficiente global de transferencia de calor***

Se sabe que el coeficiente global de transferencia de calor se puede calcular a partir de la geometría de la caldera y de las condiciones de los fluidos empleando la siguiente ecuación:

$$\frac{1}{U} = \frac{1}{h_{gef}} + \frac{b}{k} + \frac{1}{h_i} \quad (29)$$

Donde  $h_{gef}$  es el coeficiente de película en el exterior de los tubos incluyendo las aletas, su valor se puede obtener utilizando la correlación propuesta por Weir (1988).

$$h_{gef} = h_g \Phi_a' \quad (30)$$

En la cual  $h_g$  se obtiene a partir de la correlación.

$$h_g = (a_g Re_g^m Pr_g^n \Phi_d^p) \frac{k_g}{D_g}; \quad (\text{con } m + p = 1) \quad (31)$$

Donde  $k_g$  es conductividad térmica del gas; el coeficiente  $a_g$  para flujo a través de un banco de tubos escalonados es 0.45 y para tubos alineados, 0.30, mientras

que los exponentes  $m$  y  $n$  tienen valores de 0.625 y  $1/3$  respectivamente (Weir, 1988; Bejan y Krauss, 2003).

En la misma ecuación (30)  $\Phi_a'$  se define en función de la efectividad de las aletas como:

$$\Phi_a' = \eta_\Phi \Phi_a + (1 - \eta_\Phi) \left( 1 - \frac{e}{p_\Phi} \right) \tag{32}$$

Donde:

$p_\Phi$  = paso de las aletas

$\eta_\Phi$  = efectividad de la aleta, que es la relación de calor disipado por la superficie de la aleta con respecto a la que se dispararía si la temperatura de la aleta fuera la misma que la de la base.

$\Phi_a$  = es la relación entre el área total del tubo y el área de un tubo liso con el mismo diámetro que el tubo aleteado, y se expresa de la siguiente manera:

$$\Phi_a = \frac{\text{Área intercambio total}}{\text{Área intercambio del tubo liso}} = \frac{\pi d_0(L - ze) + \pi \frac{d_2^2 - d_0^2}{4} 2z + \pi d_2 ze}{\pi d_0 L} \tag{33}$$

Como es de esperarse, la aleta aumenta la superficie de intercambio y, por tanto, mejora la transmisión de calor.

En la ecuación (31) se puede observar que todos los parámetros incluyen un subíndice  $g$ , lo que indica que se está trabajando con las propiedades del gas que es el fluido que pasa a través del banco de los tubos.

Por su parte, el coeficiente de película interior se determina a partir de la correlación de Dittus-Boelter (Bejan y Krauss, 2003) para establecer el número de Nusselt de un fluido que circula en un tubo circular delgado:

$$Nu = 0.023 Re^{m'} Pr^{n'} \tag{34}$$

En esta correlación, dadas las condiciones del fluido en el interior de los tubos, se considera que  $m'=0.8$ ,  $n'=0.4$  y  $a_f = 0.0243$  (Bejan, 2003). Siendo el número de Nusselt :

$$Nu = \frac{h_i D_i}{k_i} \tag{35}$$

En la ecuación (29) la resistencia ofrecida a la transmisión de calor por parte del material en los tubos de la caldera es siempre despreciable frente a la resis-

tencia que existe al paso del calor entre las caras del tubo y los fluidos que las rodean. Por otro lado, el coeficiente de película de los líquidos es muy grande con respecto al de los gases; por tanto, se puede decir que  $U$  es muy próximo a  $h_e$  para el economizador y evaporador y muy próximo a  $\frac{h_i h_e}{h_i + h_e}$  para el sobrecalentador.

En el siguiente apartado se aplican las ecuaciones de transferencia de calor para obtener el coeficiente global de transferencia de calor a partir de los parámetros geométricos que determinan el diseño de la caldera aplicando la ecuación (29).<sup>3</sup>

### **IV.3 Herramienta para el cálculo de las secciones de la CRC**

Como ya se mencionó, para realizar el cálculo de las áreas de la caldera se elaboró una herramienta informática que también aplica la metodología de algoritmos genéticos para minimizar el error entre el valor del producto  $UA$  obtenido a partir de los parámetros geométricos y el obtenido a partir de los parámetros térmicos. Se decidió aplicar esta metodología porque la cantidad de variables que se manejan en este problema es muy grande y, además, porque esta herramienta ya ha sido aplicada para la optimización de algunos sistemas térmicos y ha dado buenos resultados<sup>4</sup> (Valdés *et al.*, 2003; Attala *et al.*, 2001 y Toffolo y Lazzaretto, 2002). El diagrama de flujo del programa de cálculo se presenta al final del capítulo.

Al aplicar la metodología de AG es muy importante definir cuidadosamente las variables de diseño y las restricciones.

#### ***IV.3.1 Variables de diseño***

En la Tabla IV.6 se muestran las variables de diseño y se indica a la longitud como un parámetro común a todas las secciones, mientras que los otros serán independientes, esto hace que sea un total de 16 variables independientes, para una planta de CC de un nivel de presión.

<sup>3</sup> La relación entre los parámetros geométricos y el coeficiente global de transferencia de calor se explica más ampliamente en Weir (1988) y Rapún (1999).

<sup>4</sup> En Valdés *et al.* (2003), que es antecedente del presente trabajo, se aplica la herramienta de algoritmos genéticos para hallar los parámetros termodinámicos que proporcionen un diseño con un costo mínimo. Se parte de los diseños optimizados mediante esta metodología para hallar la configuración geométrica con la herramienta que aquí se propone.

**Tabla IV.6**  
**Parámetros de diseño de la caldera de recuperación de calor**

Parámetro	Sección
Número de tubos en paralelo	Evaporador Economizador Sobrecalentador
Número de tubos en perpendicular	
Número de aletas	
Diámetro del tubo	
Diámetro de la aleta	
Longitud	Común a todas las secciones

Como se indicó anteriormente, el ancho de la caldera ( $b$ ) no se considera una variable independiente (por lo que no se presenta en la tabla anterior), su valor se propone a partir de los parámetros de diseño (generados por el algoritmo) aplicando la fórmula:

$$b = \frac{m_g}{L \cdot \rho_{gas} v_{gas}} \quad (36)$$

A partir de este valor se calcula la separación entre tubos ( $l_{perp}$ ) con la ecuación (23) y el coeficiente de película exterior (ecuación 30); si estos valores no cumplen con las restricciones establecidas en el apartado IV.1.1 se incrementa el valor de  $b$ , si no se cumplen nuevamente las restricciones el procedimiento se repite hasta alcanzar un valor de  $b$  adecuado. Este procedimiento se hace para cada sección de la caldera teniendo cuidado de mantener el mismo valor de  $b$  para todas las secciones.

Una vez que se ha determinado un valor de  $b$  se obtiene el área  $y$ , por tanto, el valor del producto  $UA$  para cada sección (aplicando la ecuación 29). Este valor será el producto  $UA_{geom}$  que se incluirá en la función salud definida en el subapartado siguiente.

### IV.3.2 Función objetivo del algoritmo genético (función salud)

Para trabajar con algoritmos genéticos es necesario incluir, además de la función objetivo (que debe ser una función a maximizar), unas penalizaciones para definir la función salud (que es la función objetivo en el algoritmo genético).

### *IV.3.2.1 Función objetivo*

Como ya se explicó, en el presente trabajo se pretende hallar un valor del producto  $UA_{geom}$  (obtenido a partir de la configuración geométrica) muy similar al  $UA_{term}$  (obtenido a partir de las ecuaciones 24-26), esto se logra minimizando el error, es decir hallando un valor máximo de la ecuación 28 para cada sección.

Como la CRC de una central de ciclo combinado de un nivel de presión se compone de tres secciones, entonces la función objetivo sería la suma de las funciones de error correspondientes a cada sección (el número de términos de la ecuación será el número de secciones que tenga la caldera), es decir:

$$f_{obj} = \frac{1}{1 + Error_{ec}} + \frac{1}{1 + Error_{ev}} + \frac{1}{1 + Error_{so}} \quad (37)$$

Se eligió esta función objetivo porque, como puede observarse, al disminuir el error esta función adquiere un valor máximo igual a tres, que al ser conocido proporciona una forma de evaluar la convergencia del algoritmo.

### *IV.3.2.2 Penalizaciones*

Deben seleccionarse cuidadosamente debido a que son muy importantes para restringir el dominio de la solución. En el problema que aquí se describe se consideraron las siguientes:

- *Penalización por incremento del área de cada sección.* Sirve para hallar una configuración que además proporcione un área de intercambio mínima.
- *Penalización por la pérdida de carga del lado del gas.* Si la pérdida de carga del lado del gas disminuye mucho, aumenta demasiado el área de la caldera. Algunos autores (Kakaç, 1991) sugieren que este valor se encuentra entre 10 y 15 mbar.
- *Penalización por la pérdida de carga del lado del vapor en el sobrecalentador.* Si la pérdida de carga de lado del vapor aumenta mucho será necesario consumir una mayor potencia en bombas y extractores para compensar esta pérdida.
- *Penalización por el incremento de la velocidad del vapor a la salida de la caldera.* La velocidad de vapor muy alta implica más pérdida de carga del lado del vapor. Se sugiere que se encuentre entre los 10 y 50 m/s (Dechamps, 1995).

Con la función objetivo y las penalizaciones establecidas se define la función salud que se evaluará durante el proceso de optimización con algoritmos genéticos.

$$f_{salud} = f_{obj} - (K_{aec} Pen A_{ec} + K_{aev} Pen A_{ev} + K_{aso} Pen A_{so}) - K_{\Delta P} Pend\Delta p - K_{v0} Pen v_v \quad (38)$$

El valor de las constantes de la ecuación (38) debe ser lo suficientemente grande como para eliminar a aquellos individuos que no cumplan con las restricciones establecidas; es decir, del mismo orden de magnitud que la función objetivo.

#### IV.4 Aplicación del algoritmo para el cálculo geométrico de una caldera de dos niveles de presión con recalentamiento

En este apartado se muestra la aplicación del algoritmo para la optimización termoeconómica y del diseño geométrico de una caldera de recuperación de calor acoplada a un ciclo combinado de 2 niveles de presión con recalentamiento.

Primeramente se realizó la optimización termoeconómica de la configuración analizada, aplicando la metodología descrita en la sección III y, posteriormente, el cálculo del área de intercambio de la CRC. Los datos de entrada del ciclo de gas, así como los resultados de la simulación del mismo, se muestran en la Tabla IV.7. Los resultados de la optimización termoeconómica de esta configuración se muestran en la Tabla IV.8.

**Tabla IV.7**  
**Datos de diseño del ciclo de gas**

<i>Datos de diseño</i>	
Caudal de gases en el escape (kg/s)	506.9
Pérdida de carga en la admisión (mbar)	10
Relación de compresión del compresor	12
Rendimiento isentrópico del compresor	0.85
Rendimiento de la cámara de combustión	0.95
Pérdida de carga en la cámara de combustión	4%
Temperatura de entrada a la turbina de gas (K)	1368.15
Rendimiento isentrópico de la turbina de gas	0.91
Pérdida de carga en el escape (mbar)	25
RESULTADOS DE LA SIMULACIÓN	
Potencia de la turbina (kW)	166,565
Rendimiento	0.344
Temperatura de escape	820.52 K

**Tabla IV.8**  
**Resultados de la optimización termoeconómica de un ciclo combinado de dos niveles de presión con recalentamiento**

<i>Variables de diseño</i>	<i>Valor</i>
Rendimiento	0.55
Potencia TV (kW)	88234.06
Potencia CC (kW)	255130.20
Flujo másico de alta (kg/s)	49.06
Flujo másico de baja (kg/s)	20.8
Presión de baja (bar)	3.8
PP de baja (K)	6.7
AP de baja (K)	3.2
DT de baja (K)	23.1
Presión de alta (bar)	65.9
PP de alta (K)	14.4
AP de alta (K)	5.2
DT de alta (K)	25.9

Una vez que se realizó la optimización termoeconómica de esta planta, se aplicó la herramienta para obtener la configuración geométrica; sin embargo, en este caso la función objetivo de la CRC tiene siete secciones de intercambio en lugar de tres, como sucede con el ciclo de un nivel de presión. De esta manera, la función objetivo descrita mediante la ecuación 37 se transforma en la siguiente ecuación para un ciclo de dos niveles de presión con recalentamiento:

$$\begin{aligned}
 f_{obj} = & \frac{1}{1 + Error_{ecb}} + \frac{1}{1 + Error_{ecla}} + \frac{1}{1 + Error_{evb}} + \\
 & \frac{1}{1 + Error_{ec2a}} + \frac{1}{1 + Error_{sob}} + \frac{1}{1 + Error_{eva}} + \frac{1}{1 + Error_{rec}} + \frac{1}{1 + Error_{so}}
 \end{aligned}
 \tag{39}$$

Los resultados del cálculo de áreas de la CRC correspondiente a esta configuración se muestran en la Tabla IV.9. Como es de esperarse, el criterio de convergencia sería el valor máximo del error que puede alcanzar la ecuación (39), que es de 8.0, y se observa en la Tabla IV.9 que el error alcanzado con el algoritmo es de 7.71.

**Tabla IV.9**  
**Reparto de áreas de una CRC de dos niveles de presión con recalentamiento**

<i>Parámetros geométricos</i>	$EC_B$	$EC_{1A}$	$EV_B$	$EC_{2A}$	$SO_B$	$EV_A$	$RE$	$SO_A$
Número de tubos	19	19	20	18	15	20	19	20
Número de filas	19	19	20	15	15	19	18	19
Número de aletas	139	187	268	209	126	229	200	253
Diámetro exterior de tubos (m)	0.062	0.063	0.060	0.051	0.063	0.065	0.067	0.065
Diámetro exterior de aletas (m)	0.078	0.085	0.084	0.064	0.071	0.088	0.05	0.085
Área de la sección (m <sup>2</sup> )	5900.32	8336.82	16110.1	13345.0	2225.07	13345.1	7828.80	7646.1
Porcentaje de cada sección	7.9%	11.15%	21.55%	17.85%	2.97%	17.85%	10.47%	10.23%
Longitud de los tubos (m)	24.4							
Ancho de la caldera (m)	6.34							
Pérdida de carga del lado del gas (mbar)	21.7							
Pérdida de carga del vapor (bar)	6.82							
Área total de intercambio (m <sup>2</sup> )	74737.3							
<i>Error total</i>	7.71							
<i>Función salud</i>	7.06							

En la tabla anterior se observa un valor de la longitud de los tubos y del ancho de la caldera común a todas las secciones, esto es muy deseable por cuestiones de ensamble y fabricación. Asimismo, se observa un valor de las pérdidas de carga que se encuentra dentro de los límites establecidos.

Por otro lado, se aprecia que tanto los resultados obtenidos como la convergencia del método son buenos, inclusive en el caso de la configuración con dos niveles de presión con recalentamiento, aun cuando se tienen 35 variables independientes. También hay que señalar que conforme la complejidad de la caldera aumenta, es más difícil encontrar una solución con un error casi nulo y además con área mínima.

## CONCLUSIONES Y COMENTARIOS FINALES

En diversos problemas de ingeniería se presentan problemas no lineales que involucran funciones trascendentes; sin embargo, en problemas de termodinámica y transferencia de calor, la situación es aún más complicada puesto que la obtención de las propiedades de los fluidos generalmente se realiza con correlaciones y/o tablas de datos. Esto complica mucho más la optimización y el análisis de los sistemas térmicos como las calderas de recuperación de calor.

Así pues, un algoritmo genético, cuya programación es muy sencilla, puede aplicarse para la optimización de distintos tipos de sistemas térmicos, principalmente en aquellos en los que no existe un criterio específico de diseño, sino que más bien se trata de seleccionar reiteradamente los parámetros hasta que se consiga un resultado satisfactorio. Asimismo, al poder incluir penalizaciones, se puede encontrar una solución óptima, restringiendo algunas variables de interés dentro de límites establecidos. Por las razones expuestas se decidió presentar en este capítulo un ejemplo de la aplicación de este tipo de metodologías.

Una cuestión importante que debe señalarse es la posibilidad de aplicación de esta metodología a otro tipo de sistemas de transferencia de calor, como pueden ser calderas con combustión o intercambiadores de calor. Esto es importante ya que el diseño de este tipo de sistemas siempre implica un procedimiento de ensayo y error que generalmente lleva a un diseño ajustado a los parámetros termodinámicos, pero no óptimo en cuanto a la superficie de intercambio. Con la metodología propuesta y aplicando los algoritmos genéticos se puede obtener un diseño ajustado pero con una superficie de intercambio de tamaño menor, optimizada termoeconómicamente.

## REFERENCIAS

- Attala L., B. Facchini y G. Ferrara (2001). "Thermoeconomic Optimization Method as Design Tool in Gas-Steam Combined Plant Realization", *Energy Conversion and Management*, 42: 2163-2172.
- Bejan, A. y Krauss A. (eds.) (2003). *Heat Transfer Handbook*, Estados Unidos, Wiley.
- Belew R. K. y L. B. Booker (eds.) (1991). *Proceedings of the Fourth International Joint Conf. on Genetic Algorithms*. Morgan Kaufmann Publishers.
- Coello, C. y A. Chistiansen (2000). "Multiobjective Optimization of Trusses Using Genetic Algorithm", *Computer and Structures*, vol. 75, 647-660.

- Dechamps, P. J. (1995). "Incremental Cost Optimization of Heat Recovery Steam Generators", *ASME COGEN-TURBO*, Viena, Austria.
- De Jong A. K. (1975). "An análisis of the Behavior of Class of Genetic Adaptive Systems". (Doctoral dissertation University of Michigan). *Dissertation Abstracts International*, Vol. 36, No. 10.
- Durán M., M. Valdés y A. Rovira (2003). "Aplicación de los algoritmos genéticos para la optimización termoeconómica de centrales de ciclo combinado", *Proc. 3as. Jornadas de Ingeniería Termodinámica*, Valencia, España.
- Durán M. (2004). "Estudio de calderas de recuperación de calor de ciclos combinados de turbinas de gas y vapor empleando la técnica de algoritmos genéticos", Tesis doctoral, ETSII, UPM.
- Franco A. y A. Russo (2002). "Combined Cycle Plant Efficiency Increase Based on the Optimisation of the Heat Recovery Steam Generator Operating Parameters", *International Journal of Thermal Sciences* 41: 843-859.
- Franco A. y N. Giannini (2006). "A General Method for the Optimum Design of Heat Recovery Steam Generators", *Energy* 31(15): 3342-3361.
- Goldberg D. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Michigan.
- Gunn y Horton (1989). *Industrial Boilers*, Longman Scientific and Technical.
- Herrera F. y Verdegay J. (1996). *Genetic Algorithms and Soft Computing*; Physica-Verlag, Michigan.
- Horlock J. H. (1991). "The Use of Feed Heating in the Steam Cycle of Combined Cycle Power Plant", *Proc. Instn. Mech. Engrs*, 205: NA3.
- Kakaç S. (1991). *Boilers, Evaporators and Condensers*, Wiley Interscience, Miami, Florida.
- Kelnofer R. (Ed.) (1a. ed. E. U, 1999). *Combined-Cycle Gas and Steam Power Plants*. The Fairmont Press.
- Martín del Campo C., J. L. François y A. López (2001). "AXIAL, A System for Boiling Water Reactor Fuel Assembly Axial Optimization Using Genetic Algorithms", *Annals of Nuclear Energy*, 28: 1667-1682.
- Martorell S., A. Sánchez, S. Carlos, V. Serradel (2002). "Simultaneous and Multi-criteria Optimisation of TS Requirements and Maintenance at NPPs", *Annals of Nuclear Energy*, 29: 147-168.
- Mohagheghi M. y J. Shayegan (2009). "Thermodynamic Optimization of Design Variables and Heat Exchangers Layout in HRSGs for CCGT, Using Genetic Algorithm", *Applied Thermal Engineering*, 29 (2-3); 290-299.
- Rapún J. L. (1999). "Modelo matemático del comportamiento de ciclos combinados de turbinas de gas y vapor", Tesis doctoral, ETSII-UPM.

- Rovira A. (2004). “Desarrollo de un modelo para la caracterización termoeconómica de ciclos combinados de turbinas de gas y de vapor en condiciones de carga variable”. Tesis Doctoral, ETSII, UPM.
- Staff Report (2000). “Use of the HXSDN Program to Size and Estimate the Cost of a Heat Recovery Steam Generator”, *Technical Report EPI TB-102*, Energy Planners International. Recuperado de <www.energyplan.com>.
- Stultz S. C. y J. B. Kitlo (1994). “Steam: its Generation and Use”. Babcock and Wilcox Co., 40a. ed.
- Toffolo A. y Lazzareto A. (2002). “Evolutionary Algorithms or Multi-Objective Energetic and Economic Optimization in Thermal System Design”, *Energy* 27: 549-567.
- Tsatsaronis G. (1993). “Thermoeconomic Analysis and Optimization of Energy Systems”. *Prog. Energy Combust. Sci.* 19: 227-257.
- Tsatsaronis G. y Pisa J. (1994). “Exergoeconomic Evaluation and Optimization of Energy Systems – application to the CGAM problem”, *Original Research Article Energy*, 19 (3): 287-321.
- Valdés M., M. Durán y A. Rovira (2003). “Thermoeconomic Optimization of Combined Cycle Gas Turbine Using Genetic Algorithms, Applied Thermal Engineering”, *Applied Thermal Engineering* 23: 2169-2182.
- Valdés M., A. Rovira, M. Durán (2004). “Influence of the Heat Recovery Steam Generator Design Parameters on the Thermoeconomic Performances of Combined Cycle Gas Turbine Power Plants”, *International Journal of Energy Research* 28 (14): 1243-1254.
- Valero A., M. Lozano, L. Serra, G. Isatsoranis, J. Pisa, C. Frangopoulos, M. Van Spakovsky (1994). “CGAM Problem: Definition and Conventional Solution”. *Energy* 19 (13): 279-286.
- Weir C. D. (1988). “Estimating the Performance of Gas Turbine Heat-Recovery Boilers of Design”, *Proc. Instn. Mech. Engrs.* 202 (A4).

*Procesos no lineales en la ciencia y la sociedad* se terminó de imprimir el 12 de octubre de 2013, en los talleres de Ediciones Verbolibre, S.A. de C.V., 1o. de mayo núm. 161-A, Col. Santa Anita, Deleg. Iztacalco, México, D.F., C.P. 08300. Tel.: 3182-0035. <edicionesverbolibre@gmail.com>. La edición consta de 500 ejemplares.