

Tribunal nombrado por el Mgfco. y Excmo. Sr. Rector de la
Universidad de Castilla la Mancha, el día de de
2006

Presidente D.

Vocal D.

Vocal D.

Vocal D.

Secretario D.

Realizado el acto de defensa y lectura de la Tesis el día
de de 2006 en

Calificación:

El Presidente

El Secretario

Los Vocales

Índice General

Resumen	13
1 Antecedentes	17
1.1 Introducción	17
1.2 Problema de asignación de tráfico (TAP)	22
1.2.1 Modelos estáticos de asignación de tráfico	24
1.2.2 Modelos dinámicos de asignación de tráfico	28
1.2.3 Formulación matemática del problema estático de asignación de tráfico	31
1.2.4 Algoritmos para la resolución del problema de asignación de tráfico	38
1.3 Problema de estimación de la matriz O-D	44
1.3.1 Formulación matemática de los modelos basados en programación matemática.	50
1.3.2 Formulación matemática de los modelos basados en inferencia estadística	55
1.3.3 Algoritmos para la resolución del problema de estimación de la matriz O-D	61
1.4 Sumario de la tesis	73
2 Algoritmo de generación de columnas para la estimación de matrices de viajes Origen-Destino en redes de tráfico congestionadas	75
2.2 Revisión del DAP	78
2.3 Algoritmo de generación de columnas (CGA)	80
2.4 Análisis de la convergencia del CGA	83
2.5 Experiencia numérica	86

2.5.1	Experimento 1	89
2.5.2	Experimento 2	89
2.5.3	Experimento 3	91
2.6	Conclusiones	93
3	Algoritmos heurísticos de generación de columnas para la estimación de matrices de viajes O-D en redes congestionadas	101
3.1	Introducción	103
3.2	Algoritmos heurísticos de generación de columnas	104
3.2.1	Algoritmo de generación de columnas para el DAP	105
3.2.2	Algoritmos heurísticos de generación de columnas para el DAP	107
3.2.3	Algunos ejemplos	108
3.3	Experiencia numérica	115
3.3.1	Experimento 1: evaluación de modelos DAP	116
3.3.2	Experimento 2: evaluación de algoritmos CGA heurísticos	118
3.4	Conclusiones	121
4	Ajuste de las funciones de coste en los arcos en los modelos de asignación de tráfico en equilibrio	129
4.1	Introducción	131
4.2	Modelos para el PCPCA	135
4.2.1	Notación e hipótesis empleadas	136
4.2.2	PCPCA(ρ)	137
4.2.3	PCPCA(P)	140
4.2.4	Análisis de los Modelos	141
4.3	Un algoritmo de generación de columnas para el PCPCA	143
4.3.1	El algoritmo CGA	143
4.4	Experimentos numéricos	148
4.4.1	Experimento 1: Estudio de los modelos PCPCA	149
4.4.2	Experimento 2: estudio del algoritmo CGA	153
4.5	Conclusiones	162

5	Conclusiones y sugerencias para futuras líneas de investigación	171
5.1	Conclusiones	171
5.2	Novedad y relevancia	174
5.3	Futuras líneas de investigación	175

Índice de Figuras

2.1	Topología de la red de Nguyen-Dupuis	88
2.2	Topología de la red de Friesz-Harker	88
2.3	Curvas de nivel del experimento 2	90
2.4	Experimento 2: Mínimo local	91
2.5	Experimento 3: Curvas de nivel	92
2.6	Experimento 3: Curvas de nivel	93
2.7	Experimento 3: Curvas de nivel	93
2.8	Experimento 3: Curvas de nivel	94
3.1	Topología de la red de Nguyen-Dupuis	119
3.2	Topología de la red de Friesz-Harker	120
4.1	Topología de la red de Nguyen-Dupuis	150
4.2	Desarrollo del experimento 1	150
4.3	PCPCA(ρ) curvas nivel TAP-DUE	152
4.4	PCPCA(ρ) curvas nivel TAP-SE	153
4.5	PCPCA(ρ) curvas nivel TAP-CC	153
4.6	PCPCA(ρ) curvas nivel TAP-SUE $\frac{1}{\theta} = 25$	154
4.7	PCPCA(ρ) curvas nivel TAP-SUE $\frac{1}{\theta} = 50$	154
4.8	Red de Sioux Falls	155
4.10	Problema 1: Tiempo en los pares O-D	158
4.11	Problema 1: Flujo en los arcos	158
4.12	Problema 2: Tiempo en los pares O-D	158
4.13	Problema 2: Flujo en los arcos	159

4.14 Problema 3: Tiempo en los pares O-D	159
4.15 Problema 3: Flujo en los arcos	159
4.16 Problema 4: Tiempo en los pares O-D	160
4.17 Problema 4: Flujo en los arcos	160
4.18 Red de Ciudad Real	161

Índice de Tablas

1.1	Taxonomía de los modelos de asignación de tráfico a redes urbanas . . .	25
1.2	Características de los algoritmos simpliciales	43
2.1	Notación empleada	95
2.2	Algoritmo de generación de columnas (CGA)	96
2.3	Definición de los CGAs truncados	96
2.4	Definición de los experimentos	97
2.5	Resultados numéricos del experimento 1	98
2.6	Resultados numéricos del experimento 2	99
2.7	Resultados numéricos del experimento 3	99
3.1	Notación empleada	122
3.2	Algoritmo de generación de columnas (CGA)	123
3.3	Coefficientes de escalamiento para los DAPs	123
3.4	Resultados agregados	123
3.5	Rangos y significatividad para el test de Friedman para la comparación de los DAPs	124
3.6	Prueba de los rangos con signo de Wilcoxon	124
3.7	Descripción de los DAPs empleados en el experimento 2	124
3.8	Diferencia entre las sucesiones generadas por los algoritmos SAB y el	
3.9	Resultados test 1 para la estimación mínimo error absoluto	125
3.10	Resultados test 1 para estimación mínimo cuadrática	126
3.11	Resultados test 2 para estimación mínimo error absoluto	126
3.12	Resultados del test 2 para estimación mínimo cuadrática	127

4.1	Notación utilizada para formular el TAP	165
4.2	El algoritmo de generación de columnas (CGA)	166
4.3	Predicciones 2g Modelo PCPCA(ρ) con coste en los pares O-D	166
4.4	Predicciones 2g Modelo PCPCA(ρ) con coste en la red	167
4.5	Predicciones 2g Modelo PCPCA(P) con coste en los pares y en la red	167
4.6	PCPCA(P): Predicción de los parámetros coste arcos con tiempos medios en los pares O-D	168
4.7	PCPCA(P): Predicción de los parámetros coste arcos con tiempos medios en la red	168
4.8	Definición de los problemas	168
4.9	Resultados obtenidos	169
4.10	Resultados computacionales	169
4.11	Resultados computacionales para la red de Ciudad Real	169

Resumen

El desarrollo socio-económico que han experimentado las áreas urbanas en los últimos años, tanto en los países desarrollados como en los que están en vías de desarrollo, ha traído como consecuencia un incesante incremento de la movilidad y demanda de viajes entre las diferentes zonas de las ciudades. Dicha demanda de viajes se ha traducido en constantes aumentos en la utilización masiva del vehículo privado, produciéndose los típicos problemas de congestión y contaminación atmosférica por todos conocidos.

Es en este contexto donde adquiere una especial relevancia la *planificación del transporte urbano*. Ésta tiene como objetivo predecir y controlar la evolución del equilibrio entre la oferta y la demanda de servicios de transporte, de modo que el bienestar de la sociedad sea maximizado, y teniendo en cuenta los intereses individuales de los usuarios.

Durante décadas se ha realizado un gran esfuerzo de investigación para desarrollar modelos combinados de asignación en equilibrio que recogiesen el mayor número de aspectos relevantes en el comportamiento de los usuarios y de la red de transporte (elección de modo de transporte, ruta, destino, realización del viaje, múltiple tipos de usuarios, percepción aleatoria de los costes de transportes, etc.) Este esfuerzo ha ido acompañado del desarrollo de algoritmos de resolución que permitían estimar ciertas predicciones (outputs) en función de la parametrización (input) del modelo. En la actualidad, el problema que esta adquiriendo importancia, es el problema inverso. En él, se desea calcular una parte de la parametrización del modelo a partir de un conjunto de observaciones de los distintas predicciones del modelo de equilibrio.

Sin embargo, aún no existe una teoría unificada para el problema inverso de asignación en equilibrio en redes, y en esta dirección, se ha creído que se deba de partir, al igual que se hizo históricamente con el desarrollo de modelos combinados de asignación en equilibrio, de un modelo básico de asignación de tráfico. Para determinar los parámetros de los modelos de asignación de tráfico en equilibrio, los modelos matemáticos tradicionales realizaban de grandes encuestas de movilidad, cuyo coste económico y humano es muy elevado. Recientemente se han desarrollados nuevas metodologías para la estimación de los citados parámetros, basadas en otras fuentes de información, como son las mediciones de los volúmenes de tráfico en los arcos de la red y el uso de parámetros desactualizados o estimados por otro procedimiento. Estos modelos se formulan mediante la denominada *programación matemática binivel generalizada* o también llamada

programación matemática con restricciones de equilibrio (MPEC). Estos nuevos modelos presentan dificultades en su aplicación a los problemas reales debido, fundamentalmente a:

- (i) Las grandes dimensiones de los problemas reales, del orden de varias decenas de miles de variables en el nivel superior y varios miles de variables en el nivel inferior;
- (ii) Las malas propiedades matemáticas de los modelos, como la no diferenciabilidad y la no convexidad.

Existen múltiples resultados teóricos sobre programación matemática con restricciones de equilibrio, pero son muy pocos los que tienen una repercusión directa en modelos binivel aplicados en la planificación del transporte urbano. Para que estos resultados generales sean de utilidad, y aplicables para el desarrollo de algoritmos de optimización, requieren de una alta especialización.

Por otra parte, la programación binivel es un tema de gran actualidad, frontera del conocimiento en Optimización, con gran dificultad intrínseca, que concita el interés de investigadores tanto teóricos como aplicados.

En este contexto, la finalidad de esta tesis doctoral es el estudio de las propiedades matemáticas de los modelos binivel, orientado a la elaboración de algoritmos eficientes (Incluyendo su diseño e implementación computacional) para su aplicación a los siguientes problemas:

- o *Estimación de matrices origen-destino para la demanda de tráfico (DAP)*. Este dato es fundamental para definir el patrón de movilidad. Se intenta obtener estas matrices a partir de observaciones de flujo en los arcos de la red (aforos) y de otras matrices desactualizadas, obtenidas en estudios anteriores de movilidad o mediante la realización de encuestas domiciliarias. Esta última información no es imprescindible.
- o *Determinación de los parámetros del modelo de oferta en el modelo de asignación de tráfico en equilibrio (PCPCA)*. La modelización de la red de transporte requiere de las denominadas funciones de costes en los arcos, definidas tradicionalmente por los parámetros de tiempo de viaje en el arco vacío y parámetros adicionales que reflejan el efecto de la congestión. Usualmente este ajuste es manual y basado en la experiencia de los planificadores. Aquí elaboramos un nuevo modelo binivel para este tema y su resolución a partir, fundamentalmente, de aforos

Ambos problemas pueden abordarse simultáneamente, pero el hecho de necesitar entender el problema, ha conducido a estudiarlos separadamente.

La tesis doctoral consta de los siguiente capítulos. En el capítulo 1 se ha realizado una revisión del *estado-del-arte*, haciendo especial hincapié en el problema de asignación

de tráfico en redes congestionadas (TAP) y en el problema de estimación de las matrices de viaje origen-destino en redes de tráfico congestionadas.

En el capítulo 2 se ha centrado en el desarrollo un algoritmo exacto (CGA) para resolver el DAP. El CGA se fundamenta en: i) una reformulación del DAP en un problema de un solo nivel, en el espacio de flujo en los caminos y ii) en un esquema de generación de columnas desarrollado en un contexto binivel, con el objeto de evitar la enumeración explícita de los caminos de la red de tráfico. Se ha realizado un análisis de sensibilidad del TAP, bajo cierta condición de regularidad, habiendo sido empleados los resultados para establecer la convergencia del algoritmo a óptimos locales, bajo hipótesis relativas a la continuidad de las funciones implicadas.

En el capítulo 3 se ha desarrollado un marco heurístico para la aplicación del CGA a problemas reales de grandes dimensiones basado en: i) linealización de las condiciones de equilibrio, ii) estrategias de eliminación de columnas (caminos) y iii) métricas para el DAP conducentes a funciones objetivos lineales para el DAP. Dentro de este marco están los algoritmos heurísticos clásicos SAB e IEA para el DAP. Esta nueva visión de estos algoritmos presenta fundamentalmente dos ventajas: i) permite una implementación computacional más eficiente y ii) permite el establecimiento de condiciones para la convergencia a puntos estacionarios para el SAB. Completamos este estudio con unas pruebas computacionales que evalúan el rendimiento de los algoritmos propuestos, testándose varias propuestas de modelización para el DAP que conducen a la estimación por mínimos cuadrados generalizados, minimización del error absoluto y el método MiniMax.

En el capítulo 4 se aborda el problema de ajustar/estimar los parámetros de las funciones de coste en los arcos, para los modelos estáticos de asignación de tráfico, a partir de aforos y tiempo de viajes en los pares O-D. Se desarrollan varios modelos matemáticos basados en la teoría de optimización binivel, poniéndose de manifiesto que si se usa la información recolectada en único periodo (para aforos y tiempo de viaje en los pares de demanda origen-destino) el problema de ajuste tienen infinitas soluciones, ya que solo se puede estimar el tiempo de congestión en el arco. Esta deficiencia requiere disponer de observaciones efectuadas en más de un periodo. Estos modelos presentan gran complejidad en su resolución debido, entre otros motivos, a sus grandes dimensiones. Se propone un algoritmo de generación de columnas para dicho problema que se ha probado que es convergente. Este algoritmo es la adaptación del CGA desarrollado para el DAP (modelo binivel en el que las variables del nivel superior parametrizan las restricciones del nivel inferior) al PCPCA en el que las variables del nivel superior parametrizan la función objetivo del nivel inferior. Se han desarrollado las propiedades de los modelos, tanto en problemas reales como en redes simuladas.

El trabajo se termina realizando una exposición de las conclusiones obtenidas y presentando posibles líneas de futuras investigaciones.

Por otra parte es necesario hacer notar que se ha intentado que los capítulos principales de esta tesis sean autocontenido, pues se han redactado con la intención de

publicarlos en revistas especializadas. De hecho, ya han sido enviados los capítulos 2 y 3.

Finalmente, es necesario hacer notar que los mayores beneficios para la sociedad, que esta Tesis pueda reportar, se obtendrían desde el punto de vista tecnológico: Cada uno de los modelos aquí planteados es de sumo interés para los planificadores de transporte y el hecho de que puedan ser abordados mediante la misma metodología y de que exista la posibilidad de desarrollar un paquete de métodos computacionales que resuelvan estos modelos, le confieren un gran potencial práctico.

Palabras claves: Estimación de matrices origen-destino, programación matemática con restricciones de equilibrio, métodos de generación de columnas, optimización de gran escala, optimización no diferenciable, problema de asignación de tráfico, estimación de parámetros de modelos de equilibrio, métodos heurísticos, problemas de flujo en redes, desigualdades variacionales.

Capítulo 1

Antecedentes

1.1 Introducción

El tráfico urbano es uno de los problemas que más influyen en la calidad de vida de los residentes en las ciudades y áreas metropolitanas de los países desarrollados. Un problema acentuado, en las últimas décadas, por el desarrollo de dos fenómenos simultáneos. El primero, una creciente movilidad que ha potenciado el uso del automóvil privado en detrimento del transporte público. El segundo, la generalización del modelo de ciudad difusa, una tipología urbana genuinamente norteamericana pero imitada en Europa por las grandes ciudades y las ciudades de tamaño medio, que han alcanzado su desarrollo actual tras un proceso descentralizador (Seguí and Martínez (2004)).

Este nuevo y generalizado modelo urbano acentúa la relocalización de la población, el comercio y los servicios hacia la periferia próxima, en tanto que relega la actividad industrial al espacio exterior a la urbe y en la que el centro tradicional queda reducido a una función simbólica. Una ciudad que, como sostiene Rueda (1997), *"tiene de todo y mucho pero disperso, separado funcionalmente"* y, por tanto, presenta ciertos retos en términos de accesibilidad sólo superables por una movilidad motorizada, capaz de abarcar progresivamente mayores distancias.

Asistimos, pues, a un proceso expansivo en lo urbano y en la movilidad sobre la base de un creciente número de desplazamientos y el incremento de la amplitud éstos. Ambos extremos redundan, necesariamente, en la propia expansión de las infraestructuras viarias. Se cierra así un círculo vicioso en el que la persona ha ido perdiendo protago-urbanos y la distribución modal de los desplazamientos motorizados, son buenos indicadores de la situación descrita.

No es extraño, pues, que el transporte urbano e interurbano por carretera sea, en la actualidad, el más importante de los modos de transporte de pasajeros en términos de demanda. En Europa, por ejemplo, las redes de tráfico urbanas e interurbanas canalizan el noventa por ciento de los pasajeros y el setenta por ciento de las mercancías. Ello

supone, asimismo, un liderazgo económico en términos de inversiones y volumen de negocio generado, pero también un liderazgo tecnológico que se traduce en el volumen de proyectos I+D generados.

El resultado del proceso expansivo descrito ha colocado al medio urbano en una situación difícil en términos de sostenibilidad, en donde la polución, la congestión del tráfico y los accidentes constituyen externalidades negativas de fuerte impacto en la salud y la economía de los ciudadanos. Así, la congestión del tráfico viene provocando importantes pérdidas de tiempo en los desplazamientos privados y el incumplimiento de horarios en el transporte público. Este incumplimiento en el servicio público redundará, en último término, en la pérdida de atractivo para la demanda y la consecuente dificultad de mantener una oferta cuyo coste económico y energético llega a resultar insostenible. Por su parte, la pérdida de tiempo en los desplazamientos de los usuarios tiene una transcendencia económica nada desdeñable, como lo demuestran los diversos estudios realizados. Por citar sólo un ejemplo paradigmático, en Tokio se calcula que la congestión del tráfico causa pérdidas del orden de los 50 millones de euros/día (System (2001))

En lo que se refiere a las externalidades medioambientales, la situación es asimismo preocupante. El transporte es, en este ámbito, el responsable de más del sesenta por ciento de las emisiones de monóxido de carbono, del cincuenta por ciento de los óxidos de nitrógeno y del treinta y tres por ciento de los hidrocarburos, emisiones que no sólo inciden en el ámbito local -provocando que entre un setenta y un ochenta por ciento de las ciudades europeas con una población superior a los 500.000 habitantes no reúnan hoy las condiciones mínimas de calidad atmosférica estipuladas por la Organización Mundial de la Salud-, sino que contribuyen también a la polución a escala regional y al efecto invernadero¹. Además, las consecuencias de la polución en la salud de los ciudadanos europeos suponen un coste de medio punto del PIB de la Unión Europea. Por otra parte, en torno al veinte por ciento de esos mismos ciudadanos padecen niveles inaceptables de ruido debido al tráfico. Ello suma, en el debe de la UE, otro tercio de punto de su PIB en concepto de gastos indirectos del transporte².

Todos estos problemas aconsejan el desarrollo de nuevas estrategias globales para un transporte urbano sostenible. Estrategias que contemplen no sólo una amplia gama de medidas paliativas, si no, también, el uso de tecnologías e infraestructuras innovadoras, aspecto éste que constituye una de las características más relevantes de los transportes. El concepto de un transporte eficiente, eficaz y seguro se impone hoy en las nuevas políticas de transporte.

...Especialmente; implementación de zonas de bajas emisiones, implementación de 'puntos de acceso' en determinadas áreas urbanas, etc. Entre las tecnologías e infraestructuras innovadoras, hay que señalar todo un conjunto de recursos que posibilitan rutas y modos de

¹Agencia Europea del Medioambiente, 2001

²Background document to Emerging Thematic Priorities for Research in Europe Scoping Document, 2000

transporte que contribuyen a reducir el tiempo y el recorrido de los desplazamientos al tiempo que procuran mayor fluidez de tráfico, mejor nivel de servicio de las vías y la racionalización del consumo y de las emisiones. Se trata, pues, de una serie de cambios en la concepción del desarrollo del transporte, de carácter tecnológico y que, al no implicar restricciones en la movilidad de la población, son generalmente bien aceptados por ésta.

Es en este contexto donde adquiere relevancia la *planificación del transporte* cuyo objetivo es responder adecuadamente a las necesidades de movilidad de la población, por medio del dimensionamiento de la infraestructura y de la utilización de los medios de transporte, estableciendo los costes de la solución adoptada, las fuentes de los recursos necesarios, y los plazos respectivos de las etapas propuestas.

En general, el transporte es una actividad derivada de otras actividades que tienen lugar en un área geográfica determinada, sea esta un país, una región, una ciudad, un distrito o un barrio (Manheim (1984)). El conjunto de actividades de residencia, producción, consumo, ocio, etc. que se produce en un lugar, denominado *sistema de actividades*, es el que genera ciertas necesidades de viajes o demanda por transporte. Esta demanda es satisfecha por el conjunto de vías, vehículos, terminales y por la forma como funcionan estos elementos; es decir, por el *sistema de transporte*. Todo ello se traduce en una interacción entre ambos sistemas (actividades y transporte) que produce un *patrón de flujos*, constituido por viajes entre diversos orígenes y destinos, en diferentes medios o modos de transporte, por variadas rutas y en distintos períodos que, finalmente, produce un equilibrio entre la demanda y oferta de transporte. Su manifestación más evidente es el transporte urbano, es decir, la circulación de personas y vehículos por los espacios públicos de la ciudad (calles, autopistas, líneas de metro, etc.).

La complejidad de los sistemas de transporte actuales ha mostrado la conveniencia de contar con herramientas analíticas que permitan al planificador disponer de un adecuado conocimiento del sistema y poder hacer predicciones.

El uso de modelos matemáticos, ha sido de gran utilidad en la identificación de soluciones o políticas orientadas a lograr los objetivos propuestos para el sistema, constituyéndose en una gran ayuda para los procesos de planificación y de toma de decisiones, ya que el objetivo de los análisis, en la planificación del tráfico urbano, es la obtención de descripciones macroscópicas de los flujos de tráfico a partir del comportamiento de los conductores o usuarios.

Tradicionalmente se ha considerado que el proceso de planificación del transporte

- *Recopilación de datos*. Comprende cualquier tipo de datos que explique como se comporta el sistema de transporte. Generalmente se incluye en este apartado:
 - a. Datos relativos a la infraestructura de transporte y su evolución.
 - b. Datos relativos a las necesidades de transporte y a la utilización de la in-

fraestructura.

- c. Factores tales como la utilización del terreno y distribución de la renta y tipo de empleo y estructura de las ciudades y conurbaciones.
- *Análisis y ajuste de modelos.* Los modelos que se utilizan, y que son generados a partir de los datos de la fase anterior, habitualmente se aplican secuencialmente y se clasifican en:
 - *Modelos de generación/atracción de tráfico.* Explican los viajes que se producen o atraen en una zona en función de variables socioeconómicas de la zona. El esquema empieza considerando una zonificación del área de estudio, una codificación de la red de transporte y la obtención de una base de datos para el estudio. Estos datos están referidos al nivel de actividad económica y demográfica de cada zona, que incluye el nivel de empleo, localización de centros comerciales, zonas recreativas, centros escolares, etc. y son empleados para estimar el número de viajes generados y atraídos por cada zona considerada en el estudio.

Tras esta fase se obtiene una modelización de la red de transporte mediante un grafo $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ donde \mathcal{N} y \mathcal{A} son el conjunto de nodos y arcos (dirigidos) respectivamente. El significado de los arcos depende de si la red es de tráfico o de transporte público. En el primer caso los arcos están asociados a las calles y los nodos a las intersecciones. En el segundo caso cada nodo está asociado a una parada y cada arco representa los posibles desplazamientos entre paradas que un usuario puede efectuar, además, hay otros arcos asociados al tiempo de viaje en el vehículo público, andando o esperando.
 - *Modelos de distribución zonal.* Con ellos se pretende obtener la distribución de los viajes sobre el espacio, esto es, se obtiene el número de viajes que se efectúan de una zona a otra, obteniéndose la denominada matriz de viajes origen-destino (O-D). En esta fase se obtiene un conjunto de pares ordenados de $\mathcal{N} \times \mathcal{N}$ y la demanda de viajes (que inicialmente consideraremos fija). Denotamos este conjunto de pares de demandas por I y cada par O-D por i . Denotamos por g_i la demanda total de viajes para el par $i \in I$.
 - *Modelos de distribución modal.* Explican la elección de modo, es decir, cómo eligen los usuarios el modo de transporte con el que satisfacen sus necesidades de viaje. Con estos modelos se pretende obtener una matriz
 - *Modelos de asignación.* Explican el itinerario que elegirá cada usuario en función del modo elegido anteriormente. Es decir, cada matriz de demanda O-D, para cada modo de transporte, es asignada a su respectiva red de transporte. Usualmente se efectúa una asignación de tráfico por un lado (vehículos privados) y, por otro lado, una asignación a la red de transporte público.

En la actualidad este esquema secuencial ha sido superado por métodos que integran dos o más de estas etapas o modelos simultáneamente. No obstante, este esquema sigue siendo de utilidad a la hora de describir los modelos o de comparar entre modelos alternativos.

Los modelos que integran la fase de asignación en redes congestionadas, ya sean éstas de tráfico o de transporte público, modelizan un equilibrio de Cournot-Nash, en la que los usuarios del sistema de transporte son los jugadores del *juego de equilibrio*, y en el que sus estrategias pueden consistir en la elección de la ruta, el modo de transporte, la realización o no del viaje, etc.

- *Previsiones de la demanda.* A partir de los modelos generados en la fase anterior se hacen previsiones de cómo van a crecer o disminuir las necesidades de transporte (atracciones o generaciones) y de cómo éstas van a originar una futura matriz origen-destino.
- *Evaluación de futuros escenarios.* Conocida la matriz origen-destino futura, se asigna ésta sobre diferentes escenarios para la futura red de transporte y se evalúan las situaciones producidas.

Como se puede deducir del esquema presentado, para caracterizar el funcionamiento de un sistema de transporte es básico conocer, entre otros parámetros, *la matriz origen-destino* y cómo los viajes, que ésta describe, se realizan por la red de tráfico (*asignación de tráfico*).

Históricamente, las *matrices origen-destino* han sido estimadas en base a tres metodologías diferentes:

1. *Estimación directa:* Llevando a cabo un proceso de encuestación, domiciliario o viario (directo sobre una muestra de vehículos o por ficha proporcionada a conductores, o indirecto por captación de números de matrículas).
2. *Estimación por modelos de demanda:* Haciendo uso de un modelo de distribución (e.g.: modelo de gravedad).
3. *Estimación mediante conteos de tráfico:* Utilizando conteos de tráfico a fin de actualizar una matriz origen-destino preexistente.

De las tres aproximaciones anteriores, la primera de ellas, que utiliza resultados de cabo una encuesta amplia y detallada, que permita elaborar matrices O-D completas, resulta ser un proceso costoso, desde el punto de vista económico y social (las encuestas a vehículos sobre el viario son altamente distorsionantes, consumen mucho tiempo y dan lugar a resultados con ciertos sesgos).

La segunda de ellas aplica modelos sintéticos, de demanda de viajes, al sistema de transporte actual para obtener estimaciones de los flujos Origen-Destino. Estos

modelos pueden ser calibrados específicamente para el área de estudio o transferidos desde áreas similares.

La tercera alternativa es la más profusamente utilizada en los últimos treinta y cinco años y ha dado lugar a innumerables contribuciones desde el trabajo pionero de Low (1972). La metodología en la que se basa exige, como prerrequisito, un conjunto de conteos de flujos de tráfico (aforos) y una matriz O-D previa (obtenida de algún estudio cronológicamente anterior o de una encuesta básica que predefina la estructura de la matriz final a obtener). Los aforos de tráfico es un tipo de información que se recoge de forma automática en un conjunto de puntos asociados a arcos del viario y, por consiguiente, se encuentran disponibles en áreas urbanas a muy bajo coste. De aquí que esta última alternativa presenta las ventajas económicas de eliminar encuestas, las manipulaciones de datos, los chequeos y las validaciones en gran medida. Sin embargo, presenta multitud de problemas matemáticos: problemas de gran escala, malas propiedades matemáticas, etc. Es en esta alternativa donde se va a centrar esta tesis.

1.2 Problema de asignación de tráfico (TAP)

Un problema que se puede ver como inverso de la estimación de la matriz O-D, es el *problema de la asignación* de los viajes que esta describe a la red de tráfico urbana. Durante este proceso se utiliza un conjunto de reglas o principios para asignar una matriz O-D en la red de tráfico y así producir un conjunto de flujos en los arcos de ésta. Esto no es, sin embargo, el único producto relevante de la etapa de asignación; esta tiene varios objetivos los cuales es útil considerar en detalle. No todos reciben, en la práctica, el mismo énfasis en todas las situaciones y tampoco se puede alcanzar en todos el mismo nivel de certeza.

Los objetivos primarios de los métodos de asignación son:

- Obtener buenas medidas *agregadas* de la red, por ejemplo flujos totales en los arcos, etc.
- Estimar costes (tiempos) de viajar de zona a zona para un nivel dado de demanda de viajes.
- Obtener flujos *razonables* en los arcos e identificar arcos muy congestionados.
- Estimar las rutas utilizadas para cada par origen-destino.
- Analizar cuáles pares O-D utilizan un arco o un camino particular.

Los datos básicos necesarios para los modelos de asignación son:

- Una matriz O-D, que expresa la demanda de viajes estimada para un intervalo de tiempo determinado (horas pico, horas valle, ...).
- Una red, específicamente conformada por los arcos y sus propiedades.

La premisa básica en la asignación es la suposición de un viajero racional, esto es, un viajero que escoge la ruta que le ofrece los costes menores percibidos. Se cree que hay un buen número de factores que influyen en la elección de ruta entre dos puntos; estos incluyen el tiempo de viaje, la distancia, el coste monetario, la congestión y atascos, los tipos de maniobras requeridas, el tipo de camino, el paisaje, la presencia de señales, las obras en las vías, la constancia en el tiempo de viaje y los hábitos. La construcción de una expresión *generalizada* de costes que incorpore todos estos elementos es una tarea difícil. Además, no es práctico tratar de modelar todos los factores en un modelo de asignación de tráfico, y, por tanto, las aproximaciones son inevitables.

La aproximación más común es considerar sólo dos factores en la elección de rutas: el *coste de tiempo* y el *coste monetario*; además, se considera el coste monetario proporcional a la distancia del viaje. La mayoría de los programas de asignación del tráfico permiten que el usuario asigne pesos al tiempo de viajar y la distancia para representar las percepciones de los conductores en cuanto a estos dos factores. La suma ponderada de estos dos valores llega a ser un coste generalizado utilizado para estimar la elección de ruta.

Por otra parte, hay evidencias que sugieren que, por lo menos en el tráfico urbano de vehículos privados, el tiempo es el factor dominante en la elección de rutas. Sin embargo, es un hecho que conductores diferentes escogen rutas diferentes cuando viajan entre los mismos dos puntos, esto puede deberse a dos tipos de razones:

- Diferencias en las percepciones individuales de lo que constituye la "ruta mejor"; individuos diferentes no sólo pueden incorporar factores diferentes en su función generalizada de costes sino percibirlos de formas distintas.
- Los efectos de la congestión que primero afectan a las rutas más cortas y hacen sus costos generalizados comparables a rutas que inicialmente fueron menos atractivas.

Los distintos tipos de modelos de asignación son más o menos adecuados para considerar estas influencias.

Para la elección de las rutas, la más natural es la de suponer que éstos eligen la ruta que les represente un coste mínimo hacia su destino bajo las condiciones de tráfico reinantes. El resultado de la adopción de este criterio por el colectivo de los conductores, lleva a una situación en la que ningún conductor puede reducir el coste de su desplazamiento mediante cambio de su ruta, llegándose así a un equilibrio, que es consecuencia de su competición en *igualdad* de conocimiento de las posibilidades que

ofrece el sistema de transporte. Este criterio se ha venido denominando de "usuario-óptimo", caracterizado éste por el hecho de que todas las rutas usadas poseen iguales costes marginales. Como luego veremos, Wardrop (1952a) fue el primero en establecer dos criterios de elección de ruta a los que se conoce como los dos *principios de Wardrop*.

Sin embargo, han surgido algunas objeciones acerca de las hipótesis sobre el comportamiento de los usuarios y, por tanto, sobre la aplicabilidad del modelo de equilibrio de Wardrop:

- Se supone que todo conductor percibe el mismo coste bajo las mismas condiciones de tráfico y que todos los conductores tienen completa información acerca de las diferentes alternativas posibles cuando toman la decisión de cómo y por donde van a efectuar su viaje. Es decir, se asume un contexto *determinista*. Frente a esta suposición se han ideado modelos en contextos *estocásticos*
- Los modelos matemáticos de asignación de tráfico suponen que durante todo un determinado periodo de tiempo (p. ej., hora punta por la mañana) se mantienen las mismas condiciones de demanda. Es decir, que son *estáticos* y presuponen la existencia de situaciones estacionarias de equilibrio.

La evolución de los modelos estáticos hacia los *dinámicos* se ha producido por varias causas, entre ellas porque los primeros son incapaces de explicar cualquier evolución de los flujos de tráfico en períodos cortos de tiempo. Esto plantea inconvenientes cuando, por ejemplo, se intenta interrelacionar los modelos de asignación y los problemas de optimización de la regulación semafórica. También, según Ben-Akiva and Lerman (1985) los modelos *estáticos* subestiman los tiempos totales de viaje. La evolución hacia modelos dinámicos tiende a subsanar la segunda de las dos objeciones que se han planteado a los modelos estáticos, casi por construcción. La primera de las objeciones es más difícil de resolver. Los modelos dinámicos más populares mantienen la hipótesis de que los conductores reciben y siguen un conjunto de recomendaciones resultado de un análisis de las condiciones presentes y futuras de la red.

1.2.1 Modelos estáticos de asignación de tráfico

Son los más populares y conocidos y en ellos se va a centrar el desarrollo de esta tesis. Los modelos estáticos asumen una situación estacionaria de la demanda. Es-estratégica o táctica, mientras que para aspectos operacionales se debería recurrir a modelos dinámicos.

Una vez conocida la movilidad en la zona de estudio, actual o futura, y definida la red de transporte que ha de soportarla, se trata de simular el comportamiento de ésta a través de la obtención de las cargas en la red que se deriven de la matriz de movilidad, o matriz OD.

Tabla 1.1: Taxonomía de los modelos de asignación de tráfico a redes urbanas

Evolución de la demanda	Percepción de costes por el usuario	Tipo de Coste	Herramienta matemática empleada
Estáticos	Deterministas	C. separables C. asimétricos	Optimización Desigualdades variacionales
	Estocásticos	C. separables C. asimétricos	Optimización Desigualdades variacionales
Dinámicos	Deterministas	Funciones de salida en los arcos Ecuaciones en derivadas parciales	Simulación Prog. matemática Control óptimo
	Estocásticos	Funciones de salida en los arcos Ecuaciones en derivadas parciales	Simulación Prog. matemática Control óptimo

(*) *Coste Separable*: El coste en un arco sólo depende del volumen de tráfico en ese arco. *Coste Asimétrico*: El coste en un arco depende del volumen de tráfico en él y en los demás arcos.

Existen modelos diferentes de asignación para redes de transporte público y privado, habiendo experimentado un mayor desarrollo los empleados para el transporte privado que son los que aquí se van a analizar.

A partir de la red viaria real, se selecciona un grafo, que garantice todos los posibles itinerarios coherentes entre cada par de demanda.

La modelización de la red supone la descripción de ésta de acuerdo con unas normas, las propias del programa informático que se emplee y que, en general, suelen ser:

- Nudos extremos del arco
- Distancia entre nudos
- Velocidad o tiempo de viaje
- Tipo de área atravesada

A grandes rasgos, los modelos de asignación de tráfico estáticos se centran en unas pocas horas del día, como las horas puntas, y trabajan con valores medios (demandas, tiempos, flujos, etc.) durante el período de estudio.

El problema partirá del número de usuarios que van a viajar en cada par origen-destino e intentará modelizar el comportamiento de éstos en la red de transporte, intentando predecir las rutas que elegirán dentro de las distintas rutas o caminos posibles. Con esto conseguiremos modelizar el flujo de tráfico en cada uno de los tramos que componen la red viaria, evaluando la congestión de cada tramo.

Los distintos modelos de asignación *estáticos* se pueden clasificar en función de cómo se considere que el usuario percibe el coste⁸ que le supone atravesar un arco. Si se asume que todos los usuarios perciben el coste de la misma manera y además, conocen los costes de todas las rutas (tienen información perfecta) estaremos considerando modelos de *asignación determinista*. Sin embargo, en la realidad, las percepciones de los costes están sujetas a variaciones y los usuarios eligen la ruta de acuerdo con su percepción, si se consideran los costes de esta forma, surgen los modelos de *asignación estocástica*.

Modelos de asignación determinista

Existen numerosos modelos que pertenecen a esta clase, todos ellos se pueden clasificar en función de que los costes en los arcos se consideren *separables* o *no separables* (*asimétricos*).

Diremos que un modelo considera los costes en los arcos como *separables* cuando el coste de atravesar un arco de la red no depende del nivel de flujo en los restantes arcos; en caso contrario, consideraremos que los costes son *no separables*.

Además, estos modelos se pueden clasificar según se considere o no que el coste de viaje en un arco depende o no del flujo en los arcos de la red, lo que originará modelos *con y sin congestión*. Los primeros son adecuados para centros urbanos y los segundos para zonas interurbanas.

Los modelos *sin congestión* son modelos también conocidos como modelo de asignación *todo o nada*. En ellos primero se determinan los caminos o rutas de coste generalizado mínimo entre todos los pares origen-destino. Para cada par origen-destino se asignan todos los viajes de dicho par por dichos caminos. Finalmente, conocidos los flujos de cada ruta se obtienen los flujos en los arcos de la red.

Cuando se considera el efecto de la *congestión*, los modelos matemáticos más utilizados son los modelos de *asignación de tráfico en equilibrio* que postulan que, en el equilibrio, todos los caminos utilizados, que unen un mismo par origen-destino, tienen el mismo coste. Estos modelos son conocidos desde 1924, cuando Knight describió, de forma simple e intuitiva el comportamiento del tráfico bajo condiciones de congestión. Unos 28 años después Wardrop (1952a) estableció sus dos principios, ya esbozados, que formalizaron esta noción de equilibrio e introdujeron el postulado de comportamiento alternativo en

⁸Coste percibido por el usuario en base a consideraciones subjetivas y objetivas de tiempo, seguridad, comodidad, etc.

el sentido de minimizar el coste total de todos los viajes efectuados en la red.

Así, el *primer principio de Wardrop* establece que: “*Los costes de viaje de todas las rutas usadas en el equilibrio son iguales y menores que en los que incurriría un único vehículo que utilizara una ruta distinta a la que utiliza en esos momentos.*”.

Es decir, que en el equilibrio, ningún usuario puede reducir el coste de su viaje mediante cambio de ruta.

Este principio, que es idéntico a la descripción de Knight, llegó a ser aceptado después de pasar 40 años como muy útil para describir el comportamiento en la ocupación de rutas cuando existen varias alternativas. Los flujos de tráfico que satisfacen este principio se les llama *flujos óptimos para el usuario*, pues cada usuario elige la ruta que percibe como mejor para él. En el otro extremo, los *flujos óptimos para el sistema* son caracterizados por el *segundo principio de Wardrop* como estados en los que el tiempo promedio de viaje es mínimo, es decir, que *los usuarios eligen la ruta de modo que se minimice el tiempo total de transporte en la red*. El primer principio es utilizado para modelizar el comportamiento de los usuarios, mientras que el segundo es usado como un criterio para diseñar la red de transporte.

El primer principio de Wardrop, también denominado DUE, asume que todos los usuarios perciben el coste de la misma manera y, además, conocen los costes de todas las rutas (tienen información perfecta).

En la realidad las percepciones de los costes están sujetas a variaciones y los usuarios eligen la ruta de acuerdo con su percepción.

En este contexto, se han elaborado modelos de equilibrio en los que los costes de viajes son la suma de una parte fija más una componente aleatoria, en este caso los usuarios eligen una u otra ruta dependiendo de la distribución de probabilidad de los costes aleatorios. Esta forma de asignación recibe el nombre *asignación estocástica* (SUE⁴).

La primera formulación de las condiciones de equilibrio fue realizada por Beckman (1956) demostrando que resolver la conocida “Transformada de Beckman” era equivalente a resolver un problema de asignación DUE, además, se demuestra que la solución en términos de flujos en los arcos, bajo ciertas condiciones de monotonía en el coste de los arcos, es única, no así en términos de flujos en los caminos. Por otra parte, el algoritmo de Frank Wolfe se mostró útil para resolver, en la práctica, este problema.

Análogamente a la clasificación anterior, podemos distinguir entre modelos con *congestión* y *no congestión*. En un entorno de *no congestión*, el esquema es común para todos los modelos. Primero realizan una determinación de todas las rutas existentes entre cada par origen-destino; especificando su coste. A continuación, para cada par

⁴*Stochastic user equilibrium*

origen-destino, en función del coste de las rutas, se crea la matriz de probabilidad de elección de ruta, después, para cada par, se asignan los viajes a la rutas creadas, en función de la matriz de probabilidad de elección de ruta; finalmente, conocidos los flujos de ruta, se obtienen los volúmenes en los arcos.

La diferencia fundamental entre los distintos métodos de este tipo estriba en la forma en que calculan las probabilidades de elección de los caminos (rutas). En general aplican conceptos de la teoría de la utilidad, considerando los costes en los caminos como variables aleatorias. Así, si se consideran los costes en los caminos como una variable aleatoria *uniforme* tendremos el *modelo de Burrel*, si se los considera como una v.a. *Gumbel* surge el *modelo Logit* y, finalmente, si es una v.a. *Normal* tendremos el *modelo Probit*.

El modelo de Burrel no tiene grandes dificultades de aplicación. En el modelo Logit, para calcular las probabilidades de elección de rutas se puede optar por realizarlo analíticamente o mediante la aplicación del *algoritmo de Dial*. Sin embargo, en los modelos Probit hay que recurrir a aproximaciones analíticas (Algoritmos de integración numérica, método de aproximaciones sucesivas (o de *Clark*), etc.) o simulación de Monte Carlo.

En un entorno con *congestión*, de la misma forma que para los modelos deterministas, se demuestra que existe un problema matemático cuya resolución es equivalente a realizar una asignación estocástica desde el punto de vista del usuario, con costes de congestión. Además, se demuestra que tanto la solución, en términos de flujos en los arcos, como de flujos en las rutas es única. El método de las *medias sucesivas* ha sido eficaz en la resolución de este problema.

1.2.2 Modelos dinámicos de asignación de tráfico

El problema de asignación dinámica de tráfico consiste en la estimación de los flujos de vehículos que utilizan los diferentes tramos de la red viaria, de manera variable con el tiempo, constituyendo por tanto una extensión del problema de asignación convencional o en equilibrio López Lois (1997) .

Principalmente se han desarrollado modelos deterministas de *asignación dinámica de tráfico* bajo tres aproximaciones:

- ◆ Simulación
- ◆ Programación matemática

- ◆ Control óptimo sobre redes

Si bien, algunas de las formulaciones en programación matemática pueden contemplarse como problemas en control óptimo discreto y, a su vez, las formulaciones en

control óptimo (continuas) se aproximan mediante programación matemática para su resolución quedando reducidas, en definitiva, a programación matemática.

Estos modelos consideran que los usuarios minimizan sus tiempos de viaje actualizando continuamente sus rutas elegidas de acuerdo con las condiciones del tráfico. El problema es especialmente importante para los *sistemas avanzados de información sobre el viajero* (ATIS), los cuales requieren información para recomendar rutas de acuerdo con el tiempo de viaje futuro en los arcos que serán incluidos en las rutas. Su objetivo es mejorar el comportamiento del tráfico, reduciendo la congestión al proporcionar condiciones de tráfico uniformes. En estos sistemas existe un controlador central que recomienda las rutas a los usuarios en tiempo real, lo cual satisface objetivos individuales o del sistema mediante equilibrio de usuario o mediante optimización del sistema y considera, en tiempo real, variaciones inesperadas en las condiciones de la red.

El primer modelo de simulación por ordenador de la asignación dinámica de tráfico, de acuerdo con los principios de Wardrop, fue presentado por Yagar (1971).

El primer modelo que utilizó la programación matemática fue el de Merchant and Nemhauser (1978a) y Merchant and Nemhauser (1978b). Este modelo, planteado inicialmente para un único destino en toda la red, puede caracterizarse como no lineal y no convexo. Si embargo, Carey (1987) reformuló el problema de Merchant Nemhauser como un problema no lineal convexo, con ventajas computacionales y analíticas.

El problema de Merchant Nemhauser fue implementado y resuelto por Ho (1990) mediante linealización parcial y una aplicación multinivel del principio de descomposición de Dantzig-Wolfe. También, de entre las aproximaciones que usan la programación matemática, cabe destacar la heurística desarrollada por Janson (1991)

Las formulaciones de control óptimo para el problema de asignación dinámica de tráfico fueron iniciadas por Friesz y Luque entre los años 1980 y 1989 (Friesz et al. (1989)). Los modelos planteados por estos autores aparecen como una clara extensión del problema de Merchant y Nemhauser para el caso continuo. Wie et al. (1990) formularon, para un modelo lineal, la extensión al caso de redes con más de un destino.

Todos los modelos dinámicos, mencionados hasta el momento, incluyen una modelización del fenómeno de la propagación del tráfico mediante restricciones que relacionan el flujo y el número de vehículos en los arcos a partir de las llamadas "*funciones de salida en los arcos* (Exit link functions)". Ran and Shimazaki (1989a), Ran and Shimazaki (1989b) consideraron inicialmente modelos, también bajo control óptimo determinista, que no contienen estas restricciones, son conocidos como *modelos sin funciones de salida* y del clásico método *MaxMin*, basado en la dualización parcial de las restricciones para la solución del problema de optimización resultante. Posteriormente, Boyce et al. (1991) propusieron el método de Frank-Wolfe para resolver la misma discretización, aprovechando la estructura de red de las restricciones del problema. Sin embargo, en todos los trabajos efectuados hasta el momento, no se han presentado experiencias

computacionales en redes de tamaño apreciable, a excepción del de Janson Janson (1993) con redes de algunos centenares de nodos.

Los modelos de control óptimo o programación no lineal, mencionados hasta aquí, han sido desarrollados por sus autores siguiendo la línea marcada por los modelos estáticos utilizados en planificación de transporte urbano y son catalogados como "usuario-óptimo" o "sistema-óptimo", si bien se debe destacar que las correspondientes extensiones de los principios de Wardrop efectuadas por autores diferentes son también *distinguidas*, sobre todo en el caso de los modelos considerados como "usuario-óptimo". Algunas de las propiedades de estos modelos han sido objeto de discusión y objeciones por parte de autores como Ghali and Smith (1993). Cabe mencionar también que, existen otros modelos, para la evolución dinámica de los flujos de tráfico, que no son considerados como evolución de los modelos de asignación estáticos, puesto que, o bien no hay en ellos una generalización de los principios de Wardrop, o bien no contienen como caso particular a los modelos estáticos. Entre ellos cabe citar a los modelos de Papageorgiou (1989) en cuya formulación en control óptimo discreto se apunta la modelización conjunta de redes de tráfico urbanas e interurbanas y se examina la posibilidad de utilizar control con realimentación para la determinación del guiado en dichas redes y los modelos de Sarachik y Ozgüner (1982) y los de Charbonnier, Farges y Henry (1991). Vythoulkas (1989) y Ran (1992) han propuesto sendos modelos estocásticos de asignación dinámica de tráfico.

Montero et al. (1998) muestran como los modelos de asignación de tráfico, basado en las aproximaciones de equilibrio de usuario son una de las herramientas más usadas en el planeamiento del transporte. También muestran como la infraestructura prevista es probablemente suficiente para la demanda media, pero las variaciones del flujo de tráfico con el tiempo, combinado con la influencia de la geometría de las vías, pueden producir una congestión indeseada que no puede ser pronosticada o analizada con las herramientas estáticas. Ante esta situación proponen la combinación del modelo EMME/2, con un simulador de tráfico microscópico, el AIMSUN2. En Montero et al. (2001) profundizan en el mismo tema, con énfasis sobre la descripción de los interfaces específicos que hacen consistente la combinación de ambas herramientas en el entorno GETRAM. Asimismo, muestran que los modelos para los sistemas complejos de transporte deben ser una combinación de modelos matemáticos y modelos de computación, para superar, por ejemplo, las dificultades de la integración entre modelos. El entorno GETRAM tiene una arquitectura de computación abierta y flexible, conveniente para tales propósitos.

Barceló et al. (1998) proponen dos sistemas para la gestión del tráfico. Ambos están la toma de decisiones en dos tipos de situaciones: una red gestionada por un único centro de control de tráfico, y un área, posiblemente metropolitana, subdividida en zonas, cada una bajo el control de diferentes centros de control, con posibles estrategias de control diferentes, que requerirían un cierto nivel de coordinación para mejorar el funcionamiento del sistema. Los sistemas propuestos se basan en un pronóstico a corto plazo de la demanda y un simulador microscópico para estimar estados de la red,

actuales y futuros.

Finalmente, Barceló et al. (2005) realizan una revisión de alguno de los desarrollos realizados recientemente en el simulador de tráfico microscópico AIMSUN para proporcionar una mejor respuesta a los requerimientos de los sistemas de ITS, de análisis avanzados de tráfico y ATMS.

1.2.3 Formulación matemática del problema estático de asignación de tráfico

El problema de optimización

Un problema de optimización viene caracterizado por tres elementos: *las variables del problema* que definen el conjunto de decisiones, la *función objetivo* que evalúa el coste o el beneficio de la decisión y el *conjunto de soluciones* que determina las decisiones válidas que pueden llevarse a cabo. Más formalmente, sea $f : S \mapsto \Re$ una función, X un subconjunto de S que se denomina *conjunto factible de soluciones*, entonces el problema de optimización (versión minimización) se formula

$$\begin{aligned} &\text{Minimizar } f(x) \\ &\text{sujeto a: } x \in X. \end{aligned}$$

La formulación anterior es demasiado general para que su estudio conduzca a métodos satisfactorios de resolución, ya que S puede ser un conjunto cualquiera y por tanto esta formulación recoge la situación donde S es un espacio de dimensión infinita (por ejemplo un espacio de funciones). Este caso se denomina optimización de *funcionales* y el funcional f recibe (usualmente) el nombre de *función de energía*.

La forma de abordar el problema anterior es analizar situaciones particulares. Se van a exigir, tanto a f como a X , propiedades que sean suficientemente generales para poder ser utilizadas en las aplicaciones en estudio, y lo suficientemente *fuertes* para obtener resultados de interés. La hipótesis más débil que suele exigirse a X es que sea un conjunto cerrado. En los modelos que estudiaremos asumiremos que el conjunto X es convexo, incluso en la mayoría de los casos, tendremos propiedades aún más fuertes como la de ser un conjunto poliédrico (definido mediante restricciones lineales). La regularidad local para poder caracterizar los extremos (mínimos) locales del problema y por otro, se exigen propiedades acerca del comportamiento global de la función, de modo que permitan garantizar que tales extremos locales son también extremos globales.

Para el primer grupo de hipótesis, las propiedades más empleadas son la F - diferenciabilidad, la teoría de los *subdiferenciales* (subgradientes) (Rockafellar (1970)) para

funciones convexas y la teoría de los *gradientes generalizados* para funciones localmente lipschitzianas (Clarke (1990)). Las propiedades globales más empleadas son la *convexidad* o una generalización de este concepto (bajo la hipótesis de F-diferenciabilidad) denominado *pseudoconvexidad* (Bazaraa et al. (1993)). Un tipo de modelo abordado es el *problema de optimización convexa diferenciable* (CDP). Más formalmente,

$$\begin{aligned} & \text{Minimizar } f(x) \\ & \text{sujeto a: } x \in X, \end{aligned} \quad [\text{CDP}(f, X)]$$

donde la función $f : \mathbb{R}^n \mapsto \mathbb{R}$ es continuamente diferenciable, pseudoconvexa (por ejemplo cuando es convexa) y el conjunto X es un subconjunto convexo de \mathbb{R}^n . Las soluciones a este problema vienen caracterizada por el teorema siguiente

TEOREMA 1.2.1 (Condiciones de optimalidad de $\text{CDP}(f, X)$). *El punto $x^* \in X$ es una solución óptima de $\text{CDP}(f, X)$ si y sólo si cumple*

$$\nabla f(x^*)^T(x - x^*) \geq 0, \quad \forall x \in X. \quad (1.1)$$

donde ∇f es el gradiente de la función f . (Ver, por ejemplo, el teorema 3.3.4 de Bazaraa et al. (1993))

Las condiciones anteriores pueden ser reformuladas cuando el conjunto X está definido explícitamente por $X = \{x \in \mathbb{R}^n : g_i(x) \leq 0, i \in \mathcal{I} \text{ y } h_j(x) = 0, j \in \mathcal{J}\}$. Supongamos que las funciones g_i con $i \in \mathcal{I}$ son funciones convexas y que todas son diferenciables en \mathbb{R}^n y que las funciones h_j con $j \in \mathcal{J}$ son funciones afines Slater. La condición de optimalidad (1.1) es equivalente a las condiciones de Karush-Kuhn-Tucker Gale et al. (1951). Un punto x^* es una solución óptima para el $\text{CDP}(f, X)$ si y sólo si el sistema

$$\begin{aligned} \nabla f(x^*) + \sum_{i \in \mathcal{I}} \mu_i \nabla g_i(x^*) + \sum_{j \in \mathcal{J}} \lambda_j \nabla h_j(x^*) &= \mathbf{0}, \\ g_i(x^*) \mu_i &= 0, \quad i \in \mathcal{I}, \\ g_i(x^*) &\leq 0, \quad i \in \mathcal{I}, \\ h_j(x^*) &= 0, \quad j \in \mathcal{J}, \\ \mu_i &\geq 0, \quad i \in \mathcal{I}, \end{aligned}$$

tiene solución.

Modelos de desigualdades variacionales

ciones más generales que las posibilitadas por la programación convexa diferenciable. Formalmente este problema se plantea del siguiente modo. Se considera un conjunto cerrado y convexo $X \subset \mathbb{R}^n$ y una función $F : X \mapsto \mathbb{R}^n$ continua en X . El problema de desigualdades variacionales consiste en encontrar un $x^* \in X$ cumpliendo

[VIP(F, X)]

$$F(x^*)^T(x - x^*) \geq 0, \quad \forall x \in X$$

Este problema también es conocido con los nombres de *ecuación generalizada* o de *problema de punto estacionario*.

Han sido caracterizados varios conjuntos de hipótesis que garantizan la existencia de soluciones a este problema. Por ejemplo bajo la suposición de que X es un conjunto compacto (Hartman and Stampacchia (1966), Brézis (1968)) o bajo la hipótesis de que la aplicación F es *coercitiva* (Hartman and Stampacchia (1966), Moré (1974)).

Si la aplicación F es *fuertemente monótona* entonces el problema tiene solución (Stampacchia (1969)), si además F es *pseudomonótona* entonces el conjunto de soluciones, $SOL(F, X)$, es un conjunto convexo, y bajo la *cualificación de restricciones de Slater* o la hipótesis de coercitividad de F el conjunto de soluciones está acotado. Se puede demostrar la unicidad de la solución bajo la hipótesis de que es *estrictamente monótona* (Stampacchia (1969)).

Las condiciones de optimalidad (1.1) muestran que el problema $CDP(f, X)$ puede ser formulado como un problema de desigualdades variacionales $VIP(\nabla f, X)$. Cabría plantearse estudiar el camino inverso, es decir, determinar bajo qué condiciones la función F es un gradiente, esto es, $F = \nabla f$. Una condición suficiente (teorema 4.1.6 Ortega and Rheinboldt (1970)) es la siguiente:

TEOREMA 1.2.2 (Condición suficiente para que $F = \nabla f$). *Sea $F : X \mapsto \mathbb{R}^n$ de clase C^1 sobre un conjunto convexo y abierto $X_0 \subset X$. Entonces F es el gradiente de una aplicación en X_0 si y sólo si $\nabla F(x)$ es simétrica para todo $x \in X_0$.*

Bajo la condición de simetría, la integral definida por

$$f(x) = \int_0^1 \sum_{j=1}^n F_j(\hat{x} + s(x - \hat{x}))(x_j - \hat{x}_j) ds \quad (1.2)$$

donde \hat{x} es un elemento arbitrario de X_0 , es independiente del camino elegido, y de acuerdo al teorema de Green, la función F es integrable. En este caso el problema $VIP(F, X)$ puede ser planteado como un programa matemático cuya función objetivo es f . El siguiente teorema muestra las relaciones entre $F = \nabla f$ y la función f .

TEOREMA 1.2.3 (Relaciones entre la monotonía de f y de F). *Sea $F \equiv \nabla f$. Entonces se cumple:*

1. F es monótona en $X \Leftrightarrow f$ es conveza en X .
2. F es estrictamente monótona en $X \Leftrightarrow f$ es estrictamente conveza en X .

Las demostraciones de estas relaciones se pueden consultar en Patriksson (1993).

Las condiciones de optimalidad de un problema en desigualdades variacionales, $VIP(F, X)$, se sintetizan con el siguiente teorema (Ver, por ejemplo, Lim (2002)):

TEOREMA 1.2.4 (Condiciones de optimalidad de un $VIP(F, X)$). *Sea X un conjunto no vacío, compacto y convexo de \mathbb{R}^n y sea F una aplicación continua del conjunto X un conjunto \mathbb{R}^n . Entonces existe solución del problema $VIP(F, X)$. Además, si F es estrictamente monótona en X , la solución es única.*

Formulaciones matemáticas del problema de asignación de tráfico

Ahora vamos a abordar las formulaciones matemáticas del problema de asignación de tráfico. Denotamos por K_i el conjunto de rutas para el par O-D i , con h_k el flujo en el camino (o ruta) k y por $C_k := C_k(h)$ el coste de viaje en la ruta k experimentado por un usuario para un vector de flujo $h \in \mathbb{R}^{|K|}$, siendo $|K|$ el número total de rutas en la red. Empleando esta notación un vector de flujo h^* está en equilibrio si y sólo si

$$\begin{aligned} \text{Si } h_k^* > 0 &\Rightarrow C_k = U_i, \quad k \in K_i, \quad i \in I. \\ \text{Si } h_k^* = 0 &\Rightarrow C_k \geq U_i, \quad k \in K_i, \quad i \in I. \end{aligned} \quad (1.3)$$

donde el valor de $U_i := U_i(h^*)$ es el mínimo coste de transporte en las rutas del par O-D $i \in I$

Para formular matemáticamente las condiciones de equilibrio dadas en (1.3) se necesita describir los requerimientos de factibilidad de los flujos. Denotamos los flujos en los arcos como $v \in \mathbb{R}^{|A|}$ y por K_i el conjunto de rutas para el par O-D i . El primer requerimiento es que la demanda para cada par O-D i debe ser satisfecha, esto es

$$\sum_{k \in K_i} h_k = g_i, \quad \forall i \in I. \quad (1.4)$$

Además los flujos deben ser no negativos

$$h_k \geq 0, \quad \forall k \in K. \quad (1.5)$$

La relación entre los flujos en los arcos y los flujos en las rutas viene definida por

$$\sum_{i \in I} \sum_{k \in K_i} \delta_{ak} h_k = v_a, \quad \forall a \in A,$$

donde $\delta_{ak} = 1$ si la ruta k utiliza el arco a y cero en caso contrario. Esta restricción indica que el flujo en un arco a es la suma del flujo de todos los caminos que emplean dicho arco.

La primera formulación de las condiciones de equilibrio (1.3) mediante un modelo de optimización fue realizada por Beckman et al. (1956). Estos autores asumieron que el coste en cada arco depende exclusivamente de su flujo (*costes separables*). La solución

del siguiente problema de optimización caracteriza la situación de equilibrio derivada del primer principio de Wardrop

$$\begin{aligned}
 &\text{Minimizar } Z = \sum_{a \in \mathcal{A}} \int_0^{u_a} c_a(x) dx \\
 &\text{sujeto a: } \sum_{k \in K_i} h_k = g_i, \quad \forall i \in I, \\
 &\quad \sum_{i \in I} \sum_{k \in K_i} \delta_{ak} h_k = u_a, \quad \forall a \in \mathcal{A}, \\
 &\quad h_k \geq 0, \quad \forall k \in K_i, \quad \forall i \in I.
 \end{aligned}
 \tag{TAP}$$

La situación de equilibrio derivada del segundo principio de Wardrop (para costes separables y crecientes) se puede caracterizar como solución del siguiente modelo de optimización.

$$\begin{aligned}
 &\text{Minimizar } Z = \sum_{a \in \mathcal{A}} c_a(u_a) u_a \\
 &\text{sujeto a: } \sum_{k \in K_i} h_k = g_i, \quad \forall i \in I, \\
 &\quad \sum_{i \in I} \sum_{k \in K_i} \delta_{ak} h_k = u_a, \quad \forall a \in \mathcal{A}, \\
 &\quad h_k \geq 0, \quad \forall k \in K_i, \quad \forall i \in I.
 \end{aligned}
 \tag{TAP-SE}$$

Se ha recurrido a modelos matemáticos más generales que los modelos de optimización (como son las desigualdades variacionales, problemas de complementariedad o formulaciones de punto fijo) para modelizar situaciones más realistas del problema de asignación, tales como que el coste de un arco no depende exclusivamente de su flujo sino que puede depender del flujo en otros arcos de la red. Esta situación es la común en las intersecciones de calles. Otros nuevos aspectos del problema, que pueden ser tenidos en cuenta, son la existencia de varios tipos de usuarios (*modelos multi-usuarios*) o la existencia de varios modos de transporte (*modelos multimodales*) como, por ejemplo, el transporte público y privado simultáneamente. Los costes derivados de este tipo de problemas tienen un jacobiano típicamente asimétrico y por tanto no son formulables mediante un modelo de optimización.

La formulación matemática más usual de las condiciones de equilibrio Wardropianas, se basa en los problemas de desigualdades variacionales en el espacio de flujo en los caminos h_k , $k \in K$, en este caso el espacio de factibilidad de los flujos está definido por las ecuaciones (1.4) y (1.5), y puede ser expresado matricialmente por

$$h \geq 0, \tag{1.7}$$

donde $g \in \mathfrak{R}_+^{|I|}$ es el vector de demandas, Λ es la matriz de incidencia de los pares origen destino y de los caminos. El elemento λ_{ik} de esta matriz vale 1 si la ruta k une el par i y 0 en caso contrario. El espacio de flujo en los caminos es

$$\Omega_h = \{h \in \mathfrak{R}_+^{|K|} \mid \Lambda h = g\}.$$

Las condiciones de Wardrop (1.3) se pueden formular mediante el siguiente problema en desigualdades variacionales. Encontrar un $h^* \in \Omega_h$ cumpliendo la desigualdad

$$C(h^*)^T(h - h^*) \geq 0, \quad \forall h \in \Omega_h \quad [\text{TAP-VIP}(C(\cdot), \Omega_h)]$$

donde $C(h^*)$ es el coste en los caminos para el flujo h^* . Notar que si existiesen varios grupos de usuarios o modos de transporte estas situaciones son fácilmente modelizables. Es suficiente crear tantas copias de la red de transporte como grupos de usuarios o modos de transporte, relacionando sus costes de viaje, en caso necesario, mediante el vector de costes C .

En el caso de que el coste en una ruta fuese la suma de los costes en cada uno de sus arcos (*costes aditivos*), las condiciones de Wardrop pueden ser descritas en términos de los flujos en los arcos. El conjunto de factibilidad para los flujos en los arcos viene descrito por el conjunto (poliedro acotado)

$$v = \Delta h, \quad (1.8)$$

$$g = \Lambda h, \quad (1.9)$$

$$h \geq 0, \quad (1.10)$$

donde Δ es la matriz de incidencia arco-ruta, el elemento δ_{ak} toma el valor 1 si la ruta k emplea el arco a y vale 0 en caso contrario. En este caso el conjunto de factibilidad viene definido por

$$\Omega_v = \{v \in \mathbb{R}^{|\mathcal{A}|} \mid \exists h \in \Omega_h \text{ con } v = \Delta h\},$$

y el problema $\text{VIP}(C(\cdot), \Omega_v)$ es equivalente a encontrar un $v^* \in \Omega_v$ cumpliendo la desigualdad

$$c(v^*)^T(v - v^*) \geq 0, \quad \forall v \in \Omega_v \quad [\text{TAP-VIP}(c(\cdot), \Omega_v)]$$

donde $c : \mathbb{R}^{|\mathcal{A}|} \mapsto \mathbb{R}^{|\mathcal{A}|}$ es la función de costes en los arcos. La relación entre los costes en los arcos y los costes en los caminos viene dado por $C(h) = \Delta^T c(v)$.

El conjunto Ω_v puede ser formulado sin necesidad de recurrir a las variables de flujos en los caminos. Esta representación es conocida con el nombre de formulación *nodo-arco*, e impone para cada par O-D las condiciones de conservación de flujo en cada

$$Ev = g, \quad v \in \mathcal{A},$$

donde $E \in \{-1, 0, 1\}^{|\mathcal{N}| \times |\mathcal{A}|}$ es la matriz de incidencia nodo arco de la red, las componentes del vector g^i están definidas por

$$\bar{g}_q^i = \begin{cases} 1, & \text{si } i = (q, j), \\ -1, & \text{si } i = (m, q), \\ 0, & \text{si } q \text{ no es el nodo origen o destino del par } i, \end{cases}$$

y el vector v^i es el vector de flujos en los arcos producidos por el par de demanda O-D i . El flujo agregado en los arcos será la suma de todos los flujos producidos por todos los pares O-D, esto es $v = \sum_{i \in I} v^i$. Resumiendo, la región factible en los arcos puede venir definida directamente por

$$\Omega_{\omega} = \left\{ v \in \mathbb{R}^{|A|} \mid \exists v^i \in \mathbb{R}_+^{|A|} \quad \forall i \in I \text{ con } v = \sum_{i \in I} v^i \text{ y } Ev^i = \bar{g}^i \right\},$$

dando lugar al problema equivalente de encontrar un $v^* \in \Omega_{\omega}$ que cumple la desigualdad

$$c(v^*)^T(v - v^*) \geq 0, \quad \forall v \in \Omega_{\omega} \quad [\text{TAP-VIP}(c(\cdot), \Omega_{\omega})]$$

El principio de equilibrio estocástico de usuario (SUE) fue establecido por Daganzo and Sheffi (1977) como una extensión del principio de equilibrio de Wardrop para poder incluir los efectos de la percepción del tiempo de viaje por parte de los usuarios. Este principio puede enunciarse como:

Se alcanza una distribución de flujos en las rutas, según el principio de equilibrio estocástico de usuario, cuando no hay ningún usuario que crea que su tiempo de viaje puede mejorarse escogiendo otra ruta.

Esta definición refleja que los tiempos de viaje están sujetos a la percepción del usuario y que deben considerarse componentes estocásticas en la expresión de los tiempos de viaje C_{ik} de una ruta k para un par origen-destino $i \in I$: $C_{ik} = \gamma_{ik} + \xi_{ik}$, siendo γ_{ik} el tiempo de viaje medido sobre la ruta k y ξ_{ik} una variable aleatoria de esperanza nula, $E[\xi_{ik}] = 0$.

Los modelos más populares para la percepción del tiempo de viaje son las distribuciones tipo logit y la tipo probit. En el modelo logit las variables $-\theta\xi_{ik}$ son independientes y siguen una distribución Weibull-Gumbel mientras que en la tipo probit siguen una distribución normal. En consecuencia, se debe considerar una distribución $\{\Pi_k\}$ de probabilidades para los flujos en las rutas $k \in K_i$ que unen un par origen-destino $i \in I$, de forma que, si g_i es la demanda en número de viajes en ese par, entonces los flujos en las rutas vengan dados por $h_k = \Pi_k g_i$. Una condición que permite que la condición de equilibrio estocástico de usuario pueda expresarse para variables aleatorias tanto discretas como continuas, es:

$$P(C_{ik} < C_{ik'}, \forall k' \neq k, k' \in K_i) \leq \frac{h_k}{g_i} \leq P(C_{ik} \leq C_{ik'}, \forall k' \neq k, k' \in K_i) \quad (1.11)$$

Si C_{ik} son variables aleatorias continuas, entonces $\Pi_k = P(C_{ik} \leq C_{ik'}, \forall k' \neq k, k' \in K_i)$. Si los costes reales γ_{ik} son independientes de la congestión, entonces la

determinación de las probabilidades Π_k y de los flujos totales en los arcos se conoce como el problema de carga estocástica de la red. En el caso de un modelo tipo logit el problema de carga estocástica de la red puede resolverse de forma bastante eficiente, mediante el clásico procedimiento STOCH de Dial (Dial (1971)).

Fisk (1980), extendió el problema [TAP] para equilibrio determinista de usuario con costes separables y aditivos a un modelo estocástico tipo logit con parámetro de dispersión θ . El llamado problema de Fisk es:

$$\text{Min}_{v \in \Omega_v, h \in \Omega_h} \sum_{a \in \mathcal{A}} \int_0^{v_a} c_a(x) dx + \frac{1}{\theta} \sum_{i \in I} \sum_{k \in K_i} h_k (\ln h_k - 1) \quad (1.12)$$

En caso de un modelo tipo probit resulta conocido el método de Burrell (1968).

1.2.4 Algoritmos para la resolución del problema de asignación de tráfico

Los algoritmos desarrollados para resolver el problema de asignación de tráfico (TAP) son altamente especializados y aprovechan todas las características del problema, permitiendo resolver eficientemente problemas reales de gran tamaño.

Multitud de estrategias han sido aplicadas en la resolución del TAP, ver la monografía Patriksson (1994) en la que hay recogidas más de mil referencias bibliográficas relacionadas con este problema, y su revisión escapa a los objetivos de esta tesis doctoral. En la actualidad el algoritmo de Bar Gera constituye el *Estado-del-Arte* en la resolución del TAP (Bar-Gera (2002a), Bar-Gera (2002b)). Este algoritmo permite alcanzar precisiones del orden de 10^{-10} en redes reales de grandes dimensiones, contrastando con el algoritmo de Frank-Wolfe (Frank and Wolfe (1956)) (que es el *Estado-de-la-Práctica*) que no supera una exactitud del orden de 10^{-4} para estos problemas. Sin embargo, aún no se ha mostrado cómo aplicarlo a modelos con costes no separables (Patriksson (2002)).

Algoritmos de generación de columnas

El problema de estimación de matrices de viaje Origen-Destino en redes de tráfico congestionadas (DAP) requiere la resolución de multitud de problemas TAP y esto hace sea el más adecuado cuando se deban resolver muchos problemas TAP, ya que este algoritmo trabaja con flujos en los arcos desagregados por orígenes. Cuando la matriz de origen-destino se cambia esta información ya no es directamente aprovechable ni tampoco la gran exactitud con que se ha calculado la situación de equilibrio. Las estrategias algorítmicas que se presentan más adecuadas, y que son utilizadas en esta tesis doctoral, son los denominados *métodos de generación de columnas* (GC) que

constituyen una herramienta para la elaboración de algoritmos para la resolución de problemas de programación matemática de gran escala. Estos métodos han sido aplicados exitosamente a multitud de aplicaciones de la programación lineal, entera y no lineal; constituyendo un área de creciente interés en las investigaciones.

Un algoritmo CG puede ser descrito a *grosso modo* por dos subproblemas de optimización y un procedimiento de coordinación entre ambos subproblemas. Un algoritmo CG itera entre la resolución de una relajación del problema original, el llamado (*subproblema CG*), definido a partir de la actual aproximación a la solución y cuya solución proporciona un nuevo vector (esto es, la columna), una cota del valor óptimo del problema y la respuesta a un test de optimalidad; y el llamado *problema maestro restringido* (RMP) definido sobre la envoltura convexa de un subconjunto de las columnas generadas anteriormente y cuya función objetivo es una medida de la optimalidad de las soluciones. La solución óptima al RMP es usada para definir el nuevo subproblema CG.

Dependiendo del tipo de problema que se está resolviendo, la definición de los subproblemas puede ser bastante diferente. En *programación lineal*, un subconjunto de restricciones lineales son relajadas Lagrangianamente con la actual estimación de los multiplicadores. El CG es un problema lineal que genera como columna un punto extremo del poliedro definido por las restricciones no relajadas. El RMP es un combinación convexa de los puntos extremos generados de modo que las restricciones relajadas sean satisfechas, y las variables duales de estas restricciones son usadas para definir el próximo subproblema CG. Este algoritmo fue propuesto por Dantzig and Wolfe (1960).

En programación lineal *entera*, los subproblemas CG está usualmente basado en la relajación Lagrangiana descrita en el algoritmo de descomposición de Dantzing-Wolfe. El subproblema es un problema lineal entero mientras que el RMP está formulado como un problema lineal. Este algoritmo conduce a la solución de una convexificación del problema entero original, donde el conjunto definido por las restricciones no relajadas es remplazado por su envoltura convexa, siendo sus puntos extremos identificados en el CG.

En *programación no lineal* la relajación puede actuar como un mecanismo de linealización de las partes no lineales del problema. El algoritmo de descomposición no lineal de Dantzing-Wolfe (Lasdon (1970)) para problemas convexos generaliza el anteriormente descrito para programación lineal. En este algoritmo el problema CG es no lineal mientras que el RMP sigue siendo un problema lineal.

Holloway (1974) y Hohenbalken (1977). La forma clásica de la descomposición simplicial fue primeramente descrita para problemas no lineales con restricciones lineales y ha sido generalizada para abordar problemas de la forma

$$\underset{x \in X}{\text{minimizar}} f(x), \quad [P(f, X)]$$

donde X es un conjunto compacto, convexo y no vacío; y $f : X \mapsto \Re$ es continuamente

diferenciable y pseudoconvexa en X .

En los algoritmos de descomposición simplicial los subproblemas RMP y CG se pueden interpretar como dos aproximaciones al problema original $P(f, X)$ obtenidas respectivamente como una aproximación (*interior*) de la región factible y de la función objetivo.

- Problema maestro restringido. Estos métodos construyen y resuelven una aproximación al problema original, obtenida reemplazando la región factible por un conjunto convexo (e.g. poliedral) que es una *aproximación interior* de dicha región.

$$\underset{x \in \hat{X} \subset X}{\text{minimizar } f(x)}, \quad [\text{RMP}(f, \hat{X})]$$

La región factible del RMP \hat{X} debe contener el segmento definido por la última columna generada (\hat{y}) y por la solución al último RMP (\hat{x}), esto es

$$[\hat{y}, \hat{x}] \subset \hat{X}. \quad (1.13)$$

Esta exigencia garantiza que los RMP mejoran monótonamente la aproximación a la solución del problema.

La definición de la región factible del RMP \hat{X} se ha definido tradicionalmente de dos maneras: i) mediante la envoltura convexa de un subconjunto de las columnas generadas o ii) mediante la envoltura afin de un conjunto de las columnas generadas, intersección con la región factible. Ambas definiciones garantizan la satisfacción de la propiedad (1.13) y de la exigencia $\hat{X} \subset X$ debido a que los esquemas de descomposición simplicial han sido aplicados a regiones factibles X convexas.

- Problema de generación de columnas. Esta aproximación interior es mejorada (aumentada) mediante la generación de un vector (o columna) en el conjunto factible, a través de la resolución de otra aproximación al problema de optimización original pero en este caso es la función objetivo la que es aproximada.

$$\underset{x \in X}{\text{minimizar } \hat{f}(x)}, \quad [\text{CG}(\hat{f}, X)]$$

Los subproblemas CG en los que no se obtienen necesariamente puntos extremos, no se almacena la columna generada sino su prolongación a la frontera (relativa)

$$\hat{y} := \hat{x} + \ell(\hat{y} - \hat{x}), \quad \text{donde } \ell := \max\{\ell \mid \hat{x} + \ell(\hat{y} - \hat{x}) \in X\}, \quad (1.14)$$

siendo respectivamente \hat{x} y \hat{y} la última solución al RMP y al subproblema CG. La motivación de esta operación de prolongación radica en que el número de variables del RMP se mantiene constante pero produce una región factible más amplia.

Cabe destacar tres etapas en el desarrollo de los algoritmos simpliciales

- Subproblemas generadores de columnas en el conjunto de puntos extremos. En esta primera etapa, los algoritmos son desarrollados para problemas con restricciones lineales con estructura especial, como las de red, lo que conduce a problemas CG lineales con una estructura especial, que permite su resolución con especializaciones eficientes del método simplex. Los RMP en las primeras iteraciones tienen muy pocas variables y una estructura muy simple en las restricciones (combinación convexa) lo que permite la aplicación eficiente de algoritmos de convergencia superlineal.

La experiencia con el método de descomposición simplicial (SD) ha mostrado que hace rápidos progresos inicialmente, y rápidamente alcanza una solución casi óptima, especialmente cuando se aplica un algoritmo de convergencia cuadrática para resolver los RMP, pero esta eficiencia se reduce cerca de la solución óptima. El algoritmo SD es menos eficiente para problemas en los que la dimensión de la *cara óptima del problema* es elevada. Esto es debido a que, en estas circunstancias, el número de puntos extremos, necesarios para expresar una solución óptima como combinación convexa de ellos, es elevado y, por tanto, también lo es el número de iteraciones necesarias para identificar gran parte de ellos. Esto conduce a que se deba resolver gran cantidad de RMP y que los últimos posean gran cantidad de variables.

Una primera solución a este problema fue propuesta por Hearn et al. (1985), Hearn et al. (1987). Estos autores introducen un parámetro r para limitar el número de columnas almacenadas en el RMP. Cuando el número de columnas retenidas en el RMP alcanza esta cantidad, la última columna generada se intercambia por la columna de menor peso en la combinación convexa de la última solución del RMP. Esta solución no es enteramente satisfactoria ya que requiere elevados valores del parámetro r para mantener una convergencia superlineal presente en el esquema original de SD.

Los problemas de flujos en redes multiproducto suelen tener dos formulaciones alternativas, una en el espacio de flujos en los arcos y otra en el espacio de flujos en los caminos. El espacio de flujo en los caminos tiene una estructura de producto cartesiano, donde cada conjunto es el espacio de flujo para cada producto. Larsson and Patriksson (1992) consideran una SD en el espacio de flujos en los caminos, lo que denominan *descomposición simplicial desagregada* (DSD) donde, en lugar de de envoltura convexas para definir los RMP. Este algoritmo resuelve un mucho menor número de RMPs pero con un mayor número de variables, pero el resultado global presenta mejores resultados computacionales que los anteriores.

Marín (1995) especializa el RSD para problemas con restricciones laterales. En este algoritmo el problema maestro se obtiene como la intersección de un símplice con el conjunto definido por las restricciones laterales. El subproblema lineal es

modificado para que la función objetivo recoja una estimación de los multiplicadores de Lagrange de las restricciones laterales.

- Subproblemas generadores de columnas en la frontera relativa. La explicación de la conducta de los algoritmos SD y RSD se encuentra en que emplean una aproximación de primer orden de f para definir los problemas CG. El problema CG en estos dos algoritmos coincide con el del algoritmo de Frank Wolfe. Es sabido que la calidad de estas direcciones de búsqueda se deteriora rápidamente. La razón es que la sucesión de derivadas direccionales $\{\nabla f(\hat{x}^t)^T d^t\}$ en las direcciones de búsqueda $d^t := \hat{y}^t - \hat{x}^t$ tienden a cero pero la sucesión $\{d^t\}$ no converge a 0; lo que implica que las direcciones de búsqueda tienden rápidamente a ser ortogonales al gradiente de f y por tanto la calidad de las columnas generadas (entendida como la mejora inducida en el RMP) será rápidamente reducida. Una conclusión natural es emplear en el problema CG una mejor aproximación de f , de ello se podría esperar una mejor calidad en las columnas y por tanto una mejor aproximación interior en el RMP. Larsson et al. (1997), basándose en esta observación, extendieron el algoritmo RSD a métodos de generación de columnas no lineales. Esta clase se denomina *descomposición simplicial no lineal (NSD)*. Estos métodos poseen menor sensibilidad a la dimensión de la cara óptima debido a que emplean un menor número de columnas para describir la solución óptima. Esto hace que los métodos requieran un menor número de iteraciones y permite elegir un valor menor del parámetro r . Como los problemas CG son no lineales, las columnas ya no son necesariamente puntos extremos. Estos autores prolongan la columna obtenida a la frontera relativa.
- Algoritmos generadores de puntos en la frontera relativa. La convergencia del algoritmo NSD puede ser establecida mediante resolución truncada tanto del problema CG como del RMP. Este resultado tiene una gran importancia práctica pero mucho más desde un punto de vista teórico. Esto permitió en García (2001), García et al. (2003) considerar el propio problema original como la mejor aproximación al mismo.

La diferencia sustancial con los anteriores métodos de descomposición simplicial radica en la definición del problema CG. En los esquemas anteriores el problema CG se interpreta como una aproximación al problema original. El nuevo esquema, denominado CGA, construye las columnas resolviendo aproximadamente el problema original a través de la realización de un número de iteraciones de algún algoritmo eficiente. En este contexto, el énfasis se sitúa en la elección de los onal.

Los algoritmos CGA pueden ser vistos como algoritmos modulares de programación matemática no lineal, donde se dispone de un algoritmo para resolver eficientemente el RMP y de otro para resolver el problema original. Este último algoritmo es el procedimiento para obtener las columnas en el CGP. Esta clase puede ser interpretada como un principio para acelerar la convergencia de un al-

goritmo convergente de puntos factible (algoritmo empleado en el CGP) mediante un esquema de descomposición simplicial.

En el trabajo García (2001) se hace un estudio computacional de la clase de algoritmos CGA empleando dos problemas de flujos en redes no lineales. El primero es el problema uniprodueto de flujos en redes no lineales con restricciones de capacidad y el segundo es un problema de flujos multiprodueto para la asignación en redes multimodales de transporte. La contribución fundamental de este estudio numérico, es que se ha mostrado que la clase CGA mejora sustancialmente, y en algunos casos espectacularmente, los algoritmos de descomposición simplicial RSD, SD y NSD.

La tabla 1.2 describe las características de los principales algoritmos simpliciales desarrollados en la literatura. En la primera columna es el acrónimo del método, en la segunda se indica que tipo de región factible posee el problema de interés, en la tercera columna se indica que operación (envoltura convexa o afín) se efectúa sobre las columnas actualmente retenidas para definir la región factible del RMP, en la cuarta columna se indica si el método aplica o no la operación de prolongación descrita en (1.14) y en la última columna la función objetivo del subproblema CG.

Tabla 1.2: Características de los algoritmos simpliciales

Algoritmo	X	\hat{X}	Regla (1.14)	$\hat{f}(x)$
SD	poliedro	env. convexa	no	$\hat{f}(x) = \nabla f(\hat{x})^T x$
RSD	poliedro	env. convexa y restricción n ^o columnas por r	no	$\hat{f}(x) = \nabla f(\hat{x})^T x$
DSD	poliedro con estructura producto cartesiano	prod. cartesiano de env. convexa	no	$\hat{f}(x) = \nabla f(\hat{x})^T x$
NSD	convexo	env. convexa o env. afín	si	$\hat{f}(x) = \nabla f(\hat{x})^T x + \varphi(x, \hat{x})$ donde φ es continua, respecto a la variable x es convexa y continuamente diferenciable y cumple que $\varphi(x, x) =$
		env. convexa o env. afín	si	$\nabla f(x) - \nabla f(\hat{x})$

El algoritmo SD también ha sido extendido a problemas con restricciones no lineales. Ventura and Hearn (1993) abordan la adaptación del RSD a problemas con restricciones convexas (RSDCC). El problema CG es transformado en un problema lineal mediante

la utilización de la aproximación lineal a trozos de las restricciones no lineales, tal como en el algoritmo de Topkis-Veinott Topkis and Veinott (1967).

Una combinación de la programación cuadrática secuencial (SQP) y del algoritmo NSD se desarrolla en Patriksson (1998). En este método, los problemas CG reemplazan las restricciones no lineales con aproximaciones lineales y su curvatura se tiene en cuenta, en la función objetivo, a través de la información derivada de las variables duales.

La clase de métodos de descomposición simplicial también ha sido extendida al problema más general de *desigualdades variacionales* (VIP) (ver por ejemplo la excelente monografía sobre el tema de Patriksson (1998)) y a la programación matemática convexa *no diferenciable*. Larsson et al. (1997), Larsson et al. (1999).

Como ya se ha mencionado anteriormente, las malas propiedades matemáticas hacen que la resolución de los MPEC sea muy difícil. Incluso si esto fuese posible las grandes dimensiones de los MPECs que aparecen en las aplicaciones hacen inviables estos procedimientos. En la mayoría de las aplicaciones los métodos empleados tienen una naturaleza heurística.

La línea de investigación seguida en esta tesis doctoral ha sido la adaptación de los algoritmos de generación de columnas a los problemas MPEC. La aplicación de los algoritmos simpliciales a los MPECs pasan por reinterpretar los subproblemas CG, no como la aplicación de algoritmos convergentes al problema a resolver, sino como un mecanismo de identificación de los puntos extremos de la cara óptima del problema MPEC.

En los algoritmos de generación de columnas aplicados al MPEC, el subproblema CG obtendría una identificación del conjunto de restricciones activas en la solución óptima, $\mathcal{I}(x^*)$, mientras que el RMP consistiría en el problema original restringido al conjunto de puntos extremos contenidos en la actual estimación de $\mathcal{I}(x^*)$.

Muchas son las preguntas que se deben contestar, para disponer de un algoritmo válido para las aplicaciones y, muchas de ellas, pasan por realizar un análisis de la sensibilidad del conjunto $\mathcal{I}(x)$ a perturbaciones de la variable x , con vistas a poder garantizar la convergencia a óptimos locales. Estas ideas se han desarrollado en esta tesis doctoral aplicadas al problema DAP y a uno nuevo denominado PCPCA (Cap. 4).

Durante los últimos años, se ha observado un interés creciente por los métodos DAP, es decir, por *métodos para la estimación de las matrices de O-D que usan conteos de tráfico*. Estos métodos combinan la información obtenida en los recuentos de tráfico con otras informaciones disponible, en orden a corregir y mejorar un cálculo inicial aproximado de la matriz de viaje O-D, a veces llamada *target-matrix* O-D. En general,

la *target-matrix* O-D se obtiene aplicando los métodos de estimación directos o los modelos de demanda de viajes.

Como ya se ha esbozado, la necesidad de obtener matrices de viajes de forma rápida y económica fue detectada en los primeros años de la década de 1970, donde se comenzó con las aproximaciones teóricas a la construcción de matrices de viaje a partir de información fácilmente disponible obteniendo los volúmenes de tráfico en los arcos de la red de carreteras. En la mayoría de las áreas urbanas, estos datos eran recolectados de forma económica usando sensores de tráfico. Esencialmente, las aproximaciones existentes para sintetizar matrices de viaje a partir de conteos en los arcos se pueden clasificar en dos amplias categorías O'Neill (1987):

- o Técnicas de calibración de parámetros
- o Métodos de estimación de matrices.

Las *técnicas de calibración de parámetros* usan el análisis de regresión, lineal o no lineal, para construir modelos de demanda, asumiendo un modelo de flujos tipo gravitatorio, en orden a estimar las entradas de la matriz de viajes. Sin embargo, estas técnicas requieren datos de zonas para la calibración de los parámetros de los modelos de demanda, y son, por tanto, de limitado uso práctico, porque no solamente son los datos de zona menos fácilmente disponibles, sino también porque se quedan anticuados relativamente pronto. Willumsen (1978), Nguyen (1984), y O'Neill (1987) revisaron la literatura relacionada con esta clase de procedimientos.

Por su parte, los *métodos de estimación de matrices* solamente necesitan de conteos de tráfico en los arcos y una información a priori de la estructura de la matriz de viaje.

Se han desarrollado y calibrado varios métodos para estimar estas matrices a partir de conteos de tráfico. Para poder adentrarnos en ellos, consideremos una representación de la red de transporte de una región. El sistema de transporte consiste en un conjunto de arcos dirigidos que están conectados unos con otros en los nodos. La región se divide en distintas zonas que se representan mediante centroides en los cuales el tráfico se origina o termina. Un conjunto de arcos dirigidos \mathcal{A} conecta el sistema de nodos \mathcal{N} , siendo los centroides un subconjunto de estos nodos. Suponemos que disponemos de datos de conteo de tráfico para un subconjunto $\hat{\mathcal{A}}$ de arcos (ejemplo cada hora, cada día). El problema de la estimación es equivalente a encontrar una matriz O-D razonable g que, cuando sea asignada a la red de transporte, reproduzca los datos alcanzar exactamente en todos los conteos de tráfico. Esto se puede explicar por inconsistencias internas en los conteos de tráfico, debido a datos de circulación en distintas horas o a la agregación de la red de transporte para su estudio.

El problema de estimar la matriz O-D puede ser visto como el problema inverso al de asignación: *Encontrar una matriz O-D que, cuando sea asignada a la red, reproduzca los conteos de tráfico observados.*

Normalmente habrá un gran número de matrices O-D que reproduzcan los conteos de tráfico observados, es decir, el sistema de ecuaciones resultante es indeterminado y puede tener muchas soluciones posibles.

El número de pares O-D (elementos de la matriz O-D), normalmente excede, en mucho, el número de arcos para los cuales se han recogido los conteos de tráfico. El problema sería elegir la *mejor matriz* de entre todas las posibles que reproducen los conteos de tráfico detectados.

Algunos modelos solucionan este problema postulando un modelo general para la distribución de los viajes, por ejemplo, un modelo gravitatorio, mientras que otros adoptan técnicas de inferencia estadística.

En los arcos de la red de transporte puede existir congestión y esto se puede modelar por funciones de coste. Las funciones de coste describen la relación entre el volumen de tráfico en el arco y el coste (o el tiempo) para recorrerlo. El tratamiento de los efectos de la congestión es una característica importante para distinguir entre los modelos para la evaluación de la matriz O-D. Cada modelo asume que la congestión se puede tratar exógenamente (por asignación proporcional) o endógenamente (por asignación en equilibrio). El caso exógeno es más adecuado para bajos niveles de congestión o para buenas predicciones de los niveles de congestión. El caso endógeno es más normal en presencia de congestión significativa. La mayoría de los modelos se basan en costes separables en los arcos; también se han manejado casos más generales, donde los flujos en arcos vecinos afectan al coste de un arco, mediante el uso de modelos con desigualdades variacionales.

Generalmente, todos los modelos para la estimación de la matriz O-D utilizan información anterior sobre esa matriz O-D. Esta información puede estar disponible en términos de una *target matrix* O-D y/o en el número de viajeros que se atrajeron o se originaron en diversas zonas. La *target matrix* O-D se puede obtener por una encuesta sobre una muestra o de una vieja matriz (probablemente desactualizada). La existencia y la tipología de una *target matrix* O-D es, también, una segunda característica importante que distingue a los modelos para la estimación de la matriz O-D.

Por tanto, los modelos se pueden clasificar en función de si están o no desarrollados para considerar la congestión y de si utilizan o no *target matrix* O-D.

Se puede observar que la mayoría de los modelos existentes asumen o requieren que esté disponible una *target matrix* O-D. También se puede observar que en los últimos años el énfasis se ha puesto en el estudio de redes congestionadas.

de tráfico para un solo periodo. Para poder entender los modelos vamos a introducir la notación que usualmente utilizan. Para ello hemos de considerar que un detalle crucial, en la estimación de una matriz O-D mediante el uso de conteos de tráfico, es la técnica de asignación usada: qué caminos de la red de transporte se utilizan para realizar viajes de la zona p a la zona q . La matriz P con elementos p_i^q representa la *proporción de viajes* entre la zona p y la zona q que utilizan el arco a ,

$$0 \leq p_i^a \leq 1$$

Para un arco dado a , la suma de todos los g_i (Flujos O-D desde i a j) que atraviesan este arco es el volumen o flujo del arco, v_a . En las ecuaciones fundamentales, que relacionan los volúmenes en los arcos y los flujos O-D, se incluye, a veces, un factor R de ocupación de los coches (asumido, generalmente, en 1.25)

$$v_a = \frac{1}{R} \sum_i p_i^a g_i, \quad a \in A \quad (1.15)$$

donde A es el conjunto de arcos de la red de transporte. La matriz compuesta por los elementos p_i^a es, a menudo, designada por matriz P .

Dependiendo del tratamiento de la congestión y de la determinación endógena o exógena de la matriz O-D, se postula:

Asignación proporcional: En este caso se asume independencia entre los volúmenes de tráfico y las proporciones de tráfico p_i^a . Los volúmenes en los arcos (v_a) son proporcionales a los flujos O-D (g_i). La proporción de viajeros que eligen una ruta no dependerá de la congestión de la red y si de las características de la ruta y del viajero. Los valores de p_i^a pueden determinarse antes de que la estimación de la matriz O-D sea realizada. Puede usarse un método de asignación "todo o nada": La asignación de tráfico *todo o nada* se realiza cuando todo el tráfico, para todos los pares O-D, es asignado por las rutas de mínimo coste.

Asignación en equilibrio: Cuando los efectos de la congestión sean importantes, la asignación en equilibrio es un enfoque más realista. El coste de viajar por un arco depende del volumen de flujo mediante una relación costo-flujo. El valor de p_i^a dependerá del volumen en todos los arcos, por ejemplo, $p_i^a = p_i^a(v)$, y no puede ser determinado independientemente del proceso de estimación de la matriz de viajes. Las ecuaciones (1.15) se transforman en no lineales en v .

Un patrón de flujo, con flujo en los arcos v_a ($a \in A$), y una matriz de asignación mencionada $P(v)$, que cumple los requisitos de asignación en equilibrio según el primer principio de Wardrop, se reconoce como del tipo de *equilibrio de usuario* (user-equilibrium). Los usuarios se comportan buscando el óptimo para cada usuario. Alternativamente, puede ser empleado un concepto estocástico de equilibrio que tiene en cuenta que los viajeros tienen percepciones diferentes del coste del viaje, por ejemplo, debido a variaciones individuales. El resultado es que se usan múltiples rutas con desiguales costes. La complejidad computacional de un modelo de estimación de la matriz O-D depende, en su mayor parte, de la técnica de asignación utilizada. Utilizar una técnica de asignación en equilibrio y un explícito tratamiento de los efectos de la congestión es mucho más gravoso que recurrir a la asignación proporcional.

Dados unos volúmenes de tráfico \hat{v}_a observados en un subconjunto de arcos \hat{A} y unas proporciones de tráfico p_i^a , la matriz O-D g se determina resolviendo el sistema

de ecuaciones (1.15). Este sistema de ecuaciones es normalmente altamente indeterminado: hay muchos más elementos en la matriz O-D g que arcos donde se realizan los conteos. Se necesitan datos adicionales (información anterior) y/o realizar determinadas suposiciones sobre el comportamiento de los viajeros para encontrar una única matriz O-D.

Una de las partes más importantes de la información previa o de partida es la *target matrix* O-D. En los *modelos estadísticos* la target matrix O-D puede obtenerse de un simple estudio y se toma como una observación de la "verdadera" matriz O-D a estimar. Se asume que la verdadera matriz O-D pertenece a alguna distribución estadística y que pueden ser obtenidas por estimación de los parámetros de esta distribución. En los *modelos* de tráfico basados en *programación matemática* la target matrix se asume que es una vieja matriz O-D a la que hay que realizar ajustes para que satisfaga los conteos de tráfico observados. La diferencia entre la matriz O-D estimada y la target matrix O-D se minimiza bajo restricciones de flujo. Ambos enfoques tienen en común que el problema de encontrar una matriz O-D g , dada una target matrix O-D \hat{g} , consiste en minimizar una función $F_1(g, \hat{g})$. Bajo esta notación, toda la información previa la contiene la target matrix O-D \hat{g} .

Estadísticamente los datos del conjunto de conteos de tráfico se consideran como una observación de los verdaderos datos de conteo a estimar, relacionados con, y obtenidos como, una asignación de la matriz O-D estimada. También, por otras razones antes mencionadas, las desviaciones entre los conteos estimados y los conteos observados pueden aceptarse. Esta concepción contrasta con la suposición alternativa de la reproducción exacta de los volúmenes de tráfico observados.

Por lo tanto, para una matriz O-D se intentan producir diferencias "pequeñas" entre los flujos estimados en los arcos v y los flujos observados \hat{v} . Esta intención se puede expresar como función $F_2(v, \hat{v})$ para ser minimizada conforme a las restricciones de la asignación.

Para la estimación de la matriz O-D como un *problema de optimización*, el problema se puede expresar en general de la siguiente forma:

Determine la demanda para el tráfico entre las distintas zonas de la región, es decir, determine la matriz O-D g , que solucione el programa de optimización siguiente:

$$\begin{aligned} &\text{Minimizar } F(g, v) = r_1 F_1(g, \hat{g}) + r_2 F_2(v, \hat{v}) \\ &\text{sujeto a:} \\ &\quad g \in G \end{aligned} \quad \text{MAP1}$$

donde \hat{g} es la target matrix O-D y \hat{v} son los conteos de tráfico observados en un subconjunto de arcos, siendo F_1 y F_2 ciertas métricas. La asignación de g a la red de transporte se denota como $\mathcal{A}(g)$, repartándose los flujos O-D (g_i) sobre las rutas o caminos disponibles con flujos en dichos caminos representados por (h_w) .

Si la target matriz O-D \hat{g} es muy fiable y si el parámetro τ_1 fuera muy grande comparado con τ_2 , el resultado sería una g muy próxima a \hat{g} . En este caso, serían aceptadas grandes desviaciones entre v y \hat{v} .

Si, por el contrario, los conteos de tráfico observados fueran tan fiables como la información contenida en la target matriz O-D \hat{g} , entonces las magnitudes de los parámetros τ_1 y τ_2 deberían ser parecidas. La segunda parte de DAP dirigiría entonces la optimización, conduciendo a que los flujos estimados (v) estuvieran cerca de los valores observados (\hat{v}), mientras que serían aceptadas desviaciones más grandes entre la matriz O-D estimada (g) y la información a priori (\hat{g}). Por esto, los valores relativos de los pesos de los parámetros τ_k están muy relacionados con la concepción del modelo utilizado. En Brenninger-Gothe et al. (1989) se discuten modelos generales multicriterios, para el caso de la asignación proporcional.

La métrica F_1 es elegida del tipo información mínima o máxima entropía y la F_2 se toma, a menudo, como una medida euclidiana de la distancia. En principio, se puede concebir cualquier otra combinación de métricas.

Los modelos incluidos en la categoría presentada anteriormente, basados en la programación matemática, incluyen enfoques típicos de los modelos de "información mínima" ("maximizando la entropía") y combinan modelos para la planificación del tráfico. Una estimación de la matriz O-D se obtiene por solución directa (F_1 = medida de la entropía) o por estimación de los parámetros del modelo combinado.

La otra categoría, enfoque mediante inferencia estadística, incluye la Máxima Verosimilitud (ML-Maximum Likelihood), Mínimos Cuadrados Generalizados (GLS-Generalized Least Squares) y Técnicas de Inferencia Bayesiana. Aquí los volúmenes de tráfico y la target matriz O-D se considera que pertenecen a alguna distribución de probabilidad. La target matriz O-D \hat{g} es normalmente obtenida en los modelos de programación matemática como una antigua matriz O-D (desactualizada), mientras que las aproximaciones estadísticas confían en una target matriz O-D obtenida mediante simples encuestas.

Se han propuesto técnicas de solución basadas en los gradientes, que resuelven problemas de optimización obtenidos desde los modelos de programación matemática o desde las aproximaciones mediante inferencia estadística. Estas técnicas son importantes porque se asumen como unas técnicas muy eficientes para encontrar soluciones y pueden ser aplicadas a problemas de gran escala basados en asignación en equilibrio.

Las primeras revisiones fueron publicadas por Nguyen (1984) y Cascetta and Nguyen (1988). Cascetta y Nguyen se centran en técnicas de inferencia estadística, mientras que las revisiones de Cascetta y Nguyen no proporcionan ningún enfoque particular, pero sí una estructura general para el problema de estimar la matriz O-D a partir de conteos de tráfico. Nguyen (1984) solo considera tráfico de vehículos particulares (coches), mientras que Cascetta and Nguyen (1988) también consideran el transporte público. En Fisk (1989) se comparan tres modelos de maximización de entropía y, bajo ciertas

condiciones, se muestran los resultados en estimaciones equivalentes. También se tratan algunos aspectos de su implementación y computación.

Teniendo en cuenta lo anterior, a continuación vamos a estudiar los distintos modelos clasificándolos en dos categorías: Modelos basados en programación matemática y modelos basados en inferencia estadística, al final presentaremos una revisión sobre las técnicas de solución que se han propuesto para estos modelos.

Sin embargo, esta clasificación no es única, así, en Florian (1986) los distintos modelos se clasifican en tres grupos distintos: Modelos de equilibrio en redes, modelos de gravedad-entropía y modelos que combinan asignación con distribución.

1.3.1 Formulación matemática de los modelos basados en programación matemática.

Con estos modelos se intenta determinar la matriz de viajes más idónea que sea consistente con la información contenida en los datos del volumen de tráfico en los arcos, a la vez que, en todo caso, maximiza la entropía o usa una función objetivo basada en la mínima información con respecto a la matriz de viajes previa. Willumsen (1978) presenta la aproximación con elementos de la máxima entropía siguiendo a Wilson (1970), a la vez que Van Zuylen (1978) discute los principios de mínima información.

Como la información proporcionada por los contadores de tráfico en algunos arcos es insuficiente para determinar una única matriz O-D, es posible argumentar que se puede elegir una matriz O-D de "mínima información". Esta es una matriz O-D que utiliza tan poca información como puede de la contenida en la target matrix O-D, mientras que toma las ecuaciones que relacionan los conteos de tráfico observados con los volúmenes O-D estimados. La matriz de mínima información estimada es obtenida por minimización de la función Z_I que corresponde a F_1 en [DAP] (ver, por ejemplo, Snickars and Wibull (1977) para una derivación combinatoria):

$$Z_I = \sum_{i \in I} g_i \ln \left(\frac{g_i}{\hat{g}_i} \right) \quad (1.16)$$

Esta es la función de información mínima o máxima entropía ($\hat{g}_{ij} = 1$). Una matriz O-D que minimice (1.16), mientras que reproduce los conteos de tráfico y tiene en cuenta la información contenida en la target matrix \hat{g} puede ser derivada como:

$$g_i = \hat{g}_i e^{\lambda_1 p_i^1 + \lambda_2 p_i^2 + \dots + \lambda_K p_i^K} \quad (1.17)$$

donde cada λ_i es un multiplicador de Lagrange asociado con la restricción que relaciona el flujo en el arco con la matriz de viajes, ecuación (1.15). La expresión (1.17) para g asume una asignación proporcional, por ejemplo considera p_i^a como constante.

Van Zuylen and Willumsen (1980) propusieron dos modelos importantes de este tipo. Obtienen una matriz O-D que, asignada a la red de tráfico, reproduce los conteos de tráfico observado e imponen que sea consistente. Los modelos basados en programación matemática asumen, directa o indirectamente, que el comportamiento de los viajes es representado por cierto modelo de distribución de viajes. Los modelos de Van Zuylen y Willumsen se basan en principios de información mínima y entropía máxima y conducen a modelos de distribución de viajes del tipo gravedad utilizando asignación proporcional. En la parte teórica, se asume una target matriz \hat{g} , pero en todas las aplicaciones presentadas, todas las combinaciones origen destino se suponían igualmente probables, es decir, $\hat{g}_i = 1$.

Muchos investigadores, tales como Hall et al. (1980), Van Zuylen (1981), Van Vliet and Willumsen (1981), Willumsen (1982), Willumsen (1984), Bell (1983a), Bell (1983b), Nguyen (1984), y Hammerslag and Immers (1988), han realizado tests o han propuesto mejoras y refinamientos para este tipo de modelos.

Fisk (1988) mostró como combinar la estrategia de máxima entropía de Van Zuylen y Willumsen y el concepto de asignación, desde el punto de vista del óptimo de usuario, en un mismo modelo aplicado al caso con congestión. Los conteos de flujo se incluyen en el modelo como restricciones de equilibrio respecto al usuario, y entonces, la existencia de una matriz de demanda de flujos O-D, que reproduzca exactamente los flujos observados en los arcos, no puede ser garantizada.

Brenninger-Gothe et al. (1989) sugieren una técnica multiobjetivo sin preferencias para alcanzar un compromiso, entre la separación de la solución a los flujos observados y a la matriz O-D previa, desde un enfoque de mínimos cuadrados generalizados.

Smith (1979) ha demostrado que el modelo del comportamiento óptimo para el usuario se puede expresar con desigualdades variacionales, según se muestra más adelante. El modelo propuesto tiene una estructura binivel, en el nivel superior se maximiza la entropía y en el nivel inferior se resuelve un problema de equilibrio desde el punto de vista del usuario:

$$\begin{aligned}
 \text{Minimizar } Z &= \sum_{i \in I} g_i (\ln(g_i) - 1) \\
 \text{sujeto a: } C(h) \cdot (\bar{h} - h) &\geq 0, \quad \forall \bar{h} \in \Omega_h \\
 \sum_{k \in K_i} h_k &= g_i, \quad \forall i \in I \\
 \sum_{i \in I} \delta_{ik}^a h_k &= \hat{v}_a, \quad \forall a \in \hat{A} \\
 h_k &> 0
 \end{aligned} \tag{1.18}$$

donde $C(h)$ es el coste de viaje en los caminos, h es el flujo en los caminos y \bar{h} es cualquier solución factible del flujo en los caminos. Si el patrón observado de flujo es un patrón de flujo de equilibrio, desde el punto de vista del usuario, el modelo de entropía extendida de Fisk tendrá la misma solución que un *modelo combinado de distribución y asignación*. Esto se muestra en Fisk (1989), donde, además, se concluye que para las formulaciones basadas en el equilibrio en redes, en la máxima entropía, y en la

combinación de distribución y asignación se puede esperar que produzcan los mismos resultados, bajo condiciones de congestión en la red, siempre que los volúmenes de tráfico observados en los arcos correspondan a un patrón de flujo en equilibrio.

Para los modelos combinados existen algoritmos de resolución eficientes. Los primeros trabajos fueron el modelo de viajes combinando distribución-asignación de Erlander et al. (1979) y el modelo combinado de Fisk and Boyce (1983) donde proponen una combinación de técnicas de distribución y de asignación basadas en el modelo presentado por Erlander et al. (1979), la cual también descansa en los datos de flujo en los arcos. Estos modelos combinados toman el número de viajes que se originan en, o se atraen a, cada zona dada (O_p y D_q respectivamente) y esto puede ser expresado como restricciones. Una diferencia importante entre estos modelos y el de Fisk (1988) se manifiesta en las restricciones de asignación. Aunque los conteos de tráfico observados no aparecen en el modelo combinado, si se utilizan para determinar el parámetro μ en la fase de estimación. Uno de los, por esto, llamados modelos "dobles restringidos" considerados por Fisk y Boyce tiene la siguiente formulación como modelo de optimización:

$$\begin{aligned}
 \text{Minimizar } Z &= \mu \sum_a \left[\int_0^{v_a} c_a(x) \cdot dx \right] + \sum_{i \in I} g_i \ln(g_i) \\
 \text{sujeta a: } & \sum_{p \neq q} g_{pq} = O_p, \quad \forall p \quad (\alpha_p), \\
 & \sum_{p \neq q} g_{pq} = D_q, \quad \forall q \quad (\beta_q), \\
 & \sum_{k \in K_i} h_k = g_i, \quad \forall i \in I \quad (\chi_i), \\
 & \sum_i \delta_{ik}^a \cdot h_k = v_a, \quad \forall a \in \hat{A}, \\
 & g_i, h_k \geq 0
 \end{aligned} \tag{1.19}$$

donde, entre paréntesis, se indican los multiplicadores de Lagrange. La solución al problema de minimización (1.19) puede ser obtenida tomando el Lagrangiano del problema e igualando a cero las primeras derivadas del Lagrangiano respecto a h_k y g_i :

$$\sum_a s_a(v_a) \cdot \delta_{ik}^a - \chi_i = 0, \quad \forall k \in K_i; h_k > 0 \tag{1.20}$$

$$\frac{\partial Z}{\partial g_i} = \ln(g_i) + 1 - \chi_i = 0 \tag{1.21}$$

$$g_i = e^{-\alpha_p - \beta_q} e^{-\mu \chi_i} = A_p O_p B_q D_q e^{-\mu \chi_i} \tag{1.22}$$

Las ecuaciones (1.20) y (1.21) estipulan que las rutas que están actualmente usadas, $h_k > 0$, tendrían un coste mínimo para el recorrido entre las zonas que representa

el par i . Los multiplicadores de Lagrange χ_i son así iguales a los costes mínimos de viaje. Desde ahora renombraremos nuestro conjunto χ_i como \bar{c}_i , los mínimos costes en la red. La ecuación (1.22) incluye los multiplicadores de Lagrange α_p y β_q que son determinados para satisfacer las restricciones de producción y atracción de viajes. Los factores de peso A_p y B_q son multiplicadores de Lagrange transformados, $A_p O_p = e^{-\alpha_p}$ y $B_q D_q = e^{-\beta_q}$. El valor del parámetro μ es indeterminado en la formulación del problema presentado y se calcula en la fase de estimación.

La importancia de este modelo proviene de la relativa facilidad con la que puede ser resuelto un modelo combinado de distribución y asignación. Erlander et al. (1979) prueban que el modelo puede ser rigurosamente estimado y que el correspondiente modelo de optimización tiene una solución única si la suma de las integrales de todas las funciones de costo en los arcos es conocida:

$$\hat{C} = \sum_{a \in A} \int_0^{\hat{v}_a} c_a(x) dx \quad (1.23)$$

Los conteos de tráfico observados son reproducidos, por un modelo combinado, si están disponibles conteos para todos los arcos y si estos son consistentes con un equilibrio de usuario. Se han realizado otros refinamientos al modelo, por ejemplo por Erlander, Jörnsten y Lundgren en un artículo técnico escrito en 1984 (Erlander et al. (1984)). Se estudia el problema de la falta de datos en la base anual usada en la fase de estimación. Los autores resuelven este problema por la predicción de los valores que van a ser usados en una predicción anual.

Si los conteos de tráfico observados solamente están disponibles para un subconjunto de arcos de la red, Fisk and Boyce (1983) sugieren cómo puede ser obtenida una estimación de \hat{C} balanceando los costes medios con la importancia de varios tipos de arcos en la representación de la red. El modelo de Fisk y Boyce no reproducirá, en general, los conteos de tráfico observados. Esto puede relacionarse, en cierto modo, con el hecho de que los conteos de tráfico observados pueden no constituir un patrón de flujos en equilibrio. Se han propuesto diferentes sugerencias sobre procedimientos que eliminen inconsistencias internas, pero no procedimientos efectivos para asegurar la propiedad de equilibrio de usuario y la consistencia interna del patrón de flujos observados.

En Kawakami et al. (1992), el modelo combinado de Fisk y Boyce se extiende aplicación a una red de mediano tamaño en la ciudad de Nagoya en Japón. Parece ser que se ha dispuesto una *target matrix*, pero no ha sido incluida en la formulación del modelo.

En Tamin and Willumsen (1989) se considera un modelo gravitacional y un *modelo de intervención oportuna*, ambos modelos incorporan restricciones en la demanda, tanto en los orígenes como en los destinos. Esta metodología la han aplicado a pequeños

problemas test, sin congestión, usando un modelo de gravedad, un modelo de intervención oportuna y un modelo mixto de gravedad-oportunidad.

Nguyen (1977) (y más tarde Nguyen (1984)) presentó una de las primeras formulaciones del problema de estimación de la matriz O-D basado en el equilibrio. La red era congestionada y Nguyen analizó a fondo las propiedades de la solución obtenida. La solución reproducirá los conteos de tráfico observados. Aunque la solución es única, desde el punto de vista de las variables de flujo en los arcos, normalmente existen muchas diferentes matrices O-D que se corresponden con los flujos estimados en los arcos. Es necesario elegir un criterio para determinar una matriz única de todas las diferentes matrices O-D que reproducen los conteos observados. Turquist and Gur (1979) presentan un técnica heurística e iterativa de solución para el modelo de Nguyen basada en el algoritmo de Frank-Wolfe, y Gur et al. (1980) inventan un método similar con su bien conocido sistema LINKOD. Este modelo LINKOD ha sido extensamente testado y validado por Han et al. (1981) y Hand and Sullivan (1983). Para todas estas técnicas, pueden existir diversas matrices O-D que reproduzcan los volúmenes de tráfico observados, y por eso, usualmente se especifica una target matriz como vía de solución para elegir la que mejor dentro del patrón de flujos.

Jörnsten and Nguyen (1979) y más tarde LeBlanc and Farhangian (1982) comenzaron asumiendo la asignación en equilibrio y los modelos formulados de los tipos de maximización de la entropía y mínimos cuadrados generalizados, respectivamente. El primer trabajo no requiere de target matriz O-D mientras que el modelo de LeBlanc y Farhagian también requiere conocer una target matriz O-D. El desarrollo del modelo empieza con el trabajo sugerido por LeBlanc en Gur et al. (1980). La formulación de mínimos cuadrados está motivada en la obtención de una matriz O-D que teniendo un comportamiento de usuario-óptimo se encuentre lo "más cercana" (en el sentido de GLS) a la target matriz.

$$\begin{aligned}
 & \text{Minimizar} && \sum_{i \in I} (g_i - \hat{g}_i)^2 \\
 & \text{sujeto a:} && \sum_a \int_0^{v_a} c_a(x) dx - \sum_{i \in I} \bar{c}_i g_i = Z, \\
 & && \sum_{k \in K_i} h_k = g_i \quad \forall i \in I, \\
 & && \sum_{i \in I, k \in K_i} \delta_{ik}^a h_k = v_a \quad \forall a \in A, \\
 & && g_i, h_k \geq 0
 \end{aligned} \tag{1.24}$$

(1977). Este es un problema de demanda variable. Los test presentados en estos dos trabajos son sobre pequeños problemas.

La mayoría de las formulaciones del problema de estimación de la matriz O-D en ambiente de congestión tienen una estructura binivel. El problema de nivel *superior* estima la matriz O-D (asumiendo un volumen de flujo en los arcos dado) y el problema de nivel *inferior* es un problema de asignación en equilibrio que determina un patrón

de flujo en los arcos según el tipo de equilibrio. En el nivel inferior la demanda de viajes, la matriz O-D, se asume dada.

1.3.2 Formulación matemática de los modelos basados en inferencia estadística

Máxima verosimilitud

El enfoque de máxima verosimilitud (ML) maximiza la verosimilitud entre la target matriz O-D y los conteos de tráfico observados condicionados sobre la verdadera (estimada) matriz O-D. Se asume que los elementos de la target matriz O-D \hat{g} son obtenidos como observaciones de un conjunto de variables aleatorias. Los conteos de tráfico observados \hat{v} constituyen otra fuente de información sobre g (la matriz O-D a estimar) y \hat{v} y \hat{g} son considerados normalmente como estadísticamente independientes. La verosimilitud de las observaciones \hat{g} y \hat{v} puede ser expresada como:

$$\mathcal{L}(\hat{g}, \hat{v}|g) = \mathcal{L}(\hat{g}|g) \cdot \mathcal{L}(\hat{v}|g) \quad (1.25)$$

Debido a la asunción de independencia entre los conteos de tráfico observados y la target matriz O-D, la verosimilitud observada por ambos conjuntos es igual al producto de las dos verosimilitudes. Aplicar el principio de máxima verosimilitud para este problema equivale a seleccionar la matriz O-D g que maximice esta verosimilitud. Con la convención de que $0 \cdot \ln(0) = 0$, podemos también maximizar el logaritmo del producto.

Si la target matriz O-D es obtenida por muestreo aleatorio simple, en una región con un patrón de viajes estable, se puede asumir que la target matriz O-D sigue una distribución multinomial. Esta es dependiente de pequeñas fracciones de muestreo α_i : Si \hat{g}_i viajes son muestreados de un total de g_{ip} viajes del origen p entonces $\alpha_i = \frac{\hat{g}_i}{g_{ip}}$. Para el logaritmo de la probabilidad $\mathcal{L}(\hat{g}|g)$ tenemos:

$$\ln \mathcal{L}(\hat{g}|g) = \sum_i (\hat{g}_i \ln(\alpha_i g_i)) + \text{constante} \quad (1.26)$$

.....
suficientemente grandes, para la target matriz O-D se puede asumir una distribución de probabilidad de Poisson y para el logaritmo de $\mathcal{L}(\hat{g}|g)$ tendremos:

$$\ln \mathcal{L}(\hat{g}|g) = \sum_i (-\alpha_i g_i + \hat{g}_i \ln(\alpha_i g_i)) + \text{constante} \quad (1.27)$$

Si los conteos de tráfico observados también se les asume como generados por una distribución de probabilidad de Poisson e independientes de la target matriz O-D, se obtiene una expresión similar para la probabilidad $\mathcal{L}(\hat{v}|v(g))$ que es F_2 en [DAP]:

$$\ln \mathcal{L}(\hat{v}|v(g)) = \sum_{a \in \hat{A}} (\hat{v}_a \ln(v_a(g)) - v_a(g)) + \text{constante} \quad (1.28)$$

donde $v_a(g)$ representa el volumen de flujo en el arco a resultante de la asignación de g . Si se asume una distribución normal multivariante (MVN⁵) para los términos de error de los conteos de tráfico observados, con media cero y matriz de varianzas-covarianzas W , la expresión de F_2 sería:

$$\ln \mathcal{L}(\hat{v}|v(g)) = -\frac{1}{2} (\hat{v} - v(g))^t W^{-1} (\hat{v} - v(g)) + \text{constante} \quad (1.29)$$

Si (1.27) y (1.28) y la asignación proporcional son suposiciones válidas, el problema de estimación de la matriz O-D puede ser formulado como:

$$\begin{aligned} \text{Maximizar} \quad & \sum_{i \in I} (-\alpha_i g_i + \hat{g}_i \ln(\alpha_i g_i)) + \sum_{a \in \hat{A}} (\hat{v}_a \ln(v_a(g)) - v_a(g)) \\ \text{sujeto a:} \quad & \sum_{i \in I} p_i^a g_i = v_a, \forall a \in \mathcal{A}, \\ & g_i \geq 0 \end{aligned} \quad (1.30)$$

Este es uno de los problemas de optimización considerado por Spiess (1987). Para resolver el problema (1.30), Spiess propone un algoritmo de tipo cíclico coordinado ascendente. Los test ejemplos de Spiess son de pequeño tamaño.

En el caso de asignación en equilibrio, el problema (1.30) tendrá una estructura binivel con el problema de asignación en el nivel inferior y el de estimación de la matriz O-D en el nivel superior.

Mínimos cuadrados generalizados

La target matriz O-D \hat{v} se puede suponer obtenida por estimación, como la "verdadera" pueden ser vistos como obtenidos de una ecuación estocástica:

$$\begin{aligned} \hat{g} &= g + \eta \\ \hat{v} &= v(g) + \varepsilon \end{aligned} \quad (1.31)$$

⁵Multivariate normal distribution

donde η es el error probabilístico que relaciona \hat{g} con g y ε el error que relaciona los conteos de tráfico observados \hat{v} con $v(g)$. A menudo se asume que tanto η como ε tienen por media cero, es decir, que $E(\eta) = 0$ y $E(\varepsilon) = 0$.

Nótese que, en la derivación del estimador de mínimos cuadrados generalizados (GLS) bajo la suposición de no distribución, se necesita tener los términos aleatorios η y ε de (1.31). Hay solamente un requerimiento sobre la existencia de matrices de varianza-covarianza. En la ausencia de matrices de varianza-covarianza exactas, a menudo se han utilizado matrices unitarias (con los elementos de la diagonal igual a 1). Existen un par de artículos sobre la sensibilidad del modelo GLS para usar matrices de varianza-covarianza aproximadas. Cascetta (1984) sugiere una alternativa para seleccionar la mejor matriz origen destino de todas las posibles y concluye que la estimación, e incluso la 'aproximación dura', de la matriz de varianza-covarianza produjo mejores resultados que un estimador de máxima entropía. Esta independencia de las distribuciones supuestas es una importante ventaja de las aproximaciones por GLS. Según una experiencia, parece ser que los modelos son mucho más sensibles a la variación e inexactitudes en los datos de los conteos de tráfico y a la target matriz O-D que a los valores de los parámetros, ver, por ejemplo, Cascetta (1984) o Bierlaire and Toint (1994). Los parámetros incluyen las matrices de varianza-covarianza Z y W usadas en la formulación del problema (1.32) presentado más adelante. Como en las aproximaciones de Máxima Verosimilitud, la target matriz O-D y los conteos de tráfico observados son considerados mutuamente independientes. Si la target matriz O-D \hat{g} tiene un error con una matriz de varianza-covarianza Z y la matriz de varianza-covarianza de los conteos de tráfico es W , el estimador GLS puede ser obtenido resolviendo:

$$\begin{aligned} &\text{Minimizar } \frac{1}{2} (\hat{g} - g)^t Z^{-1} (\hat{g} - g) + \frac{1}{2} (\hat{v} - v(g))^t W^{-1} (\hat{v} - v(g)) \\ &\text{sujeto a: } g_k \geq 0 \end{aligned} \tag{1.32}$$

La matriz O-D g estimada se debe restringir, por supuesto, a no negativa. Cascetta (1984) desarrolla expresiones para la media y la varianza del estimador GLS cuando las restricciones de no negatividad no están activas. Una propiedad importante de la aproximación por GLS es que las dos fuentes de información, los conteos de tráfico observados y la target matriz O-D, se combinan fácilmente. Por ejemplo, si la matriz de varianza-covarianza es cercana a cero, reflejando una gran confianza en esta parte de la información, la matriz inversa es muy grande. Esto significa que los pesos de las

La matriz de varianza-covarianza Z puede ser aproximada de diferentes formas. Si se usa, para fijar el origen, un muestreo aleatorio simple, se puede desarrollar una aproximación que puede llegar a ser insuficiente (Véase, por ejemplo, Cascetta and Nguyen (1988))

Bell (1991) presenta un algoritmo para resolver el problema (1.32) Este algoritmo

explícitamente considera la restricción de no negatividad para la matriz O-D estimada g . Ambos, Cascetta (1984) y Bell (1991) asumen asignación proporcional. Bell deriva la solución para el problema de estimación (1.32) para ser:

$$g = \hat{g} + Z P^t (P Z P^t + W)^{-1} (\hat{v} - P \hat{g}) + \left(Z - Z P^t (P Z P^t + W)^{-1} P Z \right) \mu \quad (1.33)$$

donde P es la matriz de la asignación y μ son los multiplicadores de Lagrange asociados con las restricciones de no negatividad de la matriz O-D g . En (1.33) se puede apreciar que la target matriz O-D es 'ajustada', 'cambiada' por dos términos. El primer término se relaciona con la desviación entre los conteos de tráfico observados y lo obtenidos por una asignación de la target matriz O-D. El segundo término se relaciona con las restricciones activas de no negatividad, es decir, restricciones con multiplicadores de Lagrange no nulos que corresponden a elementos de g_i iguales a cero.

Si asumimos que los conteos de tráfico y la target matriz O-D (sus términos de error η y ε) siguen una distribución normal multivariante (MVN), se puede mostrar que el estimador GLS coincide con el estimador ML. Este mismo resultado se produce también utilizando la aproximación por inferencia Bayesiana bajo la suposición de que los volúmenes de tráfico y la target matriz O-D siguen una distribución MVN (Maher (1983)). En un temprano artículo, Bell (1984) muestra que la aproximación por mínima información de Van Zuylen and Willumsen (1980) es aproximada por la aproximación por GLS si los conteos de tráfico son conocidos en un alto grado.

Yang et al. (1992) extienden el modelo GLS mediante la integración de la asignación en equilibrio de la matriz O-D en el modelo. El problema de estimación de la matriz O-D es formulado como un programa binivel con el problema de mínimos cuadrados generalizados en el nivel superior y el problema de asignación en equilibrio en el nivel inferior. El problema resultante llega a ser difícil de resolver. Se sugiere un algoritmo heurístico y se aplica a pequeños problemas.

Si los datos de tráfico observados son del tipo equilibrio de usuario y están disponibles para todos los arcos, se puede obtener una matriz O-D que reproduzca los volúmenes observados en los arcos como una solución de un sistema de ecuaciones indeterminado. La dificultad se encuentra en determinar si los datos de tráfico observados son del tipo equilibrio de usuario. Si el sistema de ecuaciones indeterminado es compatible, Yang et al. (1994) muestran que los flujos de tráfico son del tipo equilibrio de usuario. El usuario es, sin embargo, difícil y no se ha sugerido ningún procedimiento definitivo.

En Yang (1995) vuelve a usar un enfoque mínimos cuadrados generalizados, para resolver el problema en redes congestionadas, mediante una formulación binivel y algunos factores de influencia que tienen en cuenta los resultados obtenidos en ambos niveles, primer y segundo nivel. En este artículo, Yang extiende el modelo para acomodarlo al caso de interacciones de flujo en determinadas funciones de coste en los

arcos, Maher and Zhang (1999) proporcionan más conocimientos sobre la naturaleza de las soluciones producidas por los enfoques de Yang et al. (1992) y Yang (1995).

Inferencia Bayesiana

Las aproximaciones por inferencia bayesiana consideran la target matriz O-D como una función de probabilidad a *priori* $Pr(g)$ de la matriz O-D estimada g . Si los conteos de tráfico observados son considerados como otra fuente de información sobre g con una probabilidad $\mathcal{L}(\hat{v}|g)$, entonces el teorema de Bayes proporciona un método para combinar las dos fuentes de información. Para la probabilidad a *posteriori* $f(g|\hat{v})$ de observar g condicionado a los conteos de tráfico observados tendremos:

$$f(g|\hat{v}) \approx \mathcal{L}(\hat{v}|g) \cdot Pr(g) \quad (1.34)$$

La función de probabilidad a posteriori permite, en principio, la determinación de una región de confianza para g . Pero debido a problemas computacionales, en la práctica, solamente pueden ser obtenidos estimadores de puntos. Esta puede tomar la forma del valor máximo del logaritmo de la distribución a posteriori, la g que maximiza $\ln f(g|\hat{v})$. Para el primer término de (1.34), los conteos de tráfico observados, se asume, usualmente, una distribución de probabilidad de Poisson o una MVN. Las expresiones para el logaritmo de $\mathcal{L}(\hat{v}|g)$ serán, entonces, (1.28) o (1.29). Para la función de probabilidad $Pr(g)$ puede asumirse una distribución multinomial. Entonces, para el logaritmo de $Pr(g)$, usando la aproximación de Stirling tendremos:

$$\ln Pr(g) = - \sum_i g_i \ln \left(\frac{g_i}{\hat{g}_i} \right) + constante \quad (1.35)$$

Esta es una función de información mínima. También se puede obtener una función similar con la aproximación de Poisson de la distribución multinomial. Si asumimos una distribución normal multivariante para calcular $Pr(g)$, con media q y matriz de varianza-covarianza M_q , Maher (1983) obtiene:

$$\ln Pr(g) \approx - \frac{1}{\sigma} (g - q)^t Z_q^{-1} (g - q) + constante \quad (1.36)$$

Maher asume una asignación proporcional y los conteos de tráfico, supone que siguen una MVN y muestra que, en este caso, la matriz O-D estimada también llega a estar distribuida como una MVN. El modelo de Maher es aplicado a una red de transporte muy pequeña. Los problemas de optimización de las aproximaciones por inferencia Bayesiana contienen, igual que las aproximaciones por ML y GLS, la suma de dos partes. La primera se relaciona con la target matriz O-D \hat{g} , (1.26) o (1.27), y la

segunda concierne a los conteos de tráfico observados. Las aproximaciones por inferencia bayesiana son unas técnicas de inferencia estadística con propiedades comunes con las aproximaciones por ML y GLS. Pero, como Cascetta y Nguyen mostraron, los roles asumidos por g en las aproximaciones por inferencia clásica (ML y GLS) y por inferencia Bayesiana difieren. En el primer caso, los verdaderos g_{ij} son parámetros de la función de verosimilitud $\mathcal{L}(\hat{g}, \hat{v}|g)$ y en el segundo caso, los g_{ij} son variables aleatorias con una distribución dada a priori.

Otros enfoques

Casi todos los trabajos donde se utiliza asignación en equilibrio, esta se hace desde el punto de vista de óptimo para el usuario (DUE). Sin embargo, relativamente pocos trabajos se han enfrentado con el problema considerando redes congestionadas con el criterio de asignación estocástico. Bell et al. (1996) proponen entremezclar, un estimador del flujo en los caminos con las medidas de flujo en los arcos, con un modelo logit de elección de caminos, formulado como un problema de optimización.

Maher and Zhang (1999) y Zhang et al. (1999) presentan el problema de la estimación O-D como un programa binivel, donde el nivel superior se corresponde con el problema de estimación de la matriz (MATE), mientras que el nivel inferior se corresponde con un equilibrio de usuario estocástico (SUE) basado en las proporciones de elección de arco o en los flujos de arco. Se han desarrollado métodos iterativos basados en resolver alternativamente el problema MATE y el problema SUE. Estos algoritmos se han testado en dos redes moderadamente grandes y han mostrado ser bastante efectivos en la práctica.

En un contexto diferente pero relacionado, Bell et al. (1996) han propuesto un enfoque de programación no lineal para estimar el flujo en los caminos SUE para sistemas con intersecciones reguladas por semáforos.

Finalmente, también en la literatura se incluyen algunos enfoques basados en inteligencia artificial para desarrollar matrices O-D (Chin et al. (1994); Xu and Chan (1995)). Sin embargo, la aplicabilidad de estos métodos a redes urbanas reales falta por ser testada.

origen destino basándose únicamente en la información del volumen de tráfico en un subconjunto de arcos. Esta consideración introduce no linealidades en la función de coste del modelo, por la dependencia entre el coste en los arcos con volumen en los arcos, y requiere la determinación de una solución de punto fijo (más que un óptimo) para el modelo propuesto. Tal punto fijo se determina heurísticamente por aproximaciones iterativas al modelo no lineal, usando una secuencia de programas lineales.

1.3.3 Algoritmos para la resolución del problema de estimación de la matriz O-D

Este es el aspecto que más se ha tratado y más profundamente ha sido estudiado por la comunidad científica. Esto es debido fundamentalmente a dos causas, la primera es que la resolución del modelo DAP es la base para la aplicabilidad de esta metodología y la segunda es que el DAP tiene una estructura de problema de programación matemática con restricciones de equilibrio. Esta estructura binivel puede ser utilizada en otras aplicaciones que van desde los problemas de desregulación de sistemas, de gestión de la energía, telecomunicaciones, etc., como otros problemas de producción y distribución de recursos, hasta problemas de diseño en ingeniería.

Los modelos de programación matemática con restricciones de equilibrio (MPEC) son extremadamente difíciles de resolver. Este hecho se deriva de la complejidad de su conjunto factible, que denotamos por \mathcal{F} . Esta complejidad es debida, fundamentalmente, a diversas causas:

- o *No convexidad de \mathcal{F}* . La región factible puede dejar de ser convexa, aunque todas las funciones y conjuntos que intervengan en su definición lo sean.
- o *No es un conjunto cerrado*. El conjunto factible puede ser incluso no cerrado. Esta propiedad puede hacer peligrar hasta la existencia de soluciones.
- o *Naturaleza multievaluada de la función $\Omega^*(x)$* (Conjunto de soluciones del nivel inferior para un valor x de las variables del nivel superior). En muchas aplicaciones $\Omega^*(x)$ es un conjunto que contiene más de un elemento para cada valor de la variable x del nivel superior.
- o *No diferenciabilidad de $\Omega^*(x)$* . La situación más favorable es que la aplicación multievaluada $\Omega^*(x)$ defina una aplicación, en este caso dicha función puede ser no diferenciable.
- o *Pérdida de la propiedad de conexidad*. La región factible puede ser la unión de varios conjuntos disjuntos. Esta propiedad le confiere una naturaleza combinatoria, con restricciones disyuntivas, que son de gran complejidad computacional, incluso en problemas lineales binivel.

Se podría pensar que estas patologías están presentes exclusivamente en situaciones excepcionales y que la exigencia de ciertas propiedades sobre las funciones y conjuntos exigente de *programación lineal binivel*, donde todas las funciones son indefinidamente diferenciables y convexas, y todos los conjuntos son poliedros, aparecen todas las anteriores patologías. Además, Jeroslow (1988) y Hansen et al. (1992) han demostrado que el problema lineal binivel pertenece a la clase de problemas *NP-duros*, es decir, que no existe actualmente un algoritmo que pueda resolver el problema en un número polinomial de operaciones en función de las dimensiones del mismo.

Pese a la dificultad intrínseca de los problemas binivel se han desarrollado algoritmos exactos para problemas de muy pequeña dimensión. Para resolver el problema lineal binivel se han aplicado algoritmos basados en enumeración implícita, *branch-and-bound*, penalizaciones exactas, o métodos de descomposición. Bi et al. (1989), Bialas and Karwn (1982), Bialas and Karwn (1984), Hansen et al. (1992), Júdece and Faustino (1992), White and Anandalingan (1993).

El caso binivel no lineal ha sido abordado en la monografía de Shimizu et al. (1997), en la que desarrollan dos algoritmos para un caso particular de la programación convexa binivel (todas las funciones que definen el problema son convexas). El primer algoritmo está basado en el trabajo previo de Bard (1988) y aborda el caso particular de que el nivel inferior es un problema cuadrático convexo, y el nivel superior es un problema estrictamente convexo. Shimizu et al. (1997) demostraron que este algoritmo converge a la solución óptima. El segundo algoritmo que presentaron se basa en el trabajo de Jaumard et al. (1995) y está desarrollado para el caso anterior y tiene garantizada la convergencia en un número finito de iteraciones.

Luo et al. (1996) proponen tres tipo de algoritmos iterativos, el primero (PIPA) está basado en métodos de penalización interior, el segundo en una programación implícita y el tercero es una especialización de la programación secuencial cuadrática (SQP) empleada para resolver problemas de optimización no diferenciables. Los problemas que debe resolver este algoritmo son problemas diferenciables a trozos.

Los dos primeros algoritmos convergen a puntos estacionarios bajo la cualificación de restricciones de independencia lineal, mientras que el tercero tiene una convergencia local, esto es, depende de la proximidad entre el punto inicial y un punto estacionario. Por contra, tiene una velocidad de convergencia superlineal e incluso cuadrática en algunos casos. Luo et al. (1996) realizaron un estudio computacional basado en el paquete MATLAB para pequeños ejemplos test.

Sin embargo, las técnicas, o soluciones algorítmicas, utilizadas para la resolución de los modelos binivel planteados hasta el momento por la comunidad científica, en el terreno de la ingeniería del transporte, difícilmente pueden encuadrarse dentro de los métodos generales propuestos para la resolución del MPEC descritos anteriormente. A continuación revisamos estos métodos.

A *grosso modo* una clasificación para los modelos MPEC aplicados en la planificación del transporte conduce a dos grandes grupos:

- *Problemas de estimación de matrices, precios destino (DAP)*, en este modelo las restricciones del nivel inferior son independientes de las variables del nivel superior.
- o *Modelos de gestión /diseño de redes de transporte urbano (NDP)*, en este modelo las restricciones del nivel inferior son independientes de las variables del nivel superior pero la función objetivo, al contrario de lo que ocurría en el caso primero, está parametrizada por ellas.

Esta taxonomía induce una clasificación en los algoritmos de resolución desarrollados ya que éstos están fuertemente condicionados por el problema a resolver. No obstante, existen ejemplos en el que el mismo tipo de algoritmo ha sido empleado en ambos contextos.

Una dificultad añadida es la gran dimensión de estos problema en las aplicaciones reales. Por ejemplo, si se considera la estimación de matrices O-D en las redes de Madrid o Barcelona el número de variables del nivel inferior serían respectivamente 8.659 y 2.522, mientras que el número de variables en el nivel superior sería de 26.037 y 7.922 respectivamente. En otro tipo de aplicaciones el número de variables en el nivel superior varía entre unas decenas a varios miles. Otra forma de entender el coste computacional es observando que la evaluación de la función objetivo puede llevar algunos minutos, debido a que tiene que resolver un problema de optimización de miles de variables. Esto ha hecho que, en la actualidad, no se hayan implementado algoritmos exactos para problemas reales de gran escala. Casi la totalidad de los algoritmos desarrollados son heurísticos y son escasas las pruebas computacionales realizadas en problemas de grandes redes.

El algoritmo heurístico más extendido es el denominado algoritmo iterativo de *optimización-asignación*. Una iteración de este problema tiene dos fases. En la primera se resuelve el problema del nivel inferior (*problema de asignación en equilibrio*) y en la segunda se resuelve el problema del nivel superior (*problema de optimización*) considerando que las variables nivel inferior están fijadas a los valores encontrados en la primera fase. Este algoritmo fue originariamente propuesto por Allsop (1974) para un problema de control de tráfico y por Steenbrink (1974) para un problema de diseño de redes. Tan et al. (1979) demostraron computacionalmente que este algoritmo no tiene garantizada su convergencia a un punto estacionario y Marcotte (1981) demostró teóricamente este hecho. No obstante, este algoritmo es el más extendido, y ha sido formulado en todos los modelos binivel aplicados al transporte, por ejemplo Hall et al. (1980), Yang et al. (1992) aplican una modificación de este algoritmo al problema de estimar matrices O-D.

Centrándonos en los problemas tipo DAP, estos fueron abordados inicialmente por Spiess (1990b) proponiendo un método heurístico, tipo gradiente, para resolver el problema binivel, método que, aunque de débil base teórica, proporciona buenos resultados prácticos. Diversos investigadores han buscado justificaciones teóricas que expliquen los resultados (Chen (1994))), así como algoritmos alternativos para la resolución del problema.

particularmente (Chen (1994)), en el que se utiliza la función objetivo del problema inferior, o una aproximación lineal del problema del nivel inferior en un punto determinado que reemplaza al problema inferior, lo que conduce a un problema de un solo nivel, cuya solución proporciona el próximo punto donde repetir el proceso. Esta clase de algoritmos heurísticos es conocida como *formulación implícita*. Yang (1995) utiliza esta metodología para desarrollar dos tipos de aproximaciones lineales (*factores de influen-*

cia) que son aplicados a la estimación de matrices O-D, pero sólo se efectúan pruebas computacionales en redes de pequeñas dimensiones. Friesz et al. (1990), Yang and Lam (1996) lo aplican a modelos de gestión de tráfico.

En las técnicas de solución basadas en gradientes, la target matriz O-D es tomada como una solución inicial para el problema de estimación de la matriz O-D. La target matriz O-D es 'ajustada' o 'cambiada' para reproducir los conteos de tráfico siguiendo "direcciones" calculadas iterativamente basándose en el gradiente de la función objetivo. Los volúmenes en los arcos son funciones implícitas de los flujos O-D y se obtienen por un procedimiento de asignación, $v(g) = \mathcal{A}(g)$, del tipo equilibrio de usuario. Esto significa que el problema de estimación de la matriz O-D (usando conteos de tráfico) puede ser formulado únicamente en términos de las variables g_i :

$$\begin{aligned} \text{Minimizar } & F(g) = \eta_1 F_1(g, \hat{g}) + \eta_2 F_2(v(g), \hat{v}) \\ \text{sujeto a: } & g \geq 0 \end{aligned}$$

donde las F_i son medidas apropiadas de distancia (métricas). Excepto por las restricciones de no negatividad de g , este problema es no restringido. Drissi-Kaitouni and Lundgren (1992) proponen un algoritmo general de descenso para resolver el problema anterior. Florian and Chen (1993) estudian un método de tipo Gauss-Seidel y uno del tipo Lagrangiano Aumentado. De todos estos, el anterior es el más conveniente para su aplicación a grandes redes. También Yang et al. (1992) han sugerido un método del tipo Gauss-Seidel. El algoritmo de Spiess es más aproximado y esto hace funcione bien que en las aplicaciones prácticas presentadas.

En teoría todas las contribuciones, excepto para la de Spiess, consideran ambas funciones F_1 y F_2 en el nivel superior. En las aplicaciones a redes que tengan un considerable tamaño, presentadas por Spiess (1990a) y Drissi-Kaitouni and Lundgren (1992) solamente aparece F_2 en la función objetivo. Chen (1994) y un artículo comparativo de Denault (1994) contienen resultados donde también las variables g_i de la matriz de viajes son explícitamente consideradas en la función objetivo del nivel superior. Aunque se obtiene una más alta calidad en los resultados incluyendo la target matriz O-D, una ventaja de esta aproximación es que más 'tratable' computacionalmente. Spiess presenta aplicaciones a problemas de gran escala. Los problemas incluyen una aplicación urbana a la ciudad de Bern en Suiza, con cerca de 2700 arcos y una aplicación interurbana por la red de carreteras de Finlandia con unos 12500 arcos. El método de Spiess es heurístico, puesto que asume una asignación proporcional que en cada del problema de optimización indicado. Con esta asunción el gradiente de la función objetivo llega a ser fácil de computar, alcanzable desde la solución de dos problemas de asignación en equilibrio. La relación cercana a uno de los métodos sugeridos por Drissi-Kaitouni and Lundgren (1992) es mostrada por estos últimos. Las direcciones de descenso empleadas en el método de Gauss-Seidel de Florian and Chen (1993) puede ser interpretada como más elaboradas y por lo tanto proporcionan una solución menos

aproximada al problema de estimación. Como mostraron Florian and Chen (1993), esta interpretación también se aplica al método de Gauss-Seidel de Yang et al. (1992).

Chen y Florian (1994) desarrollan las condiciones de optimalidad para resolver el problema DAP en redes congestionadas con un modelo de asignación DUE, y proponen un método de descenso coordinado que garantice la observancia de las condiciones de optimalidad, aunque la convergencia no está garantizada. Florian and Chen (1995) han propuesto una formulación similar al heurístico tipo Gauss-Seidel para la formulación binivel del problema de estimación de la matriz O-D.

Los resultados obtenidos por Spiess son razonables, con una mejora significativa de la calidad. El método de Spiess está disponible en el paquete comercial para la planificación del transporte Consultants (1990). Los métodos de Drissi-Kaitouni and Lundgren (1992) se han aplicado a problemas como la ciudad de Hull en Canadá, con alrededor de 800 arcos. Drissi-Kaitouni and Lundgren (1992) investigan diferentes gradientes basados en direcciones de descenso posibles escaladas por información de segundo orden. Los resultados indican que el esfuerzo computacional es razonable. El énfasis está en la calidad de las direcciones de búsqueda y no en lograr un estimador computacionalmente eficiente de la matriz O-D. Chen (1994) aplica un método del tipo Gauss-Seidel para la red de mediano tamaño de Winnipeg en Canadá, con unos 2983 arcos, 154 centroides y con conteos en el 2.3 por ciento de los arcos. La inclusión explícitamente de las variables de la matriz O-D, es decir, F_1 , consigue resultados más precisos.

Estos métodos heurísticos no tienen en cuenta la falta de diferenciabilidad del problema, incluso en la situación favorable que se presenta cuando el problema de nivel inferior (problema de asignación) presenta unicidad de soluciones. Estas técnicas se basan en la evaluación de las proporciones de uso de los arcos por las relaciones O-D de la red. En otros casos se evitan explícitamente los puntos de no diferenciabilidad y evalúan los gradientes siguiendo las técnicas desarrolladas por Fiacco (1983), Kolstad and Lasdon (1986) y Friesz et al. (1990), si bien estas técnicas, de carácter analítico, no permiten el tratamiento de redes de grandes dimensiones.

Se han empleado *métodos de búsqueda probabilística* como el *simulado recocido*. Este método requiere de numerosas evaluaciones del nivel inferior, esta desventaja es aprovechada para hacer una resolución muy eficiente del nivel inferior. Friesz et al. (1992) lo aplican a problemas de diseño de redes de tráfico mientras que García and Marín (2002) lo emplean en un problema binivel para el diseño de las capacidades y tarifas de aparcamientos disuasorios. Otro tipo de metaheurísticos como los algoritmos

Se han probado otro tipo de algoritmos que utilizan directamente un método de optimización sobre el problema binivel. Abdulaal and LeBlanc (1979) aplican el algoritmo de Hooke-Jeeves al problema de diseño de redes. Recientemente, Codina and Montero (2001) han desarrollado un método computacionalmente eficiente, basado en la regularización de la función objetivo de nivel superior para aproximar direcciones contenidas en el gradiente generalizado de Clarke, para el caso en que el problema de

asignación es separable y diagonal. Este método permite la utilización de algoritmos de descenso y es consistente con la no diferenciabilidad del problema, estando en curso la extensión de dicho método para el caso de la asignación asimétrica general formulada en el terreno de las desigualdades variacionales.

Por otra parte, Patriksson and Rockafellar (2002) adaptan el algoritmo PIPA para un problema binivel aplicado a la gestión del tráfico.

Bell et al. (1997) muestran un estimador del flujo en los caminos apropiado para usarse en conjunción con la monitorización, el control y la guía del tráfico urbano. El tiempo de viaje, en cada arco de la red, se divide en un tiempo fijo más un tiempo de retardo fruto de la congestión. Se asume que los arcos serán de dos tipos. Para el primer tipo de arcos, se dispone de una estimación externa de sus flujos y tiempos de viaje. El segundo tipo de arcos se caracteriza por una capacidad finita, y el retardo surge cuando la demanda se acerca, si no supera, a la capacidad. La demanda se determina mediante un modelo de elección de ruta tipo logit. Se formula un problema de programación convexa equivalente y un proceso iterativo de solución. Se discute la estimación del parámetro de dispersión del modelo logit, y se propone un método de generación de columnas para evitar la enumeración de los caminos.

Yin (2000) presentan una aproximación basada en algoritmos genéticos para resolver eficientemente los modelos binivel. El funcionamiento del algoritmo se ilustra y compara con los algoritmos previos basados en el análisis de la sensibilidad usando ejemplos numéricos.

Yang et al. (2001) proponen un modelo de optimización para la estimación simultánea de la matriz origen destino (O-D) y del coeficiente de dispersión, basándose en un equilibrio estocástico de usuario tipo logit. El modelo se formula como problema de optimización estándar, diferenciable, y no lineal con las restricciones analíticas del equilibrio estocástico de usuario. Se derivan expresiones explícitas de las derivadas de las restricciones del equilibrio estocástico de usuario, con respecto a la demanda origen-destino, flujo en los arcos y coeficientes del coste de viaje, que son computadas eficientemente a través de una aproximación estocástica a la carga de la red. Para resolver el modelo de estimación simultánea, se aplica sucesivamente un algoritmo de programación cuadrática secuencial (SQP). Este algoritmo converge, bajo ciertas condiciones, a un punto de Karus-Kuhn-Tucker del problema.

Maher et al. (2001) se ocupan de dos problemas similares en el análisis de redes de transporte: la estimación de la matriz de viajes y la optimización del tráfico de señales en una red congestionada. Estos dos problemas son formulados como problemas de en la programación del problema de segundo nivel. Estos autores distinguen dos tipos de soluciones en la estimación combinada de la matriz y el problema de asignación estocástica desde el punto de vista del usuario: Una es la solución del problema de programación binivel y la otra sería una solución mutuamente consistente, donde los dos subproblemas del problema combinado se resuelven simultáneamente.

Kim et al. (2001) desarrollan un método de solución, con un algoritmo genético, para obtener una solución global que comparan con las obtenidas por el SAB. Finalmente muestran, con los resultados de ejemplos numéricos, que el algoritmo propuesto es superior al SAB con independencia de los patrones de recorrido.

Nie and Lee (2002) sugieren que un estimador lineal del flujo en los caminos puede ser resuelto independientemente mediante el empleo del algoritmo de ordenación de caminos⁶. Este mecanismo encuentra caminos (columnas) en equilibrio de usuario y una descomposición que especifica la más verosímil matriz de viajes O-D. Presentan una versión modificada del algoritmo mencionado para garantizar que todas las columnas-caminos cíclicos y libres, que satisfacen el UE, son reconocidos. El método propuesto descompone en dos simples problemas individuales el modelo de estimación O-D convencional, basado en el equilibrio, en el que la reproducción del patron de flujos en equilibrios y la estimación de la matriz O-D se ejecutan iterativa y simultáneamente. La aproximación por descomposición asigna flujos directamente sobre los caminos que se han asegurado como unos de los caminos óptimos.

Patriksson and Rockafellar (2003) muestran cómo algunos de los modelos de desigualdades variacionales, sobre conjuntos poliédricos, pueden ser puestos en una formulación, agregada o desagregada, relacionadas por una transformación afín de la variable. Para tales problemas, los autores establecen que los resultados del análisis de la sensibilidad, bajo parametrizaciones, ni confían en las propiedades estrictamente monótonas del problema, en términos de variables desagregadas, ni en una particular elección de sus valores en la solución. También muestran cómo utilizar la transformación afín para desarrollar herramientas computacionales para calcular resultados de sensibilidad y aplicarlos al análisis de la sensibilidad de demandas elásticas de problemas de tráfico en equilibrio.

Sherali et al. (2003) desarrollan una investigación para sintetizar los flujos O-D basándose solamente en la información de un conjunto parcial de volúmenes en los arcos. Este enfoque introduce no linealidades en la función de coste del modelo debido a la dependencia de los costes de viaje en los arcos de los volúmenes en los arcos, y requiere la determinación de una solución de punto fijo para el modelo propuesto. Tal punto fijo se determina heurísticamente aproximando iterativamente el modelo no lineal, usando programación lineal.

Dempe (2003) presenta una revisión de las recientes investigaciones para resolver problemas binivel y para describir las condiciones de optimalidad, además, presenta las particularidades principales de esta clase de problemas.

origen destino para su aplicación en planificación con contadores, en tiempo real, en los arcos durante varios días. Basado en el marco de la estimación binivel iterativa, el problema de nivel superior consiste en minimizar la desviación entre los flujos estimados en los arcos y los conteos en tiempo real en los arcos, y la desviación entre la

⁶K-shortest-paths ranking algorithm

demanda, dependiente del tiempo, y la demanda estática, dada por datos históricos. Estos dos tipos de desviación son combinados en una función objetivo calibrada, donde los distintos pesos se determinan por aproximaciones iterativas para obtener la mejor solución de compromiso. La formulación de día único es ampliada, para usar conteos en los arcos, realizados durante varios días para estimar la variación en la demanda durante varios días.

Bierlaire and Crittin (2004) muestran que el problema de estimar y predecir las matrices origen-destino (O-D) es importante y difícil. En el contexto específico de los sistemas de transporte inteligentes (ITS), la naturaleza dinámica del problema y los requerimientos de datos en tiempo real, lo hacen aún más complicado. Se considera, en este artículo, un acercamiento con un modelo de mínimos cuadrados para resolver el problema de estimación y predicción de la matriz O-D, que parece ofrecer algoritmos convenientes y flexibles. La naturaleza dinámica del problema se representa por un proceso auto-regresivo, capturando las series correlativas de las variables de estado. Esta formulación se inspira en Cascetta, Inaudi and Marquis (1993) y Ashok and Ben-Akiva (1993). Se compara el algoritmo de filtro de Kalman con LSQR, un algoritmo iterativo, propuesto por Paige and Saunders (1982), para la solución de problemas de gran escala con mínimos cuadrados. LSQR explota explícitamente matrices dispersas, permitiendo considerar problemas más grandes, como los que se dan en las aplicaciones reales. Demuestran que el algoritmo LSQR disminuye perceptiblemente el esfuerzo de computación necesitado por la aproximación de filtro de Kalman para problemas de gran escala.

Nakatsuji et al. (2004) muestran como la disponibilidad de los datos sobre giros en las intersecciones es esencial en las simulaciones de tráfico que se realizan con software tal como NETSIM. Sin embargo, la recopilación de estos datos es muy costosa, en términos de tiempo y dinero, incluso para una red pequeña. Para estimar los giros en las intersecciones, se integra un modelo de equilibrio de usuario estocástico (SUE), tipo logit, con un algoritmo genético. Se desarrollan tres modelos derivados: (a) Un estimador doblemente restringido, (b) un estimador simplemente restringido y (c) Un estimador basado en los origen destino (O-D). En los dos primeros estimadores, el modelo SUE se formula como un programa común de distribución de viajes y de asignación de tráfico (TDTA), mientras que en el tercer estimador, el modelo SUE se basó en un programa estándar de distribución O-D. Estos modelos para estimar el giro se examinaron aplicándolos a tres redes de carreteras: una red de carreteras virtual con datos simulados, una red de carreteras pequeña en el campo con datos obtenidos por conteo manual, y una red de carreteras con datos medidos con detectores de tráfico. anterior sobre la distribución O-D, el estimador basado en O-D es el más efectivo en la estimación de los giros. La aplicación de los modelos a redes reales, para los que no había información previa de los flujos O-D, muestra que el modelo doble restringido basado en TDTA es el más exacto y eficiente. Los movimientos de giro en la mayoría de los arcos estimados con este modelo concordaban relativamente bien con los medidos realmente. El coeficiente de correlación del flujo en arcos excedía de 0.90.

Xu et al. (2004) proponen un modelo de programación binivel que actualiza la matriz O-D y calibra la red en una misma etapa. Desarrollan un algoritmo de solución basado en los algoritmos genéticos y presentan ejemplos numéricos para ilustrar el funcionamiento del modelo y del algoritmo de solución propuesto, bajo varias condiciones donde están disponibles diferentes fuentes de datos.

Gao et al. (2004) presentan un modelo de programación binivel para el problema de diseño de redes de tráfico, en el cual el modelo superior es un modelo de diseño de tráfico normal, y el modelo de nivel inferior es un modelo de asignación de tráfico en equilibrio. Se diseña un algoritmo de resolución heurístico basado en el análisis de la sensibilidad para el modelo propuesto.

Meng et al. (2004) presentan un estudio comprensivo de los problemas estáticos de optimización en redes de transporte con restricciones de equilibrio de usuario estocástico. Se demuestra explícitamente que la formulación del problema de punto fijo, en término de flujos de arco, para el problema general de equilibrio de usuario estocástico, en el cual el Jacobiano de la matriz de las funciones del coste de viaje en los arcos puede no ser simétrica, posee una única solución con moderadas condiciones. Desarrollando un método de análisis de la sensibilidad para el problema de equilibrio de usuario estocástico, el estudio prueba que los flujos de equilibrio en los arcos perturbados son funciones implícitas, continuamente diferenciables con respecto a los parámetros de perturbación. Además, presentan una formulación de optimización continua y diferenciable, de un solo nivel, que es equivalente al modelo de programación binivel unificado. Por otra parte, como un método de solución unificado, para resolver los problemas de optimización en redes de transporte con restricciones de equilibrio de usuario estocástico, se usa un algoritmo de programación cuadrática sucesiva basado en el método de análisis de la sensibilidad.

Nie et al. (2004) desarrollan una formulación que incorpora un estimador de flujo en los caminos (PFE) desacoplado en el marco de los mínimos cuadrados generalizados (GLS). La investigación intenta resolver un problema de mínimos cuadrados generalizados (GLS) que minimice la suma de los errores en los conteos de tráfico y en las matrices O-D basado en un mapa de asignación en equilibrio, derivado exógenamente de un procedimiento de ordenación del camino más corto. La resolución del GLS-PFE inevitablemente involucra un sistema lineal no invertible y restricciones de no negatividad. Se diseña un algoritmo de solución para identificar iterativamente las restricciones activas y resolver los sistemas lineales computando la pseudoinversa.

Codina and Barceló (2004) presentan un nuevo algoritmo alternativo para el problema de estimación de la matriz O-D. El algoritmo presentado se basa en un método de optimización no diferenciable, debido a Wolfe, que puede ser interpretado como un método de direcciones conjugadas con mejores propiedades de convergencia, como se muestra en un conjunto de test computacionales. Relacionado de cerca con el algoritmo presentado, se discute el problema de aproximar gradientes de la función del nivel superior y se

presenta y analiza una nueva y consistente aproximación con una descripción detallada de los aspectos algorítmicos implicados.

Patriksson (2004) realiza un análisis completo de la sensibilidad de la demanda elástica del equilibrio del tráfico (Wardrop). Se da una caracterización para la existencia de una derivada direccional de la solución en equilibrio (flujo en los arcos, coste mínimo de viaje, demanda) en alguna dirección, y lo mismo se hace para su gradiente. El gradiente, si existe, es interpretado más a fondo como un caso límite del gradiente de la solución del SUE basado en logit, cuando el parámetro de dispersión tiende a infinito. En ausencia del gradiente, se muestra como computar un subgradiente. Todas estas computaciones (derivada direccional, (sub)gradiente) se realizan mediante la resolución de problemas de equilibrio de tráfico similares con coste en los arcos y funciones de demanda afines, y pueden ser realizados por la misma herramienta (o una similar) como la usada por el modelo de equilibrio de tráfico original; este hecho supone una clara ventaja al aplicar el análisis de la sensibilidad dentro de una aplicación binivel, tales como precios de congestión, estimación O-D, o diseño de redes.

Stathopoulos and Tsekeris (2004) formula la estimación de la matriz de viajes origen-destino (O-D) como un problema de optimización simultánea y se resuelve mediante el empleo de tres algoritmos de optimización meta-heurísticos diferentes. Estos incluyen un algoritmo genético (GA), un algoritmo de recocido simulado (SA), y un algoritmo híbrido (GASA) basado en la combinación de GA y SA. Se evalúa el funcionamiento computacional de los tres algoritmos y se comparan mediante su implementación a una red de tráfico urbano real. Los resultados de los test de simulación muestran que SA y GASA producen soluciones finales más exactas que GA, mientras que GASA muestra un grado superior de convergencia, esto es, una mejora más rápida de la solución inicial, en la comparación con el SA y el GA.

Meng et al. (2004) primero proponen un modelo de programación binivel para estimar simultáneamente la matriz O-D y calibrar parámetros de las funciones de tiempo de viaje en los arcos, utilizando flujos observados bajo la condición de equilibrio determinista desde el punto de vista del usuario. Después, transforman el modelo de programación binivel no diferenciable en un modelo de programación no lineal diferenciable continua, mediante la construcción de una función de mérito (gap function).

Casas (2004), en su tesis doctoral, realiza contribuciones importantes sobre la viabilidad de la arquitectura de los sistemas de gestión avanzada del tráfico y sobre la estimación en tiempo real de matrices O-D mediante conteos de tráfico por redes neuronales. Asimismo muestra la validez de una asignación dinámica de tráfico heurística. Los caminos O-D razonable que son probablemente usados, los cuales pueden ser usados como input en los procesos de estimación, en tiempo real, de matrices O-D a partir de conteos de tráfico y, también, pueden ser usados para validar la arquitectura propuesta en este trabajo, las pautas para la validación de los parámetros de elección de ruta, la forma en que las funciones de coste son actualizadas y la vía en que la asignación

dinámica reactiva es computada.

Clark and Watling (2005) estudian la fiabilidad del tiempo de viaje en una red modelada bajo el supuesto de demanda estocástica. Se propone una técnica para estimar la distribución de probabilidad del tiempo total de viaje en la red, a la luz de las normales variaciones cotidianas en la matriz de demanda de viajes sobre una red de tráfico. Se propone un método de solución, basado en el uso de un solo modelo de asignación de tráfico estándar, el cual opera en dos etapas. En la primera etapa, se calculan los momentos de la distribución total del tiempo de viaje mediante un método analítico, basado en los momentos multivariantes del vector de flujos en los arcos. En la etapa dos, se construye una familia flexible de funciones de densidad para esos momentos. Se discute como la distribución resultante puede ser en la práctica usada para caracterizar la no fiabilidad.

Ceylan and Bell (2005) construyen un algoritmo genético de resolución para las redes de transporte en equilibrio estocástico bajo congestión. Proponen, previamente, unas técnicas de programación binivel mutuamente consistentes (MC), en las cuales se elige un problema de control del tráfico en un área (ATC) como problema de nivel superior, mientras que la asignación de tráfico en equilibrio de usuario se elige como problema de nivel inferior. En este estudio, se ha propuesto un algoritmo genético (GA) para resolver el problema de nivel superior en una red de tráfico con congestión. En el nivel inferior se aplica la asignación de tráfico estocástica en equilibrio de usuario (SUE). En el nivel superior, GA proporciona un conjunto factible de tiempos dentro de los límites especificados para las variables y alimenta el problema de nivel inferior. La asignación SUE se resuelve mediante el estimador de flujo de caminos (PFE) y el modelo de tráfico TRANSYT se aplica al nivel superior para obtener el índice de funcionamiento de la red (PI) y por tanto un índice de su aptitud. Se define el índice de funcionamiento de la red como la suma de una combinación lineal de retrasos y número de paradas por unidad de tiempo, bajo varios niveles de tráfico. Para este propósito, se desarrolla el optimizador genético, nombrado como GATRANSPFE, que combina el modelo TRANSYT, usado para estimar el funcionamiento, con la herramienta de asignación PFE logit, usado para predecir reasignaciones de tráfico.

Wong et al. (2005) presentan un procedimiento para la estimación de matrices origen destino múltiples para una red de tráfico. El sistema consiste en numerar las clases de vehículos. La demanda se asigna a la red en forma de equilibrio de usuario. Se propone un procedimiento de maximización de la entropía para la estimación simultánea de las matrices de demanda O-D para todas las clases de vehículos, que son consistentes con los ~~conteos de tráfico que se observan directamente en la red para todas las clases~~ de tráfico, el método de estimación, y el error asociado de una forma integrada, mientras que consideran los efectos de varias suposiciones hechas para la elección de rutas, por los modelos de asignación, y los niveles de congestión del tráfico en la red. Se demuestran, teórica y numéricamente, algunas propiedades de las reglas de localización de contadores y de las medidas de error asociado para el problema de estimación de la

matriz O-D.

Aboussoror and Mansouri (2005) tratan de los problemas de programación bini-vel lineales débiles. Concretamente analizan la existencia de soluciones mediante un método de penalizaciones. Se refieren a una clase de programas de programación lineal débiles con no única solución para el problema de nivel inferior. Para tales problemas, dan un teorema de existencia de soluciones mediante un método exacto de penalización.

Doblas and Benitez (2005) presentan un procedimiento para estimar y actualizar matrices origen destino basándose en conteos de tráfico preservando la estructura anterior de una matriz conocida. El alcance del artículo es eminentemente práctico. Trata el problema de estimar y actualizar matrices O-D observadas basándose en la información disponible sobre el flujo de arco, por medio de un método de programación no lineal, el denominado método del lagrangiano aumentado (ALM). De esta forma, se puede estimar matrices O-D de forma que se utilice un algoritmo eficiente, que reduzca al mínimo la cantidad de información almacenada requerida, para solucionar problemas de gran tamaño. Se proporciona un algoritmo de solución, que se ha desarrollado y se ha diseñado totalmente para ser utilizado dentro de códigos comerciales de asignación. Las limitaciones y las características de estos programas han condicionado, de forma natural, la solución adoptada para el modelo propuesto.

Ying and Yang (2005) estudian el análisis de la sensibilidad de los flujos en equilibrio de usuario estocástico en una red bi-modal con aplicaciones a precios óptimos. Se estudian los métodos de análisis de la sensibilidad para los sistemas de transporte que tienen una red de tráfico para automóviles y otra de transporte público físicamente separadas. Para el caso en que ambas redes, automóvil y transporte público, sean comunes y estén congestionadas en el sentido de que sus funciones de coste en los arcos se incrementan con el flujo, se presenta un método computacional general para el análisis de la sensibilidad. Para otro caso donde la red de automóviles esté congestionada, pero la red de transporte público simplemente consista en líneas independientes, que conecten pares Origen-Destino y pueda haber economías de escala con el incremento de los pasajeros, se investigan condiciones bajo las cuales el análisis de la sensibilidad puede realizarse correctamente.

Yang et al. (2006) tratan sobre modelos y algoritmos para los problemas de localización de conteos de tráfico basados en líneas cordón contadoras. Se considera un problema práctico pasado por alto, pero importante, sobre la selección óptima de las líneas cordón contadoras para el estudio del censo de tráfico en redes. El problema se puede describir como: (1) Cómo seleccionar la localización óptima de un número k de estaciones de conteo de tráfico para separar todos los pares O-D. Aquí, un par O-D se dice separado si los viajes entre este par O-D están totalmente interceptados por las actuales estaciones de conteo de tráfico. Los problemas de interés son formulados como modelos de programación lineal entera. Después de explorar los problemas de programación lineal relajados y sus correspondientes problemas duales, se desarrolla un esquema de solución

que combina una ordenación de caminos, basada en un procedimiento de generación de columnas, y una técnica branch-and-bound para encontrar una solución óptima a la localización de los contadores.

1.4 Sumario de la tesis

Siguiendo las pautas habituales en la realización de este tipo de trabajo de investigación, la tesis se ha dividido en 5 capítulos.

En el *Capítulo 1* se revisa el *estado-del-arte*, procediéndose a presentar los conocimientos actuales sobre el problema de asignación de tráfico en redes congestionadas (TAP) y realizándose una revisión del problema de estimación de matrices de viaje Origen-Destino en redes de tráfico congestionadas (DAP).

En el *Capítulo 2* se desarrolla un algoritmo de generación de columnas (CGA) para el DAP. Este algoritmo itera entre un problema de asignación de tráfico en equilibrio, con el que obtiene un conjunto de caminos en equilibrio y el *DAP restringido* al conjunto de caminos previamente generados, cuya solución proporciona una nueva estimación de la matriz de viajes O-D. El DAP restringido se formula mediante optimización de un solo nivel. Se demuestra, bajo la hipótesis relativa a la continuidad de las funciones implicadas y una condición de regularidad, la convergencia del algoritmo a un óptimo local del DAP.

Para analizar la convergencia y rendimientos del algoritmo propuesto se presentan varias pruebas numéricas sobre problemas de pequeña escala.

El *Capítulo 3* está dedicado a desarrollar un marco para la elaboración de algoritmos heurísticos para el DAP. Esta clase de algoritmos surge de la aplicación de estrategias heurísticas de resolución al CGA. Los algoritmos propuestos en la literatura SAB⁷ e IEA⁸ pertenecen a esta clase. La nueva interpretación del SAB y del IEA, que se hacen aquí, presenta fundamentalmente dos ventajas: i) permite una implementación computacional más eficiente al usar el espacio de flujo en los caminos, en lugar del flujo en los arcos originalmente empleado, y, también, ii) permite el establecimiento de condiciones para la convergencia a puntos estacionarios (puntos de Karush-Khun-Tucker) para el SAB. El capítulo se completa con un estudio computacional sobre diversas implementaciones en redes de pequeña escala en las que se evalúa el rendimiento de los algoritmos propuestos y se prueban varias propuestas de modelización para el DAP, que conducen a la estimación por mínimos cuadrados en aplicaciones de minimización del

Para completar el estudio propuesto para esta Tesis, en el *Capítulo 4* se aborda el problema de calibrar (o estimar) los parámetros de las funciones de coste en los arcos (PCPCA) de una red de tráfico congestionada mediante el uso de observaciones de flujos

⁷Sensitivity analysis based

⁸Iterative estimation assignment

en los arcos y los tiempos/costes de viaje en equilibrio en un subconjunto de pares de demanda origen-destino. En este capítulo se proponen dos programas matemáticos con restricciones de equilibrio (MPEC) para modelizar este problema. Se establece la existencia de soluciones bajo condiciones de continuidad en las funciones de coste en los arcos y se realiza un análisis de la sensibilidad de las soluciones. Se ha adaptado el algoritmo CGA para resolver los modelos MPECs. Este algoritmo se basa en dos ideas: i) convertir los MPECs en un programa matemático equivalente de un sólo nivel en el espacio de flujo en los caminos mediante la formulación del problema de asignación de tráfico como un problema de complementariedad no lineal (NCP) y ii) aplicar sobre el modelo equivalente un esquema modificado de descomposición simplicial desagregada (DSD) en el que la fase de generación de columnas es obtenida mediante la resolución de un problema no lineal de asignación de tráfico. Se ha demostrado, haciendo uso del análisis de sensibilidad, que el algoritmo propuesto converge a óptimos locales del problema. Se completa el capítulo con un estudio numérico sobre dos redes de pequeño tamaño, finalizando con una aplicación a la red urbana de Ciudad Real.

Finalmente, en el *Capítulo 5* se agrupan las conclusiones y se sugieren posibles líneas para futuras investigaciones.

Capítulo 2

Algoritmo de generación de columnas para la estimación de matrices de viajes Origen-Destino en redes de tráfico congestionadas

Resumen

En este capítulo se desarrolla un algoritmo de generación de columnas (caminos) para el problema de estimación de matrices de viaje Origen-Destino en redes de tráfico congestionadas (DAP¹)². Este algoritmo itera entre un problema de asignación de tráfico en equilibrio, con el que obtiene un conjunto de caminos en equilibrio y el *DAP restringido* al conjunto de caminos previamente generados, cuya solución proporciona una nueva estimación de la matriz de viajes O-D. El DAP restringido se formula mediante optimización de un solo nivel. Se demuestra, bajo una sola hipótesis relativa a la continuidad de las funciones implicadas, la convergencia del algoritmo a un óptimo local del DAP.

Para analizar la convergencia y rendimientos del algoritmo propuesto se han realizado varias pruebas numéricas sobre problemas de pequeña escala.

Palabras clave: estimación de matrices origen-destino, programación matemática con restricciones de equilibrio, métodos de generación de columnas, optimización de gran escala, optimización no diferenciable.

¹Demand adjustment problem

²Investigación parcialmente financiada por fondos FEDER y el proyecto CICYT BFM2002-04548-C03-03

2.1 Introducción

El problema de ajustar una matriz de viajes Origen-Destino (O-D), a partir de mediciones de aforos en una red de tráfico y de una matriz O-D desactualizada, ha suscitado un gran interés entre los investigadores en los últimos años. Esto es debido fundamentalmente a que este tipo de información se puede obtener más rápida y económicamente que la obtenida tradicionalmente mediante la realización de encuestas.

El marco para la formulación de este problema en redes de tráfico congestionadas, que abreviadamente denominaremos DAP, es la programación matemática con restricciones de equilibrio (MPEC ⁸). Un MPEC consta de dos problemas relacionados: el *problema interior* o también llamado *problema del nivel inferior* y el *problema exterior* o *problema del nivel superior*. En el DAP el problema interior está definido por un problema de asignación de tráfico en equilibrio parametrizado por la matriz de viajes O-D (variables del nivel superior) y el problema exterior está definido por un problema de estimación de la matriz de viajes O-D en el que interviene los flujos observados en la red.

Se puede realizar una taxonomía de los modelos DAP empleados en la literatura en función de tres características fundamentales:

- i) El modelo de asignación de tráfico en equilibrio empleado en el nivel inferior. Los ejemplos más notables son la asignación *todo-o-nada* para redes poco congestionadas, equilibrio determinista bajo el punto de vista del usuario (DUE), asignación estocástica (SUE), etc.
- ii) La información disponible para plantear el problema en el nivel superior. Esta información suele consistir en un subconjunto de mediciones de aforo en la red y, quizás, una matriz de viaje O-D desactualizada u obtenida por otro procedimiento.
- iii) El método estadístico empleado. Los procedimientos más habituales son el de máxima verosimilitud, mínimos cuadrados generalizados, derivaciones del principio de máxima entropía, etc.

Las elecciones ii) y iii) conducen a la especificación de la función objetivo del DAP, mientras que i) define el problema interior. Una revisión bibliográfica de estos modelos -- El DAP presenta tres dificultades esenciales. La primera deriva del hecho de poseer malas propiedades matemáticas, como la no-convexidad y la no diferenciable; la segunda dificultad es las grandes dimensiones de los problemas reales y la tercera es que la aplicación multievaluada que define los flujos en equilibrio no se conoce explícitamente, estando únicamente definida implícitamente por el problema interior.

⁸Mathematical programming with equilibrium constraints

Estas dificultades ha conducido al desarrollo de algoritmos de una marcada naturaleza heurística.

A *grosso modo* dos son las estrategias que se han desarrollado para la resolución del DAP. La primera se basa en aplicar técnicas de optimización comúnmente usadas en optimización de un solo nivel, como son los algoritmos de direcciones de descenso. Los métodos descritos en Spiess (1990b), Drissi-Kaitouni and Lundgren (1992), Florian and Chen (1993), Codina and Barceló (2000), Codina et al. (2001) y Codina and Montero (2003) constituyen ejemplos de esta línea de investigación en la que se obtienen diferentes aproximaciones al gradiente de la función objetivo

La segunda estrategia se basa en aproximar la aplicación multievaluada, que define el problema interior, por una aproximación lineal mediante los llamados *factores de influencia*. Los ejemplos más notables son el algoritmo IEA⁴ propuesto en Yang et al. (1992), que está basado en el uso de proporciones para obtener los factores de influencia, y el SAB⁵ dado en Yang (1995), en el que se calculan los factores de influencia mediante análisis de sensibilidad. En Codina et al. (2001) se plantea el cálculo de las aproximaciones lineales de una forma diferente. En el citado trabajo se propone obtener estas aproximaciones como la aplicación multievaluada resultante de aproximar la función objetivo del problema interior original. El método propuesto, denominado EAM, es el resultado de considerar una aproximación de primer orden de la función objetivo del nivel inferior.

Recientemente, Cascetta and Postorino (2001) formulan el DAP como la localización de un punto fijo, en una función definida implícitamente, mediante la solución de un problema de programación matemática. Esta formulación permite el desarrollo de algoritmos de tipo punto-fijo.

Los algoritmos de generación de columnas son una poderosa herramienta para resolver problemas de programación matemática de gran escala. Estos métodos iteran entre la resolución del llamado *problema maestro restringido*, que consiste en el problema original restringido a una aproximación interior de la región factible original, y el llamado *problema de generación de columnas*, que mediante una relajación del problema original genera un nuevo vector (columna) para mejorar la anterior aproximación interior.

En este trabajo formulamos el DAP en el espacio de flujo en los caminos y proponemos un algoritmo de generación de columnas para evitar la enumeración explícita de todos los caminos de la red. Este algoritmo resuelve iterativamente un *problema de asignación* para una matriz O-D dada, cuya solución permite obtener nuevos caminos, junto de rutas generadas. En el DAP restringido se reemplaza el problema interior restringido por sus condiciones de optimalidad de primer orden, las llamadas condiciones de Karush-Kuhn-Tucker en el caso de modelos de optimización. Esta operación conduce a un problema no lineal y no convexo de un solo nivel. Este algoritmo lo

⁴Iterative estimation assignment

⁵Sensitivity analysis based

hemos denominado CGA⁶.

El CGA ha sido desarrollado para DAPs en los que se emplea, como método de asignación de tráfico, un modelo determinista de equilibrio desde el punto de vista del usuario. No obstante, este algoritmo se puede adaptar a DAPs que empleen modelos estocásticos de asignación.

Esta idea, la de reemplazar el problema interior por las condiciones de equilibrio que representa, ha sido empleada en los problemas de gestión de tráfico y de diseño de redes en Marcotte (1983), Meng et al. (2001). Un algoritmo similar al aquí desarrollado fue propuesto por Lim (2002) para el problema de diseño de redes (NDP⁷). Este autor demuestra la convergencia del mismo a un óptimo local bajo la hipótesis de complementariedad estricta de la sucesión de iteraciones generadas y bajo la hipótesis de continuidad en la función de coste en los caminos. En este trabajo, empleamos la selección de todos los caminos de mínimo coste en lugar de los caminos con flujo en la situación de equilibrio, lo que nos permite demostrar la convergencia a un óptimo local del DAP sin necesidad de que esta sucesión de patrones de equilibrio cumplan la condición de holgura complementaria estricta.

El capítulo se completa con una experiencia numérica sobre tres ejemplos de pequeñas dimensiones, con el fin de estudiar la convergencia del CGA y testar la viabilidad de ciertas estrategias heurísticas desarrolladas para su implementación. Los resultados obtenidos son prometedores, mostrando la convergencia a un mínimo local en muy pocas iteraciones.

2.2 Revisión del DAP

En esta sección revisaremos brevemente el problema DAP, propondremos un algoritmo basado en la generación de caminos y posteriormente estudiaremos su convergencia.

La notación empleada a lo largo de este trabajo, por conveniencia, se resume en la Tabla 3.1.

El problema de asignación de tráfico en redes congestionadas (TAP⁸), en su forma más general, admite una formulación mediante el siguiente problema de desigualdades variacionales (Smith (1979), Dafermos (1980)):

Encontrar un $v^* \in \Omega'$ tal que

$$[\text{VIP}(c, \Omega')]$$

donde el espacio factible de flujos en los arcos viene definido por $\Omega' = \{v \mid v = \Delta h, \Lambda h = g, h \geq 0\}$. Una formulación equivalente en el espacio de flujo en los caminos, es:

⁶Column generation algorithm

⁷Network design problem

⁸Traffic assignment problem

Encontrar un $h^* \in \Omega$ tal que

$$C(h^*)^T(h - h^*) \geq 0, \quad \forall h \in \Omega \quad [\text{VIP}(C, \Omega)]$$

donde el espacio factible de flujos en los caminos, para una matriz de viajes O-D, g , está definido por $\Omega(g) = \{h \mid Ah = g, h \geq 0\}$.

En este capítulo asumiremos que el problema de asignación de tráfico $\text{VIP}(C, \Omega)$ corresponde a un modelo determinista desde el punto de vista del usuario (Primer principio de Wardrop).

El problema de estimación o ajuste de una matriz O-D en una red de tráfico congestionada (DAP) se formula mediante la programación matemática con restricciones de equilibrio (MPEC). Matemáticamente, el DAP puede ser formulado en el espacio de flujo en los caminos por

$$\begin{aligned} \text{Minimizar} \quad & F(g, h) = \eta_1 F_1(g) + \eta_2 F_2(v) \\ \text{sujeto a:} \quad & C(h)^T(h' - h) \geq 0, \quad \forall h' \in \Omega(g) \\ & h \in \Omega(g) \\ & v = \Delta h \\ & 0 \leq g \leq \bar{g} \end{aligned} \quad [\text{DAP}]$$

donde las funciones F_1 y F_2 son dos métricas que miden respectivamente las discrepancias entre las observaciones disponibles de la matriz de demanda O-D y de los aforos en la red con sus respectivas predicciones. Los coeficientes $\eta_1, \eta_2 \geq 0$ reflejan la confianza que tenemos en la calidad de nuestras observaciones. Bajo las hipótesis de que las funciones F_1 y F_2 y $c(v)$ sean continuas, Chen and Florian (1996) demostraron que el DAP admite una solución.

Obsérvese que si la función objetivo del nivel superior se multiplica por el número positivo $\frac{1}{\eta_1 + \eta_2}$ la solución del problema no varía. Esto pone de manifiesto que el DAP es una ponderación de dos objetivos, por un lado intentar reproducir los flujos observados y por otro intentar mantener la misma estructura que la matriz O-D desactualizada.

NOTA 2.2.1 : Función objetivo F .

Las especificaciones de las funciones F_1 y F_2 permiten recoger las diferentes clases de estimadores usados en la literatura. La estimación por mínimos cuadrados generalizados es una de las más utilizadas. La solución del DAP mediante GLS es robusta con respecto a las distribuciones adoptadas [Cascetta and Nguyen (1988)].

En este trabajo no se asume ningún criterio específico, únicamente que las funciones F_i sean continuas en la región factible, con el objeto de poder asegurar la convergencia

³Generalized least squares

del algoritmo propuesto. No obstante, se emplean dos métricas que son poco habituales en los métodos de estimación clásicos: El valor absoluto mínimo

$$F^1(x, \hat{x}) = \sum_i \omega_i |x_i - \hat{x}_i|$$

y el *MínMax*

$$F^\infty(x, \hat{x}) = \text{máximo}_i \{\omega_i |x_i - \hat{x}_i|\}$$

para medir la discrepancia entre un vector estimado x y el vector observado \hat{x} . Los parámetros ω_i toman un conjunto de valores que se han determinado previamente.

Las métricas F^1 and F^∞ transforman la función objetivo del DAP en una función objetivo lineal mediante la introducción de nuevas variables. En el caso del método del valor absoluto mínimo, la transformación puede ser expresada como:

$$\begin{aligned} \text{minimizar } Z &= \sum_i \omega_i \varepsilon_i, \\ \text{sujeto a:} \\ x_i - \hat{x}_i &\leq \varepsilon_i, \quad i = 1, \dots, n \\ \hat{x}_i - x_i &\leq \varepsilon_i, \quad i = 1, \dots, n \\ \varepsilon_i &\geq 0, \quad i = 1, \dots, n \end{aligned}$$

En el caso de la estimación *MínMax*, la transformación será:

$$\begin{aligned} \text{minimizar } Z &= \varepsilon, \\ \text{sujeto a:} \\ \omega_i(x_i - \hat{x}_i) &\leq \varepsilon, \quad i = 1, \dots, n \\ \omega_i(\hat{x}_i - x_i) &\leq \varepsilon, \quad i = 1, \dots, n \\ \varepsilon &\geq 0 \end{aligned}$$

2.3 Algoritmo de generación de columnas (CGA)

Tobin (1986) mostró que el problema de desigualdades variacionales que define el problema de optimización es el siguiente

$$C(h) = \Lambda^T U + \pi \tag{2.1}$$

$$\Lambda h = g \tag{2.2}$$

$$\pi^T h = 0 \tag{2.3}$$

$$\pi, h \geq 0 \tag{2.4}$$

donde U es el vector de costes en equilibrio, esto es (\dots, U_i, \dots) para todo $i \in I$ y π es el vector de multiplicadores asociados a las restricciones de no negatividad en el flujo de los caminos. Estas condiciones reciben el nombre, para el caso de costes separables, de condiciones de optimalidad de Karush-Khun-Tucker.

Esta representación del problema de equilibrio permite transformar el DAP en un problema de optimización de un sólo nivel reemplazando el $VIP(C, \Omega)$ mediante las condiciones (2.1) a (2.4).

$$\begin{aligned}
 \text{Minimizar } & F = \tau_1 F_1(g) + \tau_2 F_2(v) \\
 \text{sujeto a: } & C(h) = \Lambda^T U + \pi \\
 & \Lambda h = g \\
 & v = \Delta h \\
 & \pi^T h = 0 \\
 & \pi, h \geq 0 \\
 & 0 \leq g \leq b
 \end{aligned}
 \tag{DAP}$$

Este problema no se puede abordar mediante algoritmos que trabajen directamente sobre la formulación explícita, ya que el número de variables (debido fundamentalmente al número de caminos) es prohibitivo para problemas reales, incluso de mediano tamaño.

El algoritmo que presentamos en esta sección es un método iterativo que construye una sucesión de aproximaciones del DAP. Cada uno de estos problemas aproximados, denominados *DAP restringidos*, está definido por la restricción del problema DAP sobre un subconjunto de caminos $\hat{K} \subset K$ y proporcionan una estimación de la matriz de viajes O-D. Posteriormente se resuelve el problema de asignación en equilibrio para la matriz O-D estimada, lo que genera nuevos caminos, ampliando el anterior subconjunto \hat{K} . El algoritmo termina cuando ningún nuevo camino es añadido a este subconjunto. Como el algoritmo propuesto para el resolver el DAP va generando nuevas columnas (caminos) en cada iteración según van siendo necesitadas, este es el origen del nombre *algoritmo de generación de columnas* (CGA).

Vamos a introducir la siguiente notación que será empleada en la descripción del algoritmo y en la demostración de su convergencia. Se define la aplicación multievaluada *dada*, por

$$\Omega(\hat{K}, g) := \left\{ h \mid \sum_{k \in \hat{K}_i} h_k = g_i \quad \forall i \in I, h_k \geq 0, \forall k \in \hat{K} \text{ y } h_k = 0, \forall k \in K - \hat{K} \right\}$$

y la función multievaluada de *soluciones del TAP restringido al conjunto de caminos*

\hat{K} y para una matriz O-D g dada, por

$$\Omega^*(\hat{K}, g) := \left\{ h \in \Omega(\hat{K}, g) \mid C_k(h, g) = U_i + \pi_k, \forall k \in \hat{K}_i, \forall i \in I; \sum_{k \in \hat{K}} \pi_k h_k = 0; \pi_k \geq 0, \forall k \in \hat{K} \right\}.$$

Cuando el conjunto \hat{K} coincide con el conjunto original K , denota en $\Omega^*(g) = \Omega^*(K, g)$ y $\Omega(g) = \Omega(K, g)$

Notar que si el conjunto de caminos \hat{K}_i es no vacío para todo $i \in I$ entonces el conjunto de soluciones del TAP restringido es no vacío, debido a la continuidad de la aplicación $C(h)$ sobre el conjunto convexo $\Omega(\hat{K}, g)$.

El problema DAP restringido al conjunto de caminos \hat{K} se formula:

$$\begin{aligned} &\text{Minimizar } F = \eta_1 F_1(g) + \eta_2 F_2(\Delta h) \\ &\text{sujeto a: } g \in [0, b] \\ &\quad h \in \Omega^*(\hat{K}, g) \end{aligned} \quad [\text{RDAP}(\hat{K})]$$

La ventaja del subproblema RDAP(\hat{K}) respecto al problema DAP es que el número de variables es significativamente menor y esto permite resolver (aproximadamente) directamente la formulación explícita.

La tabla 2.2 recoge el algoritmo CGA.

NOTA 2.3.1 : Extensión del CGA a DAP con un modelo de asignación estocástico tipo logit

El CGA ha sido desarrollado para DAPs que emplean el equilibrio determinista, desde el punto de vista del usuario, como principio para efectuar la asignación de tráfico. El CGA se puede extender a DAPs con modelos de asignación estocásticos tales como el modelo estocástico basado en distribuciones tipo logit. La dificultad esencial viene generada por el hecho de que, en este contexto, el conjunto $K(g^e)$ coincide con el espacio completo de caminos, haciendo coincidir el DAP-restringido con el DAP original. Este problema se puede paliar usando algoritmos de resolución del modelo estocástico basado en distribuciones tipo logit (Paso 1) que operasen en el espacio de los flujos en los caminos, como el dado en Damberg et al. (1996), en los que se obtienen una aproximación al conjunto $K(g^e)$. La segunda modificación que se debe introducir al el equilibrio estocástico, basado en distribuciones tipo logit, hay que reemplazar los costes de los caminos $C_p(h)$ por $C_p(h) + \frac{1}{\theta} \log(h_p)$, donde θ es un parámetro positivo asociado con el componente aleatorio del coste.

Estas dos modificaciones conducirían a un esquema operativo para este tipo de DAPs pero en el que no se podría aplicar los resultados teóricos desarrollados en la sección 2.4 para establecer la convergencia del algoritmo.

2.4 Análisis de la convergencia del CGA

En esta sección probamos que el CGA converge en un número finito de iteraciones a un mínimo local del DAP. La demostración de este resultado no requiere de las hipótesis que habitualmente se emplean para el análisis de la sensibilidad del TAP, como la condición de holgura complementaria estricta (esto es, si $h_k = 0$ entonces $\pi_k > 0$), ni hipótesis sobre diferenciabilidad (ver Tobin and Friesz (1988)). Estas suposiciones simplificarían la demostración de la convergencia del CGA, pero son demasiado restrictivas para ser satisfechas en redes reales. Los argumentos empleados se basan exclusivamente en hipótesis de continuidad.

Obsérvese que, como no se asume la hipótesis de complementariedad estricta, puede que el algoritmo CGA, en alguna iteración y para algún par, almacene caminos de mínimo coste que no contengan flujo.

La idea fundamental sobre la que se basa la convergencia del algoritmo CGA es que si el problema TAP es perturbado respecto a la matriz g' , esto es, si reemplazamos la matriz de demanda O-D g' por $g' + \epsilon$, donde ϵ es una pequeña perturbación de la demanda, entonces la nueva situación de equilibrio se puede describir empleando únicamente los caminos contenidos en K' . Esto demuestra que existe un entorno de la demanda actual en los que el DAP y el DAP restringido son equivalentes. Por tanto, ambos problemas tienen el mismo conjunto de mínimos locales en dicho entorno. El teorema de convergencia demuestra que se alcanza uno de estos mínimos locales.

Empezaremos demostrando que para cualquier matriz O-D g' existe un entorno abierto en el que el DAP y el DAP restringido son equivalentes.

TEOREMA 2.4.1 *Dado $g' \in [0, b]$. Supóngase que $C(h)$ es continua en $\Omega = \bigcup_g \Omega(g)$. Supóngase que existe un entorno de g' , denotado por \tilde{B} , tal que para todo $g \in \tilde{B}$ se cumple las siguientes dos propiedades:*

- i) la aplicación $C(h)$ es constante en $\Omega^*(g)$,*
- ii) la aplicación $C(h)$ es constante en $\Omega^*(K(g'), g)$.*

Entonces existe un entorno abierto \tilde{B} cumpliendo que

$$\Omega^*(g) = \Omega^*(K(g'), g), \quad \forall g \in \tilde{B}$$

donde $K(g')$ es el conjunto de caminos de mínimo coste en el equilibrio para la matriz

DEMOSTRACION. Primeramente probaremos la inclusión

$$\Omega^*(g) \subset \Omega^*(K(g'), g)$$

Dado un $h_{g'}^* \in \Omega^*(g')$. Por la suposición i), existe únicamente un vector de costes en equilibrio, que denotamos por $U(g')$.

El conjunto $K(g')$ consta de todos los caminos de mínimo coste para todos los pares $i \in I$, pudiendo incluso contener caminos sin flujo en la situación de equilibrio. Entonces, existe un número positivo ε cumpliendo

$$\min \{C_k(h^*_{g'}) - U_i(g') / p \in K_i - K_i(g'), i \in I\} > \varepsilon > 0,$$

donde $U(g')$ es el vector de costes en equilibrio para la matriz O-D g' . Por la suposición ii), $C_k(h^*_{g'})$ es independiente de la elección del vector de flujo en equilibrio $h^*_{g'}$, y por tanto, la anterior relación se cumple para todo $h^*_{g'} \in \Omega^*(g')$.

Esta relación conduce a

$$U_i(g') + \frac{\varepsilon}{2} < C_k(h^*_{g'}) - \frac{\varepsilon}{2}; \quad \forall k \in K_i - K_i(g'), \forall i \in I, \forall h^*_{g'} \in \Omega^*(g'). \quad (2.5)$$

Dado $\frac{\varepsilon}{2} > 0$, por la continuidad de C en el punto $h^*_{g'}$, existe un entorno abierto $V(h^*_{g'})$ de $h^*_{g'}$ y otro entorno abierto $B(h^*_{g'})$ de g' cumpliendo:

$$\|C(h) - C(h^*_{g'})\|_{\infty} < \frac{\varepsilon}{2}; \quad \forall (h, g) \in V(h^*_{g'}) \times B(h^*_{g'}),$$

donde $\|\cdot\|_{\infty}$ es la norma supremo.

La aplicación punto-conjunto $\Omega^*(g)$ es semi-continua superiormente en g . Chen and Florian (1996) demuestran que la aplicación flujos en equilibrio en los arcos es semi-continua superiormente bajo la hipótesis de continuidad. Empleando un argumento similar se puede probar que $\Omega^*(g')$ es un conjunto cerrado contenido en el conjunto acotado Ω . Esto implica que el conjunto $\Omega^*(g')$ es un conjunto compacto. Por otro lado, $\{V(h^*_{g'})\}_{h^*_{g'} \in \Omega^*(K(g'), g')}$ es un cubrimiento por abiertos de este conjunto compacto, entonces existe un subcubrimiento finito. Denotamos por $\{h^*_1, \dots, h^*_s\}$ el conjunto de puntos asociados al subcubrimiento finito. Denotamos por $V = \cup_i V(h^*_i)$ y por $B = (\cap_i B(h_i)) \cap \bar{B}$.

Esta relación implica que todas las funciones coordenadas de C , satisfacen

$$\|C_k(h) - C_k(h^*_{g'})\|_{\infty} < \frac{\varepsilon}{2}; \quad \forall (h, g) \in V \times B, \forall k \in K$$

y ésta puede ser rescrita por

$$C_k(h^*_{g'}) - \varepsilon < C_k(h) < C_k(h^*_{g'}) + \varepsilon \quad \forall (h, g) \in V \times B \quad \forall k \in K \quad (2.6)$$

Sea $k \in K_i(g')$ y $k' \in K_i - K_i(g')$, y empleando las ecuaciones (4.10) y (4.3), se obtiene que

$$C_k(h) < C_k(h^*_{g'}) + \frac{\varepsilon}{2} = U_i(g') + \frac{\varepsilon}{2} < C_{k'}(h^*_{g'}) - \frac{\varepsilon}{2} < C_{k'}(h), \quad (2.7) \\ \forall (h, g) \in V \times B.$$

Ahora probaremos por reducción al absurdo que existe un entorno $\bar{B} \subset B$ de g' cumpliendo

$$\Omega^*(g) \subset \Omega^*(K(g'), g), \quad \forall g \in \bar{B}$$

Supongamos que todo entorno de g' contenido en B , posee un elemento g tal que $h^*_g \notin \Omega^*(K(g'), g)$ para algún $h^*_g \in \Omega^*(g)$. Entonces, podemos construir dos sucesiones $\{g^j\}$ y $\{h^j\}$ cumpliendo: i) $\{g^j\} \rightarrow g'$ y ii) $h^j \in \Omega^*(g^j) - \Omega^*(K(g'), g^j)$.

Del hecho que $\{h^j\} \in \Omega$ y que el conjunto de flujos factibles es compacto, existe una subsucesión convergente $\{h^j\}_{j \in J}$. Denotamos su límite por \bar{h} .

La aplicación multivaluada $\Omega^*(g)$ es semi-continua superiormente en g' por el Teorema 1 de Chen and Florian (1996), y se cumple por tanto

$$\text{Si } \begin{array}{l} h^j \in \Omega^*(g^j), \quad \forall j \in J \\ \{g^j\}_{j \in J} \rightarrow g' \\ \{h^j\}_{j \in J} \rightarrow \bar{h} \end{array} \quad \Rightarrow \quad \bar{h} \in \Omega^*(g')$$

Debido a que $(h^j, g^j)_{j \in J} \rightarrow (\bar{h}, g') \in V \times B$, donde $V \times B$ es un conjunto abierto, existe un número entero j_0 tal que $(h^j, g^j) \in V \times B$ para todo $j > j_0$.

Empleando la relación (4.11) para (h^j, g^j) con $j > j_0$, obtenemos que el conjunto de caminos mínimos es un subconjunto de $K(g')$. Por otro lado, $h^j \in \Omega^*(g^j)$ (es un vector de flujo de caminos en equilibrio) y eso implica que $h^j_{k'} = 0$ para todo $k' \in K_i - K_i(g')$ y para todo $i \in I$ y se tiene que $h^j \in \Omega^*(K(g'), g^j)$ para todo $j > j_0$, lo que es una contradicción.

Ahora probaremos la inclusión contraria. Sea $\bar{h} \in \Omega^*(K(g'), g)$ donde $g \in \bar{B} \cap \bar{B}$. Por la suposición ii) $C(h)$ es constante en $\Omega^*(K(g'), g)$ y $\Omega^*(g) \subset \Omega^*(K(g'), g)$, entonces existe un vector flujo en los caminos en equilibrio $h^* \in \Omega^*(g)$, cumpliendo

$$C(\bar{h}) = C(h^*). \quad (2.8)$$

Sea $h \in \Omega$, probaremos que $C(\bar{h}, g)^T(h - \bar{h}) \geq 0$. De la optimalidad de h^* , se tiene

$$C(h^*)^T(h - h^*) \geq 0. \quad (2.9)$$

De la optimalidad de \bar{h} en $\Omega(K(g'))$ y $h^* \in \Omega(K(g'))$, se tiene

Empleando (4.12), (4.13) y (4.14), se obtiene

$$\begin{aligned} C(\bar{h})^T(h - \bar{h}) &= C(\bar{h})^T[(h - h^*) + (h^* - \bar{h})] \\ &= C(h^*)^T(h - h^*) + C(\bar{h})^T(h^* - \bar{h}) \geq 0. \end{aligned}$$

□

COROLARIO 2.4.2 Condición suficiente para la existencia de un mínimo local para el DAP

Supongamos que se cumplen las hipótesis del teorema 4.3.4, y si $(h^*_{g'}, g')$ es un mínimo local para el RDAP(\hat{K}) y $K(g') \subset \hat{K}$, entonces $(h^*_{g'}, g')$ es un mínimo local para el DAP.

DEMOSTRACIÓN. De la relación $\Omega(K(g')) \subset \Omega(\hat{K}) \subset \Omega$ y del teorema 4.3.4, existe un entorno \tilde{B} de g' tal que

$$\Omega^*(K(g'), g) = \Omega^*(\hat{K}, g) = \Omega^*(g), \quad \forall g \in \tilde{B}.$$

Por hipótesis el par $(h^*_{g'}, g')$ es un mínimo local de

$$\begin{aligned} &\text{Minimizar } F(g, h) = \eta_1 F_1(g) + \eta_2 F_2(v) \\ &\text{sujeto a: } g \in [0, b] \cap \tilde{B} \\ &\quad h \in \Omega^*(\hat{K}, g). \end{aligned}$$

lo que prueba el corolario. □

El algoritmo CGA va ampliando en cada iteración el conjunto de caminos retenidos. Debido a la finitud de éstos en la red, el algoritmo debe terminar en un número finito de iteraciones, incluso cuando se toma el valor del parámetro $\varepsilon = 0$. El siguiente teorema establece una condición para garantizar que la última matriz O-D estimada es un mínimo local del DAP.

TEOREMA 2.4.3 (CONVERGENCIA DEL CGA) *Sea g' la última iteración generada por el algoritmo. Suponer que g' satisface las condiciones del teorema 4.3.4, entonces g' es un mínimo local para el DAP.*

DEMOSTRACIÓN. El algoritmo CGA genera una sucesión de conjuntos de caminos $\{K^\ell\}$ cumpliendo que $K^\ell \subset K^{\ell+1}$ (Mirar el paso 2 del algoritmo CGA). Por la finitud de caminos en la red de tráfico, el CGA debe terminar en un número finito de iteraciones. La última iteración cumple el criterio de paro.

$$K(g') \subset K^{s-1} \text{ para un cierto entero } s.$$

Además, g' es un mínimo local para el RDAP(K^{s-1}) por definición del paso 3, y, empleando el corolario 4.3.5 para $\hat{K} = K^{s-1}$, el teorema queda demostrado. □

2.5 Experiencia numérica

En esta sección se analiza numéricamente la convergencia del algoritmo CGA y de algunas estrategias heurísticas para su implementación que llamaremos métodos CGA truncados. En esencia hay tres líneas fundamentales para obtener métodos CGA truncados:

- 1 La primera simplificación se basa en suponer que todos los caminos retenidos deben de ser empleados por la nueva estimación de la matriz O-D. Esto conduce a que las restricciones $\hat{h} \geq 0$ no son activas y por tanto sus multiplicadores $\hat{\pi}$ asociados tomarán el valor de cero, lo que nos permite eliminar el conjunto de restricciones no lineales de complementariedad [4.8].
- 2 La segunda simplificación consiste en no almacenar las columnas (camino) generadas en las iteraciones anteriores. Esta estrategia reduce el tamaño de los diferentes DAP restringidos pero no se cumplen las condiciones de convergencia del CGA dadas en el teorema [4.3.6].
- 3 La tercera considera aproximaciones lineales al coste en los caminos. Esta estrategia no se aborda aquí y será considerada más adelante.

Un objetivo fundamental de esta experiencia se centra en analizar el efecto de las dos primeras suposiciones sobre la convergencia del CGA resultante. Para ello hemos considerado todas las posibilidades de los dos primeros ítem que se sintetizan en la tabla 3.2. El CGA3 ha sido contemplado con el único objeto de ser exhaustivos, aunque, a priori, podemos argumentar negativamente en contra de su efectividad ya que este algoritmo fuerza a emplear todos los caminos anteriormente generados, siendo algunos de ellos únicamente viables para situaciones de gran congestión. Esto conduce a estimaciones de matrices O-D que utilicen esos caminos con la consiguiente pérdida de convergencia.

Se han realizado tres experimentos numéricos sobre dos redes de pequeño tamaño: Red de Nguyen-Dupuis (Nguyen and Dupuis (1984), en adelante ND) y Red de Friesz-Harker (Friesz and Harker (1985), en adelante FH). Sus topologías se muestran en las figuras 3.1 y 2.2 respectivamente. La red ND tiene 13 nodos, 19 arcos y cuatro pares O-D: El par $(1 \rightarrow 2)$ con demanda g_1 , $(1 \rightarrow 3)$ con demanda g_2 , $(4 \rightarrow 2)$ con demanda g_3 y $(4 \rightarrow 3)$ con demanda g_4 . La red FH tiene 6 nodos, 16 arcos y dos pares O-D: El par $(1 \rightarrow 6)$ con demanda g_1 , $(6 \rightarrow 1)$ con demanda g_2 .

Hemos considerado cuatro modelos de DAP, caracterizados por el uso o no de una matriz de referencia (target matrix) y por el uso de aforos en todos los arcos de la red o sólo en un subconjunto de ellos. Todas las combinaciones se describen en la tabla 3.7

Los flujos observados para los experimentos 1 y 3 corresponden a la asignación de una matriz O-D (400, 800, 600, 200) desde el punto de vista del usuario (TAP-UE). En el experimento 1 se han usado los costes en los arcos descritos en Nguyen and Dupuis (1984). Los costes en los arcos del experimento 3 se han construido mediante añadido una penalización en los arcos a_{14} y a_{18} cuando sus flujos respectivos superan las 400 y 300 unidades. Estas penalizaciones son 8.914 para el arco a_{14} y 11.501 para el arco a_{18} . Estos valores se corresponden con los multiplicadores de Lagrange para las restricciones de capacidad utilizadas en el TAP. Esta modificación genera funciones de coste discontinuas. Esta elección significa que si usamos el TAP para estas funciones de coste modificadas, los resultados obtenidos coinciden con el que genera los mismos

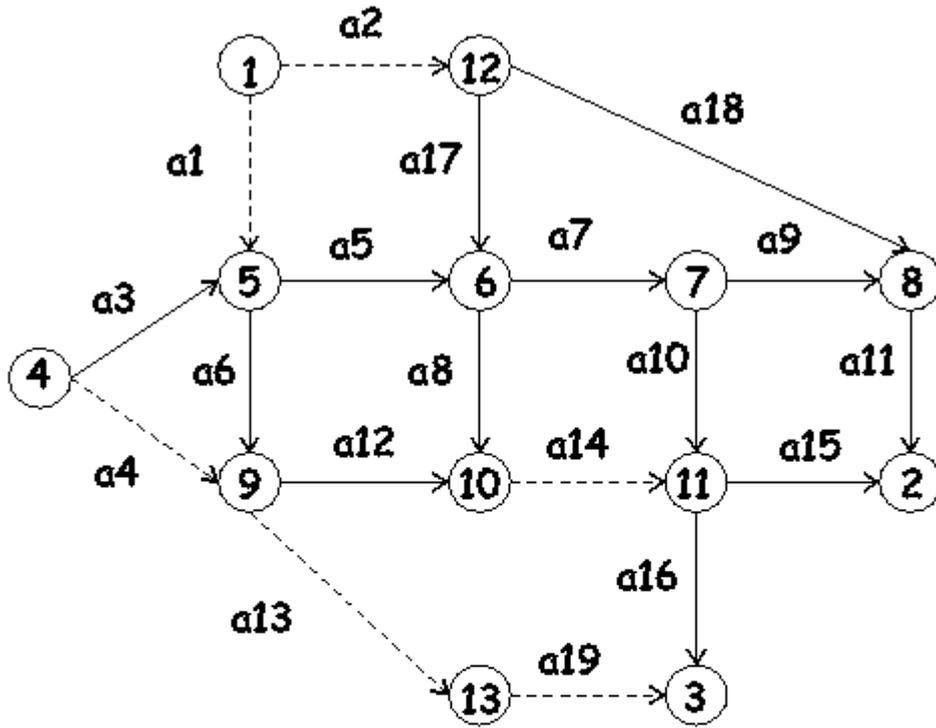


Figura 2.1: Topología de la red de Nguyen-Dupuis

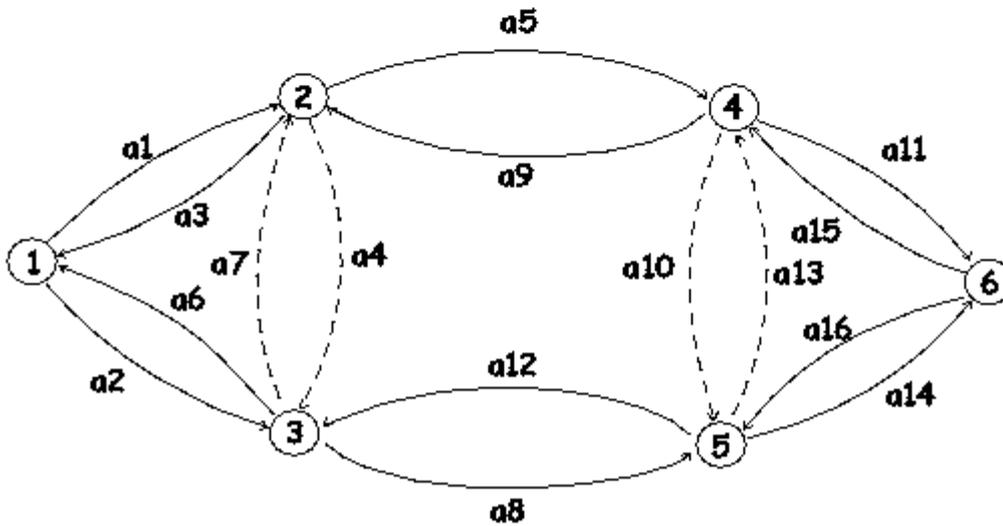


Figura 2.2: Topología de la red de Friesz and Harker

flujos en equilibrio que el TAP con restricciones de capacidad y para las funciones de coste en los arcos usadas en el ejemplo 1.

Los flujos observados para el experimento 2 corresponden a la asignación de una matriz O-D (10, 20) desde el punto de vista del usuario utilizando los costes de arco dados en Friesz and Harker (1985).

Los resultados obtenidos en los experimentos 1, 2 y 3 se muestran respectivamente en las tablas 2.5, 2.6 y 2.7. Todas estas tablas tienen la misma estructura. Cada fila muestra los resultados obtenidos para cada combinación de tipo de DAP (ver Tabla 3.7) y algoritmo usado para resolverlo (ver Tabla 3.2). La primera columna define el DAP usado y la segunda el algoritmo empleado. Se han utilizado dos puntos iniciales, uno asociado a una situación de gran congestión de las redes y otro a una de baja congestión. Para cada punto inicial se ha reportado la última O-D matriz estimada, el valor de la función objetivo y el número de iteraciones empleadas. Esta información es recogida en el resto de columnas.

Todos los algoritmos se han codificado usando el programa GAMS (General Algebraic Modeling Systems). Una rápida introducción a este programa se puede ver en el libro de Castillo et al. (2002). Se ha elegido como criterio de paro el ratio de cambio relativo entre sucesivas soluciones de la matriz O-D, g^{f+1} y g^f , y el valor del parámetro de tolerancia fue de $\varepsilon = 1.0 \times 10^{-4}$, es decir,

$$\sum_{i \in I} \frac{|g_i^{f+1} - g_i^f|}{g_i^f} \leq \varepsilon = 1.0 \times 10^{-4},$$

2.5.1 Experimento 1

En la tabla 2.5 se muestran los resultados obtenidos para la última iteración de los algoritmos, el valor de la función objetivo y el número de iteraciones empleado. La primera conclusión es que los CGA, CGA1 y CGA2 alcanzan la solución óptima en todos los casos, y además el número de iteraciones empleadas está entre 2 y 3 para el CGA y CGA1, mientras entre 2 y 5 para el CGA2. También se constata que el CGA3 converge a un punto muerto en el DAP2 para la inicialización asociada a una situación de alta congestión.

Los problemas DAP2 y DAP3 tienen un óptimo global en la matriz (400, 800, 600, 200) y el valor óptimo es 0. Esta solución es degenerada, esto es, la matriz O-D tiene caminos en equilibrio sin flujo y no satisface la condición SC. Existen evidencias numéricas de la no existencia de mínimos locales para los DAPs considerados en este experimento.

2.5.2 Experimento 2

principales de este experimento hemos dibujado las curvas de nivel de los cuatro modelos DAP que se muestran en la figura 2.3. Se observa que el DAP2 posee dos mínimos locales, de los cuales, únicamente uno es global. Este problema es no convexo y no diferenciable. El problema DAP3 posee un único mínimo global y en un entorno suyo parece que se trata de un problema convexo y diferenciable. Los problemas DAP0 y DAP1 se obtienen sumando la función $F_1(g)$ a la función objetivo de los problemas

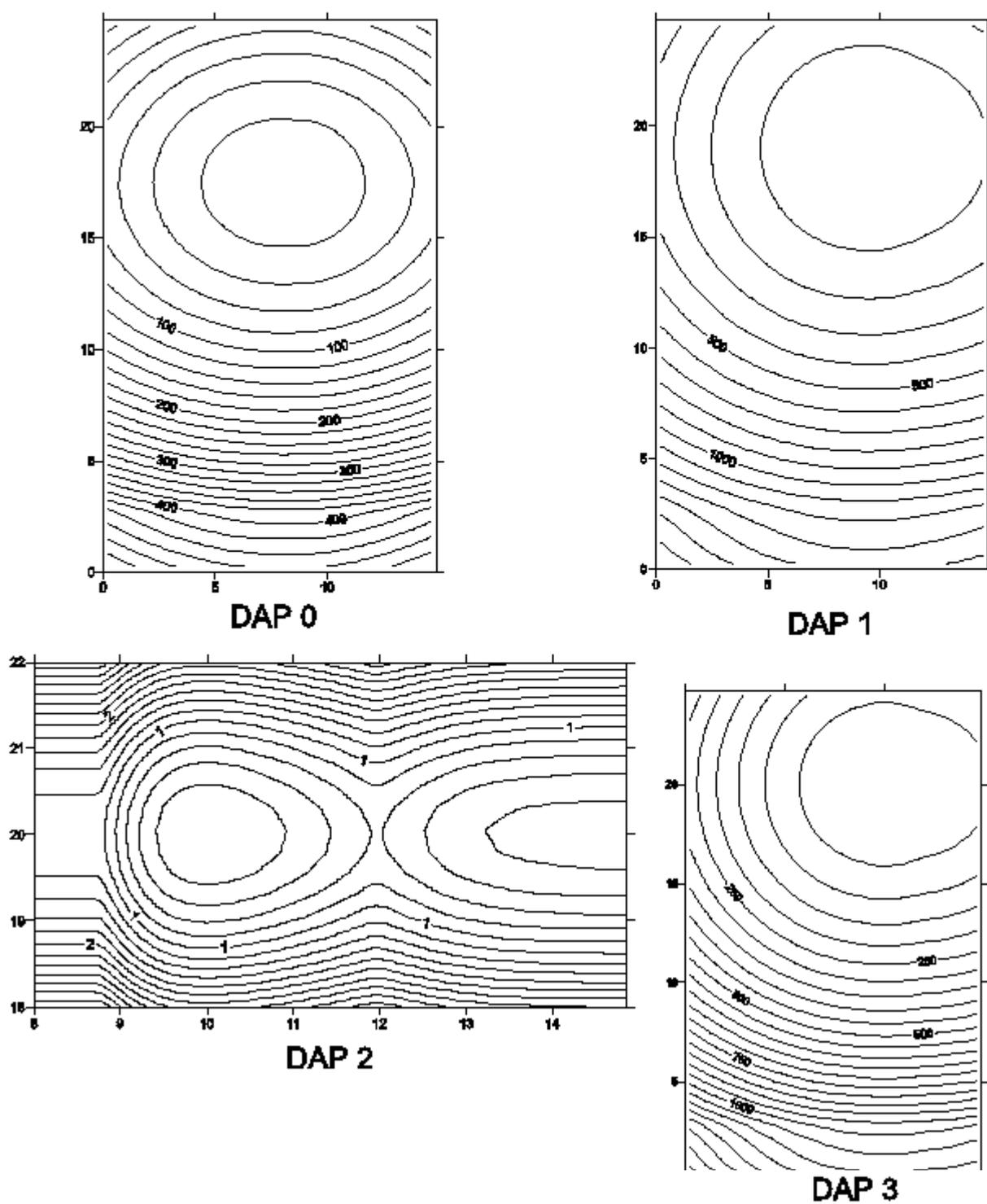


Figura 2.3: Curvas de nivel del experimento 2

DAP2 y DAP3. Este término se puede interpretar de forma parecida al término re-

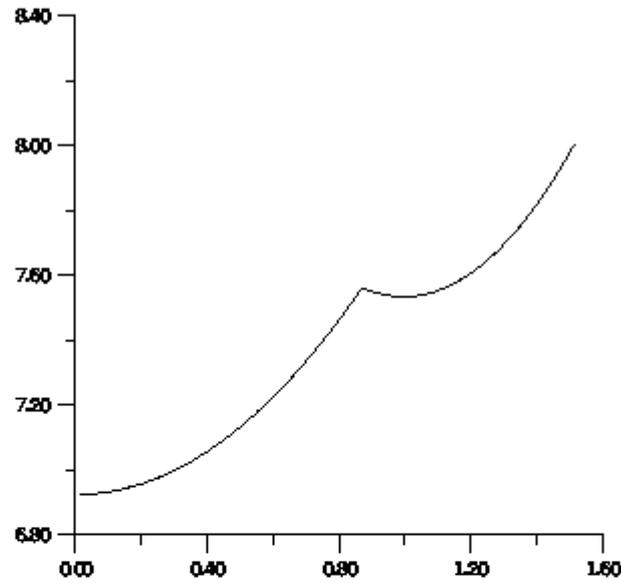


Figura 2.4: Experimento 2: Mínimo local

gularizador de los algoritmos de punto próximo. Esto haría que el problema DAP0 mejorase las propiedades de convexidad y diferenciabilidad del problema DAP2. En la figura 2.3 parece que el problema DAP0 se vuelve convexo, diferenciable y con un solo mínimo local. Pero esto no es así. Hemos representado la función objetivo del DAP0 sobre el segmento que une los puntos $(8, 17.47)$ con $(8.92, 17.47)$ y se obtiene la gráfica de la figura 2.4, la cual muestra que estamos nuevamente ante un problema no convexo, no diferenciable y que posee dos óptimos locales.

En los resultados computacionales mostrados en la tabla 2.6 se observa que los CGA, CGA1 y CGA2 vuelven a converger a óptimos locales, algunos de los cuales no son globales.

Respecto al número necesario de iteraciones para alcanzar la convergencia se observa que el CGA utiliza 2 ó 3 iteraciones. Esto es así debido a que si partimos de una matriz altamente congestionada se generan casi todos los caminos que describen la solución del DAP y por tanto el DAP restringido y el DAP casi coinciden, por tanto se resuelve en muy pocas iteraciones. Esto nos indicaría que para redes de gran tamaño, aunque el coste computacional por iteración es alto, serían necesarias pocas iteraciones.

La motivación del tercer experimento es evaluar la robustez de los algoritmos propuestos cuando la hipótesis de continuidad de la función de coste en los arcos no se cumple. Nótese que esta hipótesis es necesaria para la convergencia del CGA y también lo es para garantizar la existencia de soluciones de $VIP(C, \Omega)$ (existencia de equilibrio). El concepto de equilibrio se ha extendido a funciones discontinuas, a través del llamado

equilibrio discontinuo desde el punto de vista del usuario. Se puede garantizar la existencia de este tipo de equilibrio, por tratarse de funciones de coste positivas, acotadas, semicontinuas superiormente y regulares (ver, por ejemplo, Teorema 3.18, Patriksson (1994)) para el Ejemplo 3. Este ejemplo constituye una evidencia numérica de que, quizás, se pudiera relajar la hipótesis de continuidad de las funciones de coste en los arcos (por ejemplo, trabajando con funciones de coste positivas, acotadas, semicontinuas superiormente y regulares) sin afectar a las propiedades de convergencia del CGA.

El esquema del algoritmo del CGA dado en la Tabla 2.2 sigue siendo válido en este caso, pero aparece una dificultad en la resolución del TAP en Paso 1. Para obtener la situación de equilibrio hemos tenido que introducir restricciones adicionales y variables artificiales utilizadas en la programación matemática clásica.

Para entender los DAP empleados en este ejemplo, hemos calculado las curvas de nivel de la función objetivo para dos pares O-D fijos. Los valores fijados han sido $g_2 = 667.81$ y $g_4 = 208.3$. Las gráficas obtenidas se muestran en la figura 2.5 y siguientes. Se tratan de problemas de optimización discontinua junto a la existencia de mínimos locales.

La principal conclusión es que los algoritmos CGA, CGA1 y CGA2 siguen convergiendo a óptimos locales y el número de iteraciones sigue manteniéndose pequeño.

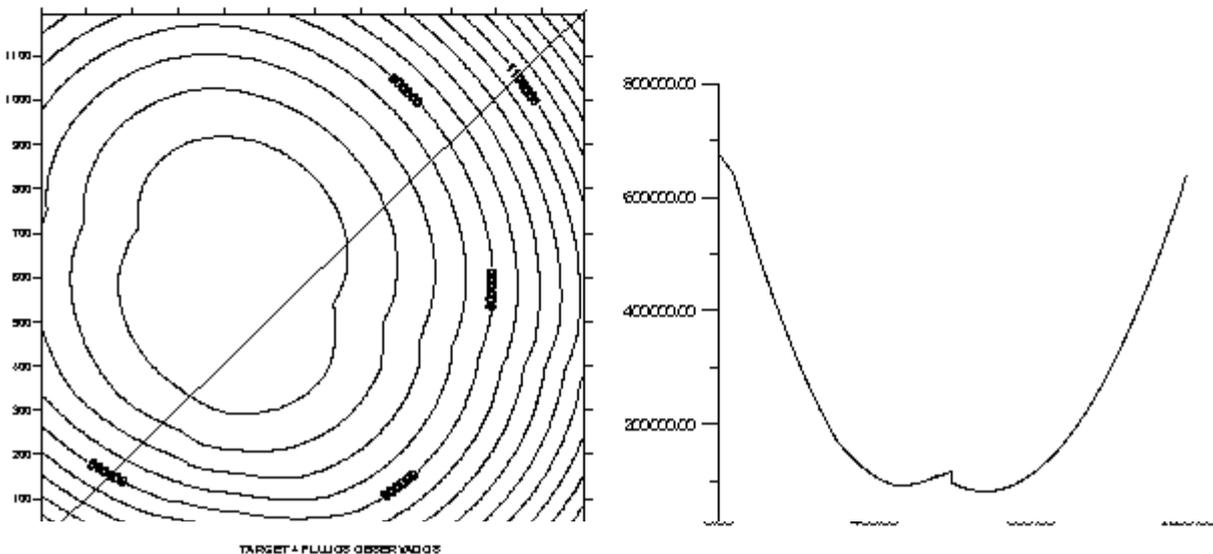


Figura 2.5: Experimento 3: Curvas de nivel

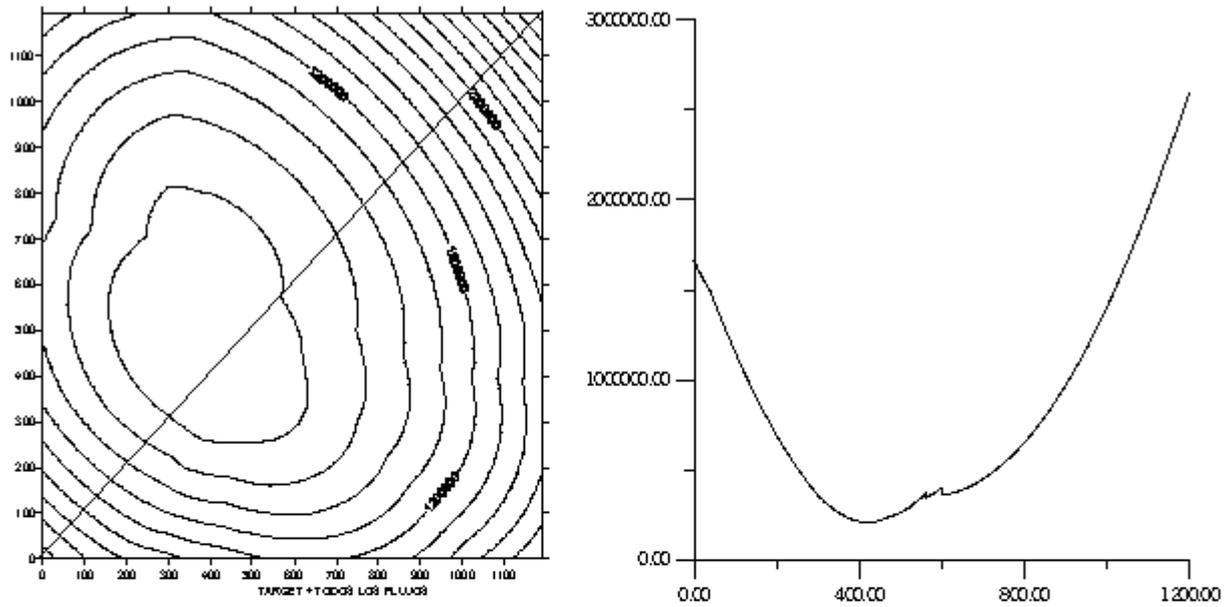


Figura 2.6: Experimento 3: Curvas de nivel

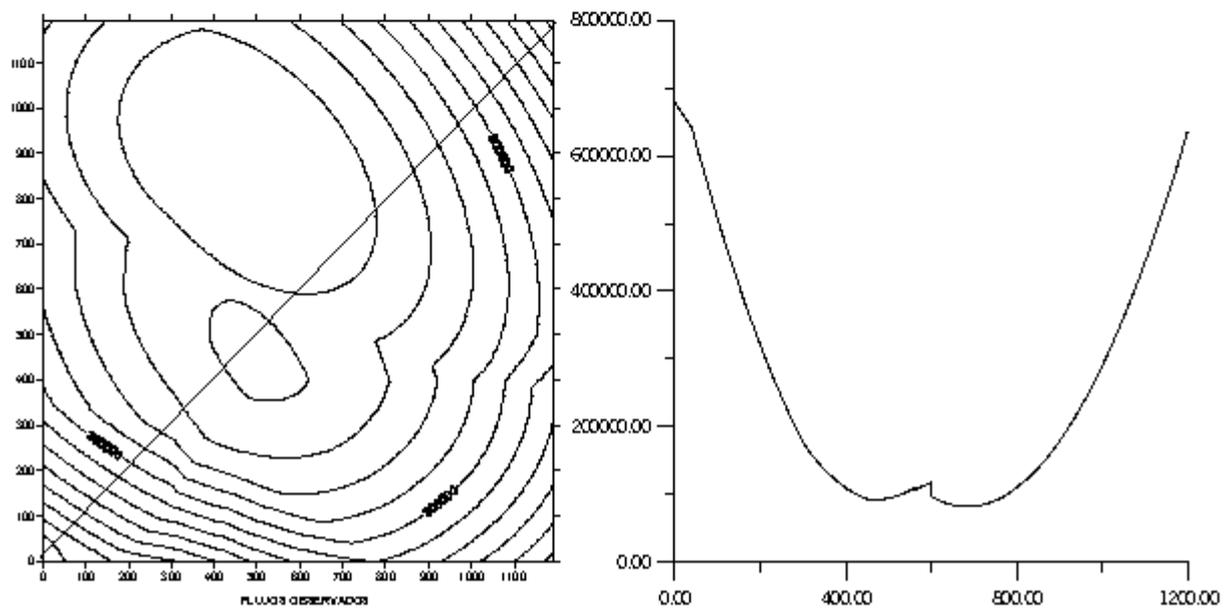


Figura 2.7: Experimento 3: Curvas de nivel

4.0 CONCLUSIONES

En este trabajo hemos propuesto un algoritmo de generación de columnas para el DAP que tiene garantizada su convergencia a óptimos locales bajo únicamente hipótesis de continuidad en las funciones de coste en los caminos y la norma usada en el problema de nivel superior. En los experimentos numéricos hemos contrastado este tipo de

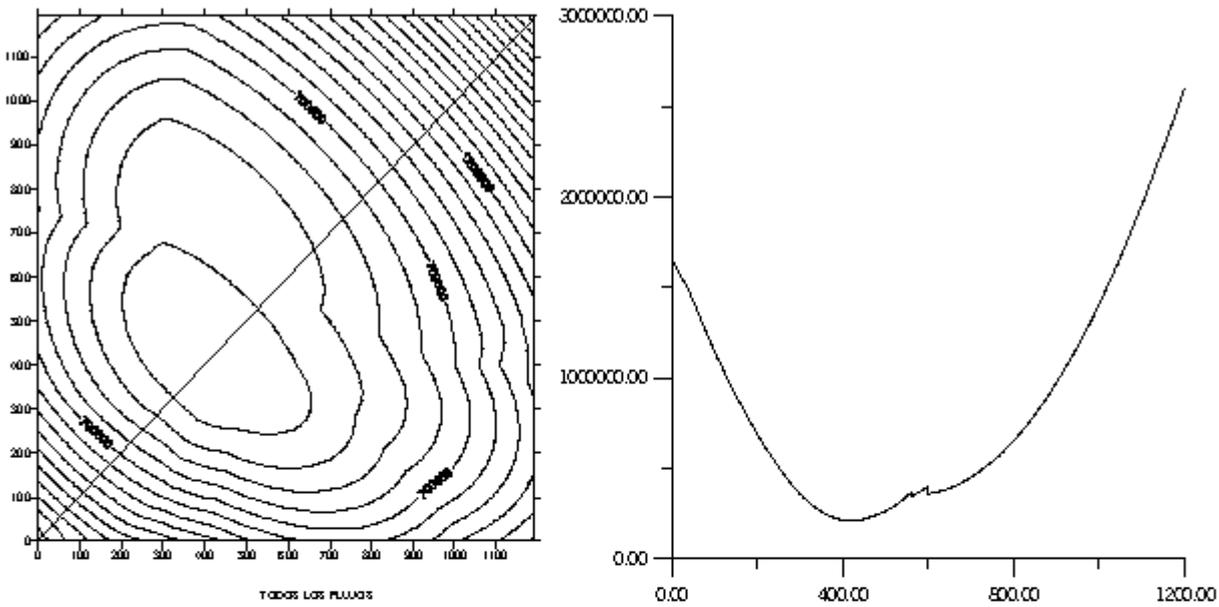


Figura 2.8: Experimento 3: Curvas de nivel

convergencia, y hemos visto que incluso puede converger sin necesidad de esta hipótesis.

En los test numéricos hemos comprobado que el número de iteraciones necesario para alcanzar esta convergencia es pequeño, y este número depende de la identificación del conjunto de caminos en equilibrio para los mínimos locales. Este hecho, unido a que las estrategias CGA1 y CGA2 funcionan bien (reducen el tamaño de los DAPs restringidos), hace que pensemos que estos métodos se puede aplicar a problemas de gran tamaño. Para ello, en el siguiente capítulo se van a desarrollar estrategias que permitan aproximar linealmente los costes en los arcos.

También hemos comprobado, en los test numéricos, que la convergencia se da solamente a óptimos locales. Esto justifica que se deba realizar un esfuerzo de investigación sobre como abordar problemas de gran escala con múltiples óptimos locales.

Tabla 2.1: Notación empleada

\mathcal{N} :	conjunto de nodos de la red de tráfico
\mathcal{A} :	Conjunto de arcos de la red
I :	Conjunto de pares O-D en la red
K :	Conjunto de caminos en la red
K_i :	Conjunto de caminos que unen en el par O-D i
$i \in I$:	Un par O-D de la red
$k \in K$:	Un camino de la red
$a \in \mathcal{A}$:	Un arco de la red
g_i :	Demanda de viajes para el par O-D i
g :	Matriz de viajes O-D, esto es, el vector con todas las demandas de viajes (\dots, g_i, \dots).
b :	Un límite superior de g .
\hat{g} :	Matriz O-D de referencia (target O-D matrix), por ejemplo (\dots, \hat{g}_i, \dots).
h_k :	Flujo en el camino k
h :	Vector de flujo en los caminos (\dots, h_k, \dots).
v_a :	Flujo en el arco a
v :	Vector de flujo en los arcos (\dots, v_a, \dots).
\hat{v} :	Vector de flujo en los arcos observados mediante aforos(\dots, \hat{v}_a, \dots).
$\Delta = [\delta_{ak}]$:	Matriz de incidencia arco-camino, en la que el elemento δ_{ak} vale 1 si el camino k emplea el arco a y 0 en otro caso
$\Lambda = [\lambda_{ik}]$:	Matriz de incidencia par-camino, en la que el elemento λ_{ik} vale 1 si el camino k une el par i y 0 en otro caso
$c_a(v)$:	Coste de viaje en el arco a en función del flujo de todos los arcos de la red
$c(v) = [c_a(v)]$:	Vector de funciones de coste en los arcos
$C_k(h)$:	Coste de viaje en el camino k en función del flujo de todos los caminos de la red
$C(h) = [C_k(h)]$:	Vector de funciones de coste en los caminos

Tabla 2.2: Algoritmo de generación de columnas (CGA)

-
0. (*Inicialización*). Determinar una matriz inicial de viajes O-D y denotarla por g^1 , inicializar el conjunto $K^0 = \emptyset$ y tomar $\ell = 1$. Fijar un parámetro de tolerancia $\varepsilon > 0$.
 1. (*Problema de asignación de tráfico*). Resolver el TAP para la matriz de viajes origen-destino g^ℓ . Sea $K(g^\ell)$ el conjunto de caminos de mínimo coste en el equilibrio para la matriz O-D g^ℓ .
 2. (*Chequeo de la optimalidad*). Si $K(g^\ell) \subset K^{\ell-1}$, entonces parar g^ℓ es un mínimo local del DAP. En caso contrario tomar $K^\ell = K^{\ell-1} \cup K(g^\ell)$.
 3. (*Problema DAP restringido*). Encontrar un mínimo (local o global) del problema RDAP(K^ℓ) y denotarlo por $g^{\ell+1}$.
 4. (*Criterio de paro*). Si $g^{\ell+1}$ satisface el criterio de convergencia

$$\sum_{i \in I} \frac{|g_i^{\ell+1} - g_i^\ell|}{g_i^\ell} \leq \varepsilon,$$

parar, en caso contrario tomar $\ell = \ell + 1$ y regresar el paso 1.

Tabla 2.3: Definición de los CGAs trancados

Nombre	R. Complementariedad ($\pi^T h = 0$)	Eliminación Caminos
CGA	si	no
CGA2	no	si
CGA3	no	no

Tabla 2.4: Definición de los experimentos

Experimento	DAP	F
1, 3	DAP0	$\sum_{i \in \{1,2,3,4\}} (g_i - \hat{g}_i)^2 + \sum_{a \in \mathcal{B}} (v_a - \hat{v}_a)^2$
	DAP1	$\sum_{i \in \{1,2,3,4\}} (g_i - \hat{g}_i)^2 + \sum_{a \in \mathcal{A}} (v_a - \hat{v}_a)^2$
	DAP2	$\sum_{a \in \mathcal{B}} (v_a - \hat{v}_a)^2$
	DAP3	$\sum_{a \in \mathcal{A}} (v_a - \hat{v}_a)^2$
donde		$\hat{g} = (360, 640, 540, 160)$ $\hat{v} = (675.14, 524.86, 102.57, 697.43, 416.19, 361.53,$ $356.42, 184.63, 102.57, 253.85, 502.57, 497.43,$ $561.53, 632.06, 497.43, 438.47, 124.35, 400.00,$ $561.53)$ $\mathcal{B} = \{a_1, a_2, a_4, a_{13}, a_{14}, a_{13}\}$
2	DAP0	$\sum_{i \in \{1,2\}} (g_i - \hat{g}_i)^2 + \sum_{a \in \mathcal{C}} (v_a - \hat{v}_a)^2$
	DAP1	$\sum_{i \in \{1,2\}} (g_i - \hat{g}_i)^2 + \sum_{a \in \mathcal{A}} (v_a - \hat{v}_a)^2$
	DAP2	$\sum_{a \in \mathcal{C}} (v_a - \hat{v}_a)^2$
	DAP3	$\sum_{a \in \mathcal{A}} (v_a - \hat{v}_a)^2$
donde		$\hat{g} = (8, 16)$ $\hat{v} = (1.03433, 8.96567, 17.54423, 1.03433, 0.00000,$ $2.45577, 0.53670, 10.00000, 17.00754, 0.00000,$ $0.00000, 2.99246, 14.40415, 10.00000, 2.60339,$ $17.39661)$ $\mathcal{C} = \{a_4, a_7, a_{10}, a_{13}\}$
Problema interior		
Experimento	Red	Funciones coste en los arcos
1	Nguyen-Dupuis	Descritas en Nguyen and Dupuis (1984)
2	Friesz-Harker	Descritas en Friesz and Harker (1985)
3	Nguyen-Dupuis	Descritas en Nguyen and Dupuis (1984) exceptuando los arcos a_{14} y a_{18} que son
		$s_{14}(v_{14}) := \begin{cases} \bar{s}_{14}(v_{14}) & \text{if } v_{14} \leq 400 \\ \bar{s}_{14}(v_{14}) + 8.914 & \text{if } v_{14} > 400 \end{cases}$ $s_{18}(v_{18}) := \begin{cases} \bar{s}_{18}(v_{18}) & \text{if } v_{18} \leq 300 \\ \bar{s}_{18}(v_{18}) + 11.501 & \text{if } v_{18} > 300 \end{cases}$ Nguyen and Dupuis (1984)

Tabla 2.5: Resultados numéricos del experimento 1

		Punto arranque $g^1 = (1000, 1000, 1000, 1000)$						Punto arranque $g^1 = (100, 100, 100, 100)$					
		Matrices O-D estimadas						Matrices O-D estimadas					
DAP	Algorit.	g_1	g_2	g_3	g_4	F^o	$\#^b$	g_1	g_2	g_3	g_4	F^o	$\#^b$
0	CGA	404.27	721.35	580.72	211.03	19271.58	2	404.27	721.35	580.72	211.03	19271.58	3
	CGA1	404.27	721.35	580.72	211.03	19271.58	3	404.27	721.35	580.72	211.03	19271.58	3
	CGA2	404.27	721.35	580.72	211.03	19271.58	5	404.27	721.35	580.72	211.03	19271.58	3
	CGA3	404.27	721.35	580.72	211.03	19271.58	3	404.27	721.35	580.72	211.03	19271.58	3
1	CGA	395.44	748.13	588.33	206.32	23470.46	2	395.44	748.13	588.33	206.32	23470.46	3
	CGA1	395.44	748.13	588.33	206.32	23470.46	3	395.44	748.13	588.33	206.32	23470.46	3
	CGA2	395.44	748.13	588.33	206.32	23470.46	3	395.44	748.13	588.33	206.32	23470.46	3
	CGA3	395.44	748.13	588.33	206.32	23470.46	3	395.44	748.13	588.33	206.32	23470.46	3
2	CGA	400.00	800.01	600.01	199.99	0.00	2	400.00	800.00	600.00	200.00	0.00	3
	CGA1	400.00	800.00	600.00	200.00	0.00	2	400.00	800.00	600.00	200.00	0.00	3
	CGA2	400.00	800.00	600.00	200.00	0.00	5	400.00	800.00	600.00	200.00	0.00	3
	CGA3	799.02	775.33	617.27	148.46	84062.40	2	400.00	800.00	600.00	200.00	0.00	3
3	CGA	400.00	800.00	600.00	200.00	0.00	2	400.00	800.00	600.00	200.00	0.00	3
	CGA1	400.00	800.00	600.00	200.00	0.00	3	400.00	800.00	600.00	200.00	0.00	3
	CGA2	400.00	800.00	600.00	200.00	0.00	3	400.00	800.00	600.00	200.00	0.00	3
	CGA3	400.00	800.00	600.00	200.00	0.00	3	400.00	800.00	600.00	200.00	0.00	3

^o Función objetivo

^b Número de iteraciones

Tabla 2.6: Resultados numéricos del experimento 2

		Punto de arranque $g^1 = (6, 9)$				Punto de arranque $g^1 = (18, 17)$			
		Matrices O-D estimadas				Matrices O-D estimadas			
DAP	Algorit.	g_1	g_2	F^a	ℓ^b	g_1	g_2	F^a	ℓ^b
0	CGA	8.00	17.47	6.92	2	8.92	17.47	7.53	2
	CGA1	8.00	17.47	6.92	2	8.92	17.47	7.53	2
	CGA2	8.00	17.47	6.92	2	8.92	17.47	7.53	3
	CGA3	8.00	17.47	6.92	2	11.95	17.47	22.11	2
1	CGA	9.53	18.99	14.97	3	9.53	18.99	14.97	3
	CGA1	9.53	18.99	14.97	3	9.53	18.99	14.97	3
	CGA2	9.53	18.99	14.97	3	9.53	18.99	14.97	3
	CGA3	9.53	18.99	14.97	3	11.95	18.99	37.82	2
2	CGA	6.00	20.00	1.07	2	14.95	20.00	0.10	2
	CGA1	6.00	20.00	1.07	2	14.95	20.00	0.10	2
	CGA2	6.00	20.00	1.07	2	14.95	20.00	0.10	2
	CGA3	6.00	20.00	1.07	2	14.95	20.00	0.10	2
3	CGA	10.00	20.00	0.00	3	10.00	20.00	0.00	2
	CGA1	10.00	20.00	0.00	3	10.00	20.00	0.00	2
	CGA2	10.00	20.00	0.00	3	10.00	20.00	0.00	3
	CGA3	10.00	20.00	0.00	3	11.95	20.00	10.24	2

^a Función objetivo^b Número de iteraciones

Tabla 2.7: Resultados numéricos del experimento 3

		Punto arranque $g^1 = (1000, 1000, 1000, 1000)$						Punto arranque $g^1 = (10, 10, 10, 10)$					
		Matrices O-D estimadas						Matrices O-D estimadas					
DAP	Algorit.	g_1	g_2	g_3	g_4	F	ℓ	g_1	g_2	g_3	g_4	F	ℓ
0	CGA	406.78667.81	525.88	208.03	103492.83	5	406.78	667.81525.88	208.03	103492.83	4		
	CGA1	450.69675.65	711.68	157.54	108209.85	3	406.78	667.81525.88	208.03	103492.83	6		
	CGA2	450.69675.65	711.68	157.54	108209.85	4	406.78	667.81525.88	208.03	103492.83	7		
	CGA3	879.40904.88	560.60	301.15	779324.83	4	388.39	679.17492.95	219.58	105711.44	3		
1	CGA	358.18678.28	487.39	221.04	212264.67	3	358.18	678.28487.39	221.04	212264.67	4		
	CGA1	358.18678.28	487.39	221.04	212264.67	4	358.18	678.28487.39	221.04	212264.67	5		
	CGA2	358.18678.28	487.39	221.04	212264.67	7	358.18	678.28487.39	221.04	212264.67	5		
	CGA3	850.611037.96	848.87	159.59	1724307.24	4	358.18	678.28487.39	221.04	212264.58	4		
2	CGA	543.72653.75	1064.21	7.71	687.76	3	965.06	234.94	370.84	326.59	79555.32	5	
	CGA1	543.72653.75	1064.21	7.71	687.76	3	1185.07	0.00	123.92	565.52	81320.06	4	
	CGA2	543.72653.75	1064.21	7.71	687.76	6	1185.07	0.00	123.92	565.52	81320.08	4	
	CGA	383.74678.41	434.26	262.89	198995.02	3	383.74	678.41434.26	262.89	198995.02	5		
3	CGA1	383.74678.41	434.26	262.89	198995.02	4	383.74	678.41434.26	262.89	198995.37	12		
	CGA2	383.74678.41	434.26	262.89	198995.02	6	383.74	678.41434.26	262.89	198995.37	14		
	CGA3	914.20788.11	416.31	404.86	843812.22	4	911.10	31.04	0.00	1200.00	896430.95	6	

^a Función Objetivo^b Número de iteraciones

Capítulo 3

Algoritmos heurísticos de generación de columnas para la estimación de matrices de viajes O-D en redes congestionadas

Resumen

En este capítulo desarrollamos un marco para la elaboración de algoritmos heurísticos para el problema de ajuste de matrices origen-destino en redes de tráfico congestionadas (DAP)¹. Esta clase de algoritmos surge de la aplicación de estrategias heurísticas de resolución a un algoritmo convergente de generación de columnas para el DAP. Los algoritmos propuestos en la literatura SAB e IEA pertenecen a esta clase. Aquí se hace una nueva interpretación del SAB e IEA que presenta fundamentalmente dos ventajas: i) permite una más eficiente implementación computacional, al usar el espacio de flujo en los caminos en lugar del flujo espacio de flujo en los arcos originalmente empleado, y, también, ii) permite el establecimiento de condiciones para la convergencia a puntos estacionarios (puntos de Karush-Khun-Tucker) para el SAB. El capítulo se completa con un estudio computacional sobre diversas implementaciones en redes de pequeña escala, en las que se evalúa el rendimiento de los algoritmos propuestos y se prueban varias propuestas de modelización para el DAP que conducen a la es-
método MinMax.

Palabras claves:

¹Investigación parcialmente financiada por fondos FEDER y el proyecto CICYT BFM2002-04548-C03-03

CAPÍTULO 3. ALGORITMOS HEURÍSTICOS DE GENERACIÓN DE COLUMNAS PARA LA ESTIMACIÓN DE MATRICES DE VIAJES O-D EN REDES CONGESTIONADAS, programación matemática con restricciones de equilibrio, optimización de gran escala, optimización no diferenciable.

3.1 Introducción

La programación matemática con restricciones de equilibrio (MPEC) es una herramienta adecuada para la modelización de ciertos problemas importantes de planificación, control y gestión del transporte urbano. Los ejemplos más notables, en redes de tráfico congestionadas, son el denominado problema de diseño de redes (NDP), el problema de estimación de matrices O-D (DAP) o el problema del control semafórico.

Un MPEC consta de dos problemas relacionados: el *problema interior* o también llamado *problema del nivel inferior* y el *problema exterior* o *problema del nivel superior*. En el DAP el problema interior está definido por un problema de asignación en equilibrio parametrizado por la matriz de viajes O-D (variables del nivel superior) y el problema exterior está definido por un problema de estimación de la matriz de viajes O-D en el que interviene los flujos observados en la red.

Los MPECs presentan tres dificultades esenciales. La primera proviene del hecho de poseer malas propiedades matemáticas, como la no convexidad y la no diferenciabilidad; la segunda dificultad es el gran tamaño de los problemas reales y la tercera es que la aplicación punto-conjunto que define los flujos en equilibrio no se conoce explícitamente, estando únicamente definida implícitamente por un problema de optimización o de desigualdades variacionales. Estas dificultades han conducido al desarrollo de algoritmos de una marcada naturaleza heurística.

Dos son las estrategias que, *a grosso modo*, se han desarrollado para la resolución del DAP. La primera se basa en aplicar técnicas de optimización, comúnmente usadas en optimización de un solo nivel, como son los algoritmos de direcciones de descenso. Los métodos descritos en Spiess (1990b), Drissi-Kaitouni and Lundgren (1992), Florian and Chen (1993), Drissi-Kaitouni and Lundgren (1992), Codina and Barceló (2000), Codina et al. (2001) y Codina and Montero (2003) constituyen ejemplos de esta línea de investigación en la que se obtienen diferentes aproximaciones del gradiente de la función objetivo.

La segunda estrategia se basa en aproximar la aplicación punto-conjunto, que define el problema interior, por una aproximación lineal mediante los llamados *factores de influencia*. Los ejemplos más notables son el algoritmo IEA, propuesto en Yang et al. (1992), que está basado en el uso de proporciones para obtener los factores de influencia; y el SAB, dado en Yang (1995), en el que se calculan los factores de influencia mediante análisis de sensibilidad. En Codina et al. (2001) plantean el cálculo de las aproximaciones lineales de forma diferente. En ese trabajo se propone obtener esta la función objetivo del problema interior original. El método propuesto, denominado EAM, es el resultado de considerar una aproximación de primer orden de la función objetivo del nivel inferior.

Recientemente, Cascetta and Postorino (2001) formulan el DAP como la localización de un punto fijo de una función definida implícitamente mediante la solución de un problema de programación matemática. Esta formulación permite el desarrollo de

algoritmos de tipo punto-fijo.

En García and Verastegui (2003), el DAP se formula en el espacio de flujo en los caminos y se propone un algoritmo de generación de columnas (caminos) para evitar la enumeración explícita de todos los caminos de la red. Este algoritmo resuelve iterativamente un *problema de asignación* para una matriz O-D dada, cuya solución permite obtener nuevos caminos, y un *problema de estimación*, en el que se resuelve el DAP original restringido al conjunto de rutas generadas. En el DAP restringido se reemplaza el problema inferior restringido por sus condiciones de optimalidad de primer orden, las llamadas condiciones de Karush-Kuhn-Tucker en el caso de modelos de optimización. Esta operación conduce a un problema no lineal y no convexo de un solo nivel. Este algoritmo se denomina CGA.

El algoritmo CGA requiere la solución de un problema de estimación de grandes dimensiones, no lineal y no convexo en cada iteración. Desde un punto de vista computacional, se debe mantener un compromiso entre la exactitud de la solución obtenida, para el problema de estimación, y el esfuerzo computacional necesario para obtenerla. La implementación del algoritmo CGA pasa por resolver heurísticamente los problemas de estimación, lo que origina los algoritmos que hemos denominado *CGA heurísticos*. La estrategia más sencilla consiste en aproximar linealmente las condiciones de equilibrio de las restricciones. En este trabajo mostramos que ciertas aproximaciones lineales conducen a los algoritmos SAB, IEA y EAM.

Una estrategia complementaria, para la aplicación de los CGA heurísticos a problemas de gran escala, consiste en considerar las métricas derivadas de la $\|\cdot\|_1$ y de $\|\cdot\|_\infty$, las cuales conducen a una función objetivo lineal para el DAP (restringido) que, unido a las anteriores aproximaciones lineales, convierten los problemas de estimación en problemas lineales.

El capítulo se completa con una experiencia numérica, sobre redes de pequeñas dimensiones, con el fin de estudiar la convergencia de los CGA heurísticos y testar la viabilidad de ciertas estrategias desarrolladas. Los resultados obtenidos son prometedores, mostrando la convergencia a un mínimo local en muy pocas iteraciones.

3.2 Algoritmos heurísticos de generación de columnas

puesto en García and Verastegui (2003) para resolver el DAP. En la subsección 3.2.2 introducimos estrategias heurísticas en el CGA para su implementación a gran escala y en la subsección 3.2.3 relacionamos estas clases de algoritmos con los desarrollados en la literatura.

La notación empleada a lo largo de este trabajo, por conveniencia, se resume en la Tabla 3.1.

3.2.1 Algoritmo de generación de columnas para el DAP

El problema de asignación de tráfico en redes congestionadas, en su forma más general, admite una formulación mediante desigualdades variacionales. En este trabajo se considera la formulación en el espacio de flujo en los caminos. Encontrar un $h^* \in \Omega$ tal que

$$C(h^*)^T(h - h^*) \geq 0, \quad \forall h \in \Omega(g) \quad [\text{VIP}(C, \Omega(g))]$$

donde el espacio factible de flujos en los caminos, para una matriz de viajes O-D, g , está definido por $\Omega(g) = \{h \mid \Lambda h = g, h \geq 0\}$.

El problema de estimación o ajuste de una matriz O-D en una red de tráfico congestionada (DAP) se formula mediante la programación matemática con restricciones de equilibrio. Matemáticamente, el DAP puede ser formulado en el espacio de flujo en los caminos por

$$\begin{aligned} \text{Minimizar} \quad & F(g, h) = \eta_1 F_1(g) + \eta_2 F_2(v) \\ \text{sujeto a:} \quad & C(h)^T(h' - h) \geq 0, \quad \forall h' \in \Omega(g) \\ & h \in \Omega(g) \\ & v = \Delta h \\ & 0 \leq g \leq \bar{g} \end{aligned} \quad [\text{DAP}]$$

donde las funciones F_1 y F_2 son dos métricas que miden respectivamente las discrepancias de las observaciones disponibles de la matriz de demanda O-D y de los aforos en la red con sus respectivas predicciones. Los coeficiente $\eta_1, \eta_2 \geq 0$ reflejan la confianza que tenemos en la calidad de nuestras observaciones. Bajo las hipótesis de que las funciones F_1 y F_2 y $c(v)$ son continuas, Chen and Florian (1996) demostraron que el DAP admite una solución.

Obsérvese que si la función objetivo del nivel superior se multiplica por el número positivo $\frac{1}{\alpha, +\infty}$ la solución del problema no varía. Esto pone de manifiesto que el DAP es y por otro intentar mantener la misma estructura que la matriz O-D desactualizada. Esto nos mueve a considerar una formulación de la función objetivo como $F(g, h) = \eta F_1(g) + (1 - \eta) F_2(v)$ donde $\eta \in [0, 1]$.

García and Verastegui (2003) transforman el DAP en un problema de optimización de un sólo nivel reemplazando el $\text{VIP}(S, \Omega)$ mediante sus condiciones de optimalidad, obteniendo

$$\begin{aligned}
 &\text{Minimizar } F = \eta F_1(g) + (1 - \eta) F_2(v) \\
 &\text{sujeto a: } C(h) = \Lambda^T U + \pi \\
 &\quad \Lambda h = g \\
 &\quad \Delta h = v \qquad \qquad \qquad \text{[DAP]} \\
 &\quad \pi^T h = 0 \\
 &\quad \pi, h \geq 0 \\
 &\quad 0 \leq g \leq b
 \end{aligned}$$

Este problema no se puede abordar mediante algoritmos que trabajen directamente sobre la formulación explícita, ya que el número de variables (debido fundamentalmente al número de caminos) es prohibitivo para problemas reales, incluso de mediano tamaño.

El algoritmo que presentamos en esta sección es un método iterativo que construye una sucesión de aproximaciones del DAP. Cada uno de estos problemas aproximados, denominados *DAP restringidos*, está definido por la restricción del problema DAP sobre un subconjunto de caminos $\hat{K} \subset K$ y proporcionan una estimación de la matriz de viajes O-D. Posteriormente se resuelve el problema de asignación en equilibrio para la matriz O-D estimada, lo que genera nuevos caminos, ampliando el anterior subconjunto \hat{K} . El algoritmo termina cuando no es añadido ningún nuevo camino a este subconjunto. El algoritmo propuesto para el resolver el DAP va generando nuevas columnas (caminos) en cada iteración según van siendo necesitadas, este es el origen del nombre *algoritmo de generación de columnas* (CGA).

El problema DAP *restringido* al conjunto de caminos \hat{K} se formula:

$$\begin{aligned}
 &\text{Minimizar } F = \eta F_1(g) + (1 - \eta) F_2(v) \\
 &\text{sujeto a:} \qquad \qquad \qquad \text{[RDAP}(\hat{K})\text{]} \\
 &\quad \hat{C}(\hat{h}) = \Lambda^T U + \hat{\pi} \qquad \qquad \qquad (3.1) \\
 &\quad \Lambda \hat{h} = g \qquad \qquad \qquad (3.2) \\
 &\quad \hat{\Delta} \hat{h} = v \qquad \qquad \qquad (3.3) \\
 &\quad \hat{\pi}, \hat{h} \geq 0 \qquad \qquad \qquad (3.5) \\
 &\quad 0 \leq g \leq b \qquad \qquad \qquad (3.6)
 \end{aligned}$$

donde $\hat{\cdot}$ indica la correspondiente matriz o vector restringidos al conjunto de caminos \hat{K} .

La ventaja del subproblema RDAP(\hat{K}) respecto al problema DAP es que el número de variables es significativamente menor y esto permite resolver (aproximadamente) directamente la formulación explícita.

La tabla 3.2 recoge el algoritmo CGA. Este algoritmo tiene asegurada su convergencia a un óptimo local en un número finito de iteraciones bajo hipótesis de continuidad en las funciones F_1 , F_2 y el coste en los caminos, y bajo una débil condición de regularidad.

3.2.2 Algoritmos heurísticos de generación de columnas para el DAP

El algoritmo CGA resuelve un problema DAP restringido en cada iteración. El número de variables de este problema se incrementa conforme progresa el algoritmo. Este hecho, unido a las restricciones no lineales de equilibrio, pueden hacer inviable este algoritmo para los problemas de gran escala que aparecen en la práctica. Esto motiva que el problema DAP restringido solo se pueda resolver de modo aproximado. En esta sección se discute algunas de las estrategias que se pueden emplear para aproximar este problema.

El primer elemento de complejidad del DAP restringido es su gran tamaño, y éste se puede paliar reduciendo el número de caminos empleados en la definición del DAP restringido. La elección más natural es tomar $\hat{K}^\ell = K^\ell$, esto es, considerar únicamente los caminos en equilibrio de la iteración actual. Esta elección puede hacer perder la convergencia del algoritmo, debido a algún proceso de ciclaje, pero si este converge, esto es, que en dos iteraciones consecutivas se repita el mismo conjunto de caminos \hat{K}^ℓ , la matriz de viajes O-D estimada es un mínimo local del DAP.

El segundo elemento de complejidad son el las restricciones no lineales (3.1) y (3.4). El conjunto de restricciones (3.1) nos proporcionan el coste en los caminos y la restricción (3.4) recoge la condición de holgura complementaria que afirma que el flujo en el camino, de un par determinado, es cero, o, en caso contrario, el coste del camino es exactamente el coste de equilibrio del par. Las restricciones (3.1) se pueden reemplazar por aproximaciones lineales para el coste de los caminos. Las restricción (3.4) se puede *relajar* (eliminar), ya que si asumimos que $\hat{K}^\ell = K^\ell$, cabe esperar que los caminos actuales serán empleados en el equilibrio y la condición de holgura complementaria se

Las tres simplificaciones anteriores: i) eliminación de caminos generados con anterioridad a la iteración actual, ii) aproximación lineal del coste en los caminos y iii) relajación de la condición de holgura complementaria, conducirán a RDAPs con restricciones lineales en el que la función objetivo será no lineal. En concreto los métodos clásicos de estimación de mínimo error absoluto, mínimos cuadrados generalizados y MinMax emplean, para la función objetivo, las métricas:

$$F^1(x, \hat{x}) = \sum_i \omega_i |x_i - \hat{x}_i|$$

$$F^2(x, \hat{x}) = \sum_i \omega_i (x_i - \hat{x}_i)^2$$

$$F^\infty(x, \hat{x}) = \max_i \{\omega_i |x_i - \hat{x}_i|\}$$

para medir las discrepancia entre un vector estimado x y el vector observado \hat{x} . Los parámetros ω_i son un conjunto de pesos determinados a priori.

Las métricas F^1 y F^∞ tienen la ventaja que los RDAPs se pueden transformar a problemas lineales mediante la introducción de nuevas variables. En el caso de la estimación mínimo error absoluto la transformación tendrá la expresión:

$$\begin{aligned} \text{minimizar } Z &= \sum_i \omega_i \varepsilon_i \\ \text{sujeto a:} \\ x_i - \hat{x}_i &\leq \varepsilon_i, \quad i = 1, \dots, n \\ \hat{x}_i - x_i &\leq \varepsilon_i, \quad i = 1, \dots, n \\ \varepsilon_i &\geq 0, \quad i = 1, \dots, n \end{aligned}$$

En el caso de la estimación MinMax la transformación sería

$$\begin{aligned} \text{minimizar } Z &= \varepsilon \\ \text{sujeto a:} \\ \omega_i (x_i - \hat{x}_i) &\leq \varepsilon, \quad i = 1, \dots, n \\ \omega_i (\hat{x}_i - x_i) &\leq \varepsilon, \quad i = 1, \dots, n \\ \varepsilon &\geq 0 \end{aligned}$$

3.2.3 Algunos ejemplos

En la literatura hay varios algoritmos que pueden ser descritos como un algoritmo CGA truncado. El ejemplo más sencillo es el dado en Codina et al. (2001) en el que todos los caminos de K^ℓ se considera que tienen un coste independiente del vector de flujo, siendo éste el coste en equilibrio. Esta suposición conduce a poder eliminar las restricciones no lineales de equilibrio y quedarse únicamente con la satisfacción de la

Otros algoritmos asumen que los costes en los caminos son lineales. Bajo estas hipótesis, como muestra el siguiente lema, la situación de equilibrio se puede calcular explícitamente en función de la matriz de viajes O-D proporcionada.

LEMMA 3.2.1 *Análisis de la sensibilidad para el equilibrio con costes lineales en los caminos.*

Supóngase que los elementos del par (h^ℓ, U^ℓ) definen respectivamente el flujo y el coste en el equilibrio para una cierta matriz de viajes O - D g^ℓ . Asumimos que el vector h^ℓ satisface la condición de holgura complementaria estricta. Supongamos que el coste en los caminos viene dado por la expresión $C(h) := a + Bh$ para una cierta matriz B y un cierto vector a . Supóngase que \hat{B} es una matriz invertible, entonces existe un entorno de g^ℓ donde el flujo en los caminos en la situación de equilibrio viene dado por las ecuaciones:

$$\begin{aligned}\hat{h} &= \hat{h}^\ell + \hat{B}^{-1}(\hat{\Lambda})^T(\hat{\Lambda}B^{-1}\hat{\Lambda}^T)^{-1}(g - g^\ell) \\ U &= U^\ell + (\hat{\Lambda}\hat{B}^{-1}\hat{\Lambda}^T)^{-1}(g - g^\ell)\end{aligned}$$

donde $\hat{\cdot}$ indica la restricción del vector o matriz sobre el conjunto de caminos con flujo positivo en la situación de equilibrio para la matriz O - D g^ℓ ; y toma el valor cero en el resto de componentes del vector de flujo.

DEMOSTRACIÓN. La situación de equilibrio viene definida por la solución del siguiente sistema de ecuaciones y desigualdades parametrizado por la matriz demanda g :

$$\begin{aligned}a + Bh &= \Lambda^T U + \pi \\ \Lambda h &= g \\ h^T \pi &= 0 \\ \pi, h &\geq 0\end{aligned}$$

Veamos si el anterior sistema de ecuaciones admite una solución en la que las componentes que tenían flujo positivo para la matriz g^ℓ siguen teniendo flujo para la nueva matriz g y los caminos que no tenían flujos siguen sin tenerlo. Esta situación conduce a que $\hat{\pi} = 0$ y que el sistema restringido se escriba:

$$\hat{a} + \hat{B}\hat{h} = \Lambda^T U \tag{3.7}$$

$$\hat{\Lambda}\hat{h} = g \tag{3.8}$$

$$\hat{h} \geq 0 \tag{3.9}$$

Despejando el valor de \hat{h} de la ecuación (3.7), obtenemos

$$\hat{h} = \hat{B}^{-1}(-\hat{a} + \hat{\Lambda}^T U) \tag{3.10}$$

Sustituyendo la ecuación (3.10) en (3.8), obtenemos la expresión

$$\hat{\Lambda}\hat{B}^{-1}\hat{a} = \hat{\Lambda}\hat{B}^{-1}\hat{\Lambda}^T U - \hat{g} \tag{3.11}$$

Como la anterior ecuación también se cumple para el par (h^ℓ, U^ℓ)

$$\hat{\Lambda}\hat{B}^{-1}\hat{a} = \hat{\Lambda}\hat{B}^{-1}\hat{\Lambda}^T U^\ell - g^\ell \tag{3.12}$$

Sustituyendo la expresión (3.12) en (3.11), obtenemos

$$U = U^\ell + (\hat{\Lambda}\hat{B}^{-1}\hat{\Lambda}^T)^{-1}(g - g^\ell) \tag{3.13}$$

Sustituyendo el valor de los costes de equilibrio obtenido en (3.13) en la expresión (3.10),

$$\hat{h} = \hat{B}^{-1}(-\hat{a} + \hat{\Lambda}^T[U^\ell + (\Lambda\hat{B}^{-1}\Lambda^T)^{-1}(g - g^\ell)])$$

Teniendo en cuenta que en la situación de equilibrio para la matriz g^ℓ se tiene que $\hat{a} + \hat{B}\hat{h}^\ell = \Lambda^T U^\ell$, la anterior expresión simplifica a la relación buscada

$$\hat{h} = \hat{h}^\ell + \hat{B}^{-1}(\hat{\Lambda})^T(\hat{\Lambda}\hat{B}^{-1}\hat{\Lambda}^T)^{-1}(g - g^\ell)$$

Nos falta por probar que los flujos \hat{h} son no negativos. Notar que la dependencia de los flujos respecto de la matriz g viene dada por una función continua. Por la hipótesis de la holgura complementaria estricta, tenemos que $\hat{h}^\ell > 0$, por tanto existe un entorno de g^ℓ de modo que los flujos definidos por esa relación continua siguen siendo positivos.

Tras esta discusión hemos resuelto el anterior sistema de ecuaciones y desigualdades y, por tanto, hemos calculado explícitamente la situación de equilibrio. Notar que ésta se puede seguir expresando en el espacio de flujos restringidos. \square

La matriz $\hat{B}^{-1}(\hat{\Lambda})^T(\hat{\Lambda}\hat{B}^{-1}\hat{\Lambda}^T)^{-1}$ se puede calcular explícitamente cuando la matriz \hat{B} es diagonal. Supóngase que las funciones de coste en los caminos admiten la expresión $C_k(h_k) := A_k + B_k h_k$ con $k \in \hat{K}_i$ e $i \in I$, entonces el elemento k -ésimo de la diagonal de $\hat{B}^{-1}(\hat{\Lambda})^T(\hat{\Lambda}\hat{B}^{-1}\hat{\Lambda}^T)^{-1}$ es

$$\frac{1}{\sum_{k' \in \hat{K}_i} \frac{1}{B_{k'}}} \quad (3.14)$$

En la siguiente proposición mostramos como se puede obtener el algoritmo IEA, descrito en Yang et al. (1992), dentro de este tipo de aproximaciones. La ventaja que tiene esta formulación es que permite más fácilmente integrar algoritmos con importantes propiedades de reoptimización, como el DSD (Larsson and Patriksson (1992)), en la resolución del DAP; ya que estos algoritmos trabajan en el espacio de flujo en los caminos.

PROPOSITION 3.2.2 (ALGORITMO IEA) *El algoritmo IEA es un CGA heurístico en el que se consideran las aproximaciones lineales $C_i(\hat{h}) \approx \frac{U_i^\ell}{\hat{B}_i} \hat{h}$, con $k \in \hat{K}_i$, $i \in I$. $U_i^\ell \succ$*

DEMOSTRACIÓN. Por hipótesis, \hat{B}^{-1} es invertible y aplicando el lema 3.14 para costes en los arcos diagonales, ecuación (3.14), y se obtiene que la asignación a los caminos se hace de acuerdo a la relación:

$$h_k = h_k^\ell + \frac{h_k^\ell}{\sum_{k' \in \hat{K}_i} h_{k'}^\ell} (g_i - g_i^\ell), \quad k \in \hat{K}_i, i \in I$$

Esto indica que la variación de la demanda se asigna a los caminos proporcionalmente al flujo que llevaban los caminos en la situación de equilibrio para la matriz g^ℓ .

Por tanto la asignación a los arcos satisface la relación

$$\begin{aligned} v_a &= \sum_{i \in I} \sum_{k \in \hat{K}_i} \delta_{ak} h_k = \sum_{i \in I} \sum_{k \in \hat{K}_i} \delta_{ak} \left[h_k^\ell + \frac{h_k^\ell}{\sum_{k' \in \hat{K}_i} h_{k'}^\ell} (g_i - g_i^\ell) \right] = \\ &= v_a^\ell + \sum_{i \in I} \left[\sum_{k \in \hat{K}_i} \delta_{ak} \frac{h_k^\ell}{\sum_{k' \in \hat{K}_i} h_{k'}^\ell} (g_i - g_i^\ell) \right] = v_a^\ell + \sum_{i \in I} p_{ai}^\ell (g_i - g_i^\ell) \end{aligned}$$

donde p_{ai}^ℓ es la proporción de usuarios del par i que emplean el arco a para la matriz O-D g_i .

Lo que muestra que la asignación a los arcos se hace proporcional al flujo que llevaba en la situación de equilibrio para la matriz g^ℓ . \square

PROPOSITION 3.2.3 (ALGORITMO SAB) *Asumiendo las hipótesis i) $C(h)$ diferenciable en Ω , ii) $d^T \nabla C(h^\ell) d > 0$ para toda dirección factible de Ω en h^ℓ iii) h^ℓ satisface la condición de holgura complementaria estricta y iv) $\Omega(g^\ell)$ satisface la cualificación de restricciones de independencia lineal entonces el algoritmo SAB de Yang (1995) se puede considerar un CGA heurístico para las aproximaciones lineales $C_k(h) \approx a_k + B_k h_k$, con $k \in \hat{K}_i, i \in I$ donde los coeficientes son $a_k = U_i^\ell - \frac{1}{\partial h_k / \partial g_i} h_k^\ell$ y $B_k = \frac{1}{\partial h_k / \partial g_i}$ para todo $k \in \hat{K}_i$ e $i \in I$.*

DEMOSTRACIÓN. El algoritmo SAB aproxima, en cada iteración, las condiciones de equilibrio mediante los llamados *factores de influencia* definidos como las derivadas del flujo en los arcos respecto de las variables de demanda en los pares. Con lo que obtendríamos:

$$v_a = v_a^\ell + \sum_{i \in I} \frac{\partial v_a}{\partial g_i} (g_i - g_i^\ell) \quad \forall a \in \mathcal{A},$$

El conjunto de hipótesis i) – iv) son las suposiciones del Teorema 6. de Tobin and Friesz (1988). Este teorema garantiza que existen las derivadas parciales $\frac{\partial v_a}{\partial g_i}$ y satisfacen

$$\frac{\partial v_a}{\partial g_i} = \sum_{k \in \hat{K}_i} \delta_{ak} \frac{\partial h_k}{\partial g_i}, \quad \forall a \in \mathcal{A}, \forall i \in I \quad (3.15)$$

independientemente del punto extremo h^ℓ elegido.

Aplicando el resultado del lema 4.2.2 a la aproximación lineal dada en el enunciado de esta proposición (\hat{B} es diagonal y sus elementos no nulos), obtenemos:

$$\frac{\frac{1}{B_i}}{\sum_{k' \in K_i^e} \frac{1}{B_{k'}}} = \frac{\frac{\partial h_k}{\partial g_i}(g^f)}{\sum_{k \in K_i} \frac{\partial h_k}{\partial g_i}(g^f)}$$

Por otro lado, bajo estas hipótesis, existe un entorno de g^f en el que las funciones $\hat{h}(g)$, definidas implícitamente como la solución del problema TAP respecto a la matriz g , son funciones continuamente diferenciables que además satisfacen la relación:

$$\sum_{k \in K_i} h_k(g) = g_i, \quad \forall i \in I \quad (3.16)$$

donde $h_k(g)$ es el flujo en equilibrio para el camino k y para la matriz de demanda O-D g . Tomando derivadas, en ambos lados de la igualdad respecto a la variable g_i , obtenemos

$$\sum_{k' \in K_i} \frac{\partial h_{k'}}{\partial g_i}(g^f) = 1. \quad (3.17)$$

Teniendo en cuenta la expresión (3.17), obtenemos el resultado deseado de que la demanda es asignada a los caminos de acuerdo a la proporción $\frac{\partial h_k}{\partial g_i}$ con $k \in K_i^f, i \in I$. En este caso el lema 4.2.2 adopta la expresión:

$$h_k = h_k^f + \frac{\partial h_k}{\partial g_i}(g - g_i)$$

que, junto a la relación (3.15), nos indica que los flujos en los arcos son asignados como el algoritmo SAB. \square

La aplicación del algoritmo SAB a problemas reales es limitada ya que, para el cálculo de $\nabla_g v$, se debe invertir una matriz cuyas dimensiones vienen definidas por el número de coordenadas positivas del vector de flujo en equilibrio en los caminos. El número de caminos puede ser demasiado grande para ser computacionalmente aplicable a grandes dimensiones. En el siguiente resultado se muestra una formulación alternativa del SAB, como un CGA heurístico, en la que no es necesario obtener explícitamente dichas derivadas. Este resultado tiene una gran importancia para la aplicación del método a problemas reales, ya que en la formulación original el cálculo de los factores de influencia se efectúa mediante análisis de sensibilidad.

dérese que en los subproblemas DAP restringidos se consideran las siguientes aproximaciones de primer orden a los costes en los arcos, $c^f(v) := v^f + \nabla_v c(v^f)(v - v^f)$. Además se supone que la sucesión de flujos en los caminos en equilibrio $\{h^f\}$ generada por el algoritmo son no degenerados, esto es, cumplen la condición de holgura complementaria estricta, entonces el CGA aproximado resultante coincide con el algoritmo SAB.

DEMOSTRACIÓN. El CGA resultante asigna flujo en los caminos a través de la solución del siguiente sistema de ecuaciones lineales escrito de forma matricial:

$$\begin{aligned}\hat{C}^\ell(\hat{h}) &= (\hat{A}^\ell)^T U \\ \hat{A}^\ell \hat{h} &= g\end{aligned}$$

donde $\hat{C}^\ell(\hat{h})$ es el coste de los caminos restringido al conjunto K^ℓ para los costes en los arcos dado por $c^\ell(v)$, \hat{A}^ℓ y $\hat{\Delta}^\ell$ son respectivamente las matrices de incidencia par-camino y la matriz de incidencia arco-camino restringida a los caminos de K^ℓ . Las variables de este sistema son los flujos en los caminos de K^ℓ , \hat{h} , y el vector de costes en equilibrio U para el conjunto de los pares. Esta solución viene parametrizada por la demanda g .

El coste en los caminos C^ℓ se puede expresar

$$\begin{aligned}\hat{C}^\ell(\hat{h}) := (\hat{\Delta}^\ell)^T c^\ell(v) &= (\hat{\Delta}^\ell)^T [c(v^\ell) + \nabla_v c(v^\ell)(v - v^\ell)] = (\hat{\Delta}^\ell)^T c(v^\ell) + [(\hat{\Delta}^\ell)^T \nabla_v c(v^\ell)(\hat{\Delta}^\ell)(\hat{h} - \hat{h}^\ell) \\ &= (\hat{A}^\ell)^T U^\ell + \nabla_{\hat{h}} \hat{C}(\hat{h}^\ell)(\hat{h} - \hat{h}^\ell)]\end{aligned}$$

La expresión (3.18) permite reescribir el sistema de ecuaciones lineales por

$$\nabla_{\hat{h}} \hat{C}(\hat{h}^\ell)(\hat{h} - \hat{h}^\ell) = \hat{A}^\ell (U - U^\ell) \quad (3.19)$$

$$\hat{\Delta}^\ell \hat{h} = g \quad (3.20)$$

Vamos a demostrar que el modo de asignar el flujo según el algoritmo SAB satisface el anterior sistemas de ecuaciones y, como este sistema de ecuaciones lineales tiene solución única debido a que h^ℓ es un punto no degenerado, Tobin and Friesz (1988), entonces el algoritmo CGA aproximado coincide con el SAB.

La asignación de flujo en los caminos, según el algoritmo SAB, satisface

$$\hat{h} = \hat{h}^\ell + \nabla_g \hat{h}(g^\ell)^T (g - g^\ell) \quad (3.21)$$

$$U = U^\ell + \nabla_g U(g^\ell)^T (g - g^\ell) \quad (3.22)$$

Bajo las hipótesis de la proposición, existe un entorno de g^ℓ de modo que la situación de equilibrio viene descrita por

Derivando a ambos lados de (3.23) respecto a la variable g , obtenemos

$$\nabla_g C(h(g)) = \hat{A}^\ell \nabla_g U(g)^T \quad (3.24)$$

Por otro lado, aplicando la regla de la cadena, tenemos:

$$\nabla_g \hat{C}(\hat{h}(g)) = \nabla_{\hat{h}} \hat{C}(\hat{h}) \cdot \nabla_g \hat{h}(g)^T \quad (3.25)$$

Sustituyendo la solución SAB dada en la ecuación (3.22) y teniendo en cuenta las relaciones (3.24) y (3.25), obtenemos

$$\nabla_{\hat{h}} \hat{C}(\hat{h}^\ell) (\hat{h}^\ell + \nabla_g \hat{h}(g^\ell)^T (g - g^\ell) - \hat{h}^\ell) = \nabla_g \hat{C}(\hat{h}^\ell) (g - g^\ell) \quad (3.26)$$

$$= \Lambda^\ell \nabla_g U(g^\ell)^T (g - g^\ell) \quad (3.27)$$

y por tanto se satisface la primera ecuación del sistema.

Para verificar el cumplimiento de la ecuación (3.22), considérese la restricción de satisfacción de la demanda

$$\Lambda^\ell \hat{h}(g) = g$$

que se satisface en un entorno de g^ℓ . Derivando a ambos lados respecto de g , obtenemos

$$\hat{\Lambda}^\ell \nabla_g \hat{h}(g)^T = \hat{I} \quad (3.28)$$

donde \hat{I} es la matriz identidad de orden $|I|$

Comprobemos la ecuación (3.22),

$$\hat{\Lambda}^\ell \hat{h} = \hat{\Lambda}^\ell \left[\hat{h}^\ell + \nabla_g \hat{h}(g^\ell)^T (g - g^\ell) \right] = g^\ell - \hat{I} (g - g^\ell) = g$$

□

Una observación sobre la reformulación del SAB es que siempre está definida, incluso cuando no se satisfacen las hipótesis de holgura complementaria estricta, necesarias en el cálculo de los factores de influencia, que aparecen en la formulación original del SAB.

En la siguiente proposición demostramos como el SAB puede ser expresado como un CGA heurístico en el que el conjunto de caminos almacenados se va actualizando en cada iteración. En esta proposición modificamos el SAB original, para tener garantizada su convergencia, haciendo que todos los caminos generados en las iteraciones anteriores sean almacenados para las futuras iteraciones y empleados a la hora de definir los DAP restringidos. Además, bajo la hipótesis de holgura complementaria estricta, tenemos que el punto al que converge es un KKT del DAP.

eliminado a la hora de definir K^ , entonces el SAB converge en un número finito de iteraciones a una matriz g' . Además si los flujos en equilibrio satisfacen la condición de holgura complementaria estricta para dicha matriz, entonces la matriz O-D g' tiene asociada un punto de KKT para el DAP.*

DEMOSTRACIÓN. El algoritmo termina cuando el conjunto restringido de caminos de la red, K^ℓ , no es ampliado. Como el conjunto de caminos en la red es finito, entonces

se debe cumplir en una determinada iteración que $K^{\ell+1} = K^\ell$ y por tanto se satisface el criterio de paro del algoritmo. Esto prueba que el algoritmo converge en un número finito de iteraciones a una matriz g' .

La matriz g' cumple que es un mínimo local de $RDAP(K^\ell)$ y satisface la *cualificación de restricciones de Slater* (Bazaraa et al. (1993)), ya que el conjunto de restricciones está formado por un conjunto de restricciones lineales (y por tanto funciones convexas) y una única restricción no lineal ($\hat{\pi}^T \hat{h} = 0$) cuyo gradiente es no nulo por satisfacerse la condición de holgura complementaria estricta. Esto prueba que g' tiene asociado un punto de KKT para el $RDAP(K^\ell)$.

Ahora extenderemos este punto de KKT del $RDAP(K^\ell)$ al problema original DAP. La diferencia entre ambos problemas es que las restricciones (3.1) y las de no negatividad están dadas para todos los caminos de la red en el problema original, mientras que en el problema restringido $RDAP$ sólo están definidas para el subconjunto de caminos K^ℓ , además lo mismo ocurre para el flujo en los caminos. Para extender este punto de KKT basta con tomar los multiplicadores asociados a las nuevas restricciones iguales a 0 y el flujo de los caminos no contenidos en K^ℓ igual a cero. Con esta elección no varía la condición de estacionalidad del Lagrangiano y la condición de holgura complementaria. La única condición que faltaría por comprobar es la factibilidad de este punto en el DAP, esto es, que el flujo sigue estando en equilibrio para el conjunto de todos los caminos. Como en la última iteración no se ha añadido ningún nuevo camino esto indicaría que no hay ningún camino fuera del conjunto K^ℓ que tenga un coste menor que los caminos de K^ℓ y por tanto los multiplicadores π_p serán no negativos para todo camino p fuera de K^ℓ . Esto prueba que el punto de KKT extendido sigue siendo un punto de KKT para el DAP. \square

Notar que esta proposición también se puede enunciar poniendo como hipótesis la convergencia del SAB en lugar de suponer el almacenamiento de todos los caminos generados.

3.3 Experiencia numérica

En esta sección completamos el análisis teórico desarrollado en las secciones anteriores han perseguido dos objetivos claramente diferenciados: En primer lugar se evalúan varias propuestas de modelización para el DAP, que conducen, cada una de ellas, a un método de estimación de matrices O-D, posteriormente se estudia el rendimiento computacional de los algoritmos CGA heurísticos aplicados a estos modelos.

Esta diferencia entre objetivos nos ha conducido a separar los test numéricos en los siguientes experimentos:

- Experimento 1: *evaluación de modelos DAP*. La adopción de un criterio estadístico en la estimación de matrices O-D conduce a diferentes propuestas de modelización para el DAP. En este experimento hemos analizado mediante simulación de Monte Carlo los métodos de estimación mínimo cuadrática, mínimo error absoluto y la estimación MiniMax.
- Experimento 2: *evaluación de los algoritmos CGA truncados en los modelos DAPs*.

3.3.1 Experimento 1: evaluación de modelos DAP

En este experimento vamos a analizar tres métodos populares de estimación: mínimos cuadrados generalizados (The least square weighted (LSW) method), mínimo error absoluto (The least-absolute-value (LAV) method), y el método MiniMax (Minimax (MM) method). Cada uno de estos métodos emplea respectivamente las métricas F^1 , F^2 y F^∞ . Para esta sección se han elegidos para todos ellos los pesos $\omega_i = 1$.

La especificación de un DAP requiere de: i) una métrica para $F_1(g)$ ii) una métrica para $F_2(v)$ y iii) la calibración del parámetro η . En este experimento emplearemos el mismo tipo de métrica para i) y ii), surgiendo así tres tipos de DAPs que denominaremos, respectivamente, DAP_1 , DAP_2 y DAP_∞ . Las estimaciones obtenidas con cada uno de ellos las denominaremos respectivamente: $\hat{g}_1(\eta)$, $\hat{g}_2(\eta)$ y $\hat{g}_\infty(\eta)$ remarcando con esta notación que la estimación depende directamente del valor elegido del parámetro η .

Cuando deseamos comparar la capacidad predictiva de los modelos DAP_1 , DAP_2 y DAP_∞ entre sí surgen dos dificultades fundamentales:

- ¿Qué valor de η elegir en cada modelo? Esta cuestión la hemos resuelto eligiendo el valor de η en cada modelo DAP de modo que nos de la *mejor* estimación posible de la verdadera matriz O-D.
- ¿Cual es el estadístico más adecuado para medir la calidad de la estimación? Tradicionalmente, en la literatura, se emplea el error cuadrático medio (MSE). El hecho de que la elección de un determinado estadístico pudiera influir en los resultados obtenidos, nos ha hecho considerar también el error absoluto medio (MAE) y el máximo error absoluto (ME).

Los items a) y b) nos han conducido a considerar los estadísticos

$$MAE(\hat{g}, g) = \frac{1}{n_{od}} \min_{\eta \in [0,1]} \left\{ \sum_{i \in I} |\hat{g}_i(\eta) - g_i| \right\}$$

$$MSE(\hat{g}, g) = \frac{1}{n_{od}} \min_{\eta \in [0,1]} \left\{ \sum_{i \in I} (\hat{g}_i(\eta) - g_i)^2 \right\}$$

$$ME(\hat{g}, g) = \min_{\eta \in [0,1]} \left\{ \max_{i \in I} (|\hat{g}_i(\eta) - g_i|) \right\}$$

donde g es la verdadera matriz OD y \hat{g} es la mejor matriz estimada por un determinado DAP.

La denominada target-matriz, o matriz de referencia, se puede interpretar como una estimación inicial de la matriz O-D. El DAP mejora esta estimación inicial incorporando en el proceso de estimación la información suministrada en los aforos de la red de tráfico. Es por este motivo que unos mejores indicadores de la bondad de un determinado modelo DAP sean:

$$\begin{aligned}\%red_1(\bar{g}, \hat{g}) &= \frac{MAE(\bar{g}, g) - MAE(\hat{g}, g)}{MAE(\bar{g}, g)} \\ \%red_2(\bar{g}, \hat{g}) &= \frac{MSE(\bar{g}, g) - MSE(\hat{g}, g)}{MSE(\bar{g}, g)} \\ \%red_\infty(\bar{g}, \hat{g}) &= \frac{ME(\bar{g}, g) - ME(\hat{g}, g)}{ME(\bar{g}, g)}\end{aligned}$$

donde \bar{g} es la target-matriz. Estos estadísticos aplicados a los resultados obtenidos por un determinado DAP nos miden, de tres formas diferentes, el porcentaje de reducción del error contenido en la target-matriz al incorporar los aforos.

En este experimento nos hemos planteado una experiencia de simulación de Monte Carlo para evaluar los anteriores índices sobre los DAP₁, DAP₂ y DAP_∞. Para ello hemos considerado la red de tráfico dada en Nguyen and Dupuis (1984) (ver figura 3.1). Esta red posee 4 pares O-D, 13 nodos y 19 arcos.

Hemos perturbado aleatoriamente la verdadera matriz O-D y sus flujos, en equilibrio bajo el principio DUE, para generar una muestra aleatoria de target-matrices y de flujos observados en todos los arcos de la red. Más concretamente, los aforos fueron obtenidos sumando a los verdaderos flujos en equilibrio un término aleatorio generado mediante variables normales independientes, de media cero y coeficiente de variación cu_v , (cociente entre la desviación estándar y el valor medio de la distribución). La target-matriz O-D fue obtenida perturbando la verdadera matriz O-D mediante la adición de un término aleatorio a cada par de media cero y coeficiente de variación cu_g . Las matrices de varianza y covarianza de las anteriores variables aleatorias de flujos y pares de demanda se asumieron diagonales, calculándose las varianzas a través de las expresiones

$$\sigma_{v_s}^2 = (cu_v \cdot v_s) \quad \sigma_{g_i}^2 = (cu_g \cdot g_i)$$

Los coeficientes de variación elegidos en este experimento fueron $cu_g = 0.40$ y $cu_v = 0.05$.

Una vez generados los datos para un determinado problema, hemos resuelto todos los DAP pertinentes. El algoritmo empleado es el CGA ya que tiene asegurada una convergencia a un óptimo local. Además las pruebas numéricas realizadas indican que estos problemas sólo poseen un óptimo global que, por tanto, es alcanzado.

El cálculo del valor óptimo de γ para cada índice se ha efectuado numéricamente. Se ha discretizado el intervalo $[0, 1]$ y se han considerado los valores del parámetro $\gamma \in \{0, 0.1, \dots, 0.9, 1\}$. El valor óptimo de γ se ha considerado el mejor de estos 11 valores. Para mejorar la exactitud de este procedimiento, sin recurrir a un refinamiento de la partición del intervalo, hemos escalado las métricas asociadas a los flujos y a la matriz estimada de la función objetivo del DAP ($\gamma \frac{F_1(s)}{a_1} + (1-\gamma) \frac{F_2(v)}{a_2}$) con los coeficientes que se muestran en la tabla 3.3.

Finalmente, se generó una muestra aleatoria de tamaño 50, lo que hizo necesario resolver $50 * 11 * 3 = 1650$ DAPs y muchos más problemas de asignación de tráfico. Los resultados obtenidos se resumen en la tabla 3.4.

Con la muestra aleatoria generada, realizamos un test estadístico para testar si los tres modelos tienen la misma capacidad estimativa o si, por el contrario, los resultados obtenidos dependen del modelo usado. Para ello empleamos el test no paramétrico de Friedman para K -muestras relacionadas (Siegel (1990)). Este test lo hemos realizado para las 3 medidas de reducción referidas anteriormente, los resultados se muestran en la tabla 3.5.

El test no paramétrico de Friedman asigna a cada modelo DAP valores de 1, 2 ó 3, siendo el valor 3 asignado al modelo que obtiene la mayor reducción, 2 a la intermedia y 1 a la menor. Se observa que los modelos DAP_1 y DAP_2 parecen ser similares, siendo el DAP_∞ el que produce los peores resultados, con una diferencia estadísticamente significativa.

Los test nos indican que existen diferencias significativas entre los tres DAPs empleados, medidas éstas con cualquiera de los índices $\%red_1$, $\%red_2$ o $\%red_\infty$. Además, parece no tener sentido nuestra preocupación sobre el modo de medir la bondad del DAP, ya que son similares los rangos asignados con los tres índices (columnas). También se observa que los modelos DAP_1 y DAP_2 obtienen resultados similares. Para testar esta afirmación utilizamos el test de Wilcoxon (Siegel (1990)) para estos dos modelos, los resultados se muestran en la tabla 3.6. Los resultados obtenidos muestran que ambos modelos no obtienen resultados estadísticamente diferentes y por tanto se podrían utilizar indistintamente.

3.3.2 Experimento 2: evaluación de algoritmos CGA heurísticos

El modelo cuadrático DAP_2 obtiene estimaciones similares. Desde el punto de vista computacional el modelo DAP_1 conduce a algoritmos CGA heurísticos en los que los DAP restringidos son lineales, mientras que éstos son cuadráticos en el DAP_2 . En este experimento profundizaremos en los aspectos computacionales del uso del DAP_1 y DAP_2 .

Se han realizado dos experimentos numéricos, que denominaremos test 1 y test 2, sobre dos redes de pequeño tamaño: red de Nguyen-Dupuis (Nguyen and Dupuis

(1984), en adelante ND) y red de Friesz-Harker (Friesz and Harker (1985), en adelante FH). Sus topologías se muestran en las figuras 3.1 y 3.2 respectivamente. La red ND tiene 13 nodos, 19 arcos y cuatro pares O-D: el par $(1 \rightarrow 2)$ con demanda g_1 , $(1 \rightarrow 3)$ con demanda g_2 , $(4 \rightarrow 2)$ con demanda g_3 y $(4 \rightarrow 3)$ con demanda g_4 . La red FH tiene 6 nodos, 16 arcos y dos pares O-D: el par $(1 \rightarrow 6)$ con demanda g_1 , $(6 \rightarrow 1)$ con demanda g_2 .

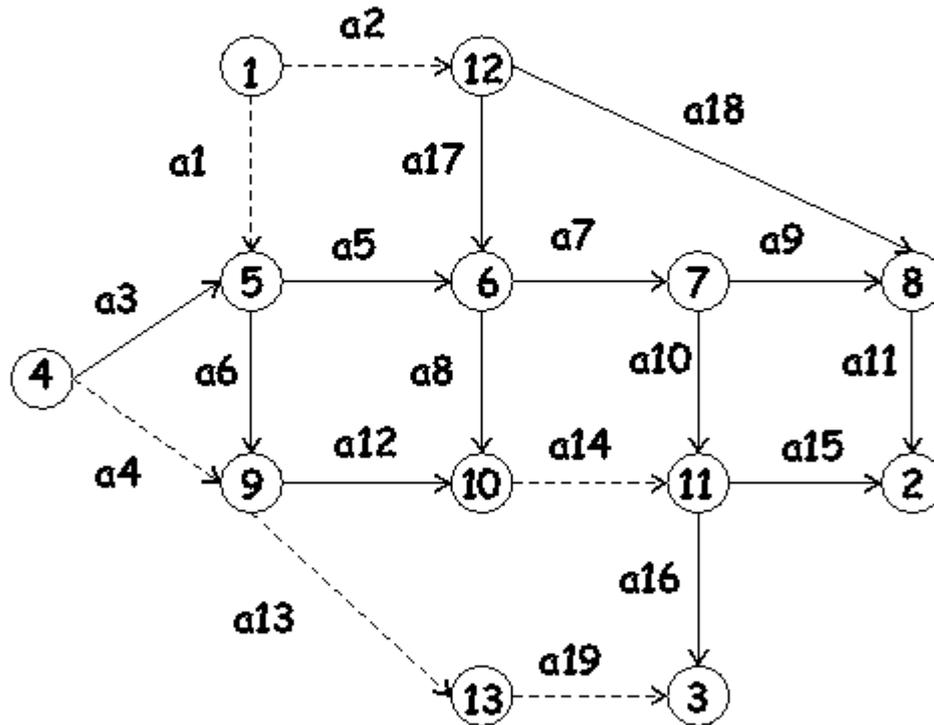


Figura 3.1: Topología de la red de Nguyen-Dupuis

Hemos considerado cuatro modelos de DAP caracterizados por el uso o no de una matriz de referencia (target matrix) y por el uso de aforos en todos los arcos de la red o sólo en un subconjunto de ellos. Todas las combinaciones se describen en la tabla 3.7. Se han elegido como parámetros de peso en la función objetivo $\tau_1 = 1$ y $\tau_2 = 1$.

Los flujos observados para el test 1 corresponden a la asignación de una matriz O-D $(400, 800, 600, 200)$ desde el punto de vista del usuario (TAP-UE), utilizando para los dos experimentos los costes de arcos dados en Nguyen and Dupuis (1984). Los flujos observados para el test 2 corresponden a la asignación de una matriz O-D $(200, 400, 600, 200)$ desde el punto de vista del usuario (TAP-UE), utilizando para los dos experimentos los costes de arcos dados en Friesz and Harker (1985).

Se han elegido dos puntos iniciales, uno asociado a una situación de gran congestión de las redes y otro a una de baja congestión.

Se han utilizado los algoritmos IEA, SAB-I (SAB implícito) y CGA. Todos ellos se han codificado usando el programa GAMS (General Algebraic Modeling Systems). El

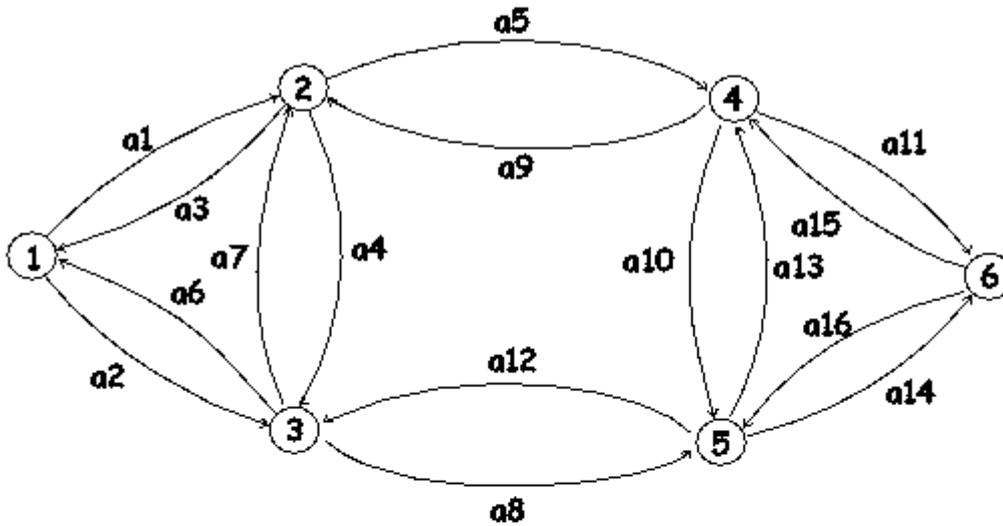


Figura 3.2: Topología de la red de Friesz-Harker

criterio de paro (Paso 4.) elegido para las pruebas computacionales fue $\epsilon = 0.0001$. Una rápida introducción a GAMS se puede ver en el libro de Castillo et al. (2002).

La proposición 3.2.3 pone de manifiesto que, bajo ciertas hipótesis, el SAB coincide con un CGA heurístico que denominaremos *SAB implícito* (SAB-I). Realmente el SAB-I está siempre definido, aunque las condiciones dadas en la proposición 3.2.3 no se satisfagan o incluso cuando las derivadas $\partial v_a / \partial g_i$ no existan. La primera prueba computacional tiene por objeto poner de manifiesto esta diferencia. Se ha codificado el SAB de dos formas. En la primera se calculan numéricamente las derivadas $\partial v_a / \partial g_i$ y la segunda forma es el SAB-I. En la mayoría de las pruebas en las que se ha variado el DAP a resolver y el punto inicial ambos esquemas generaban la misma sucesión de matrices estimadas. En la tabla 3.8 se recoge la evolución para la red de Nguyen-Dupuis. En ella se observa que, aunque generan dos sucesiones diferentes ambos convergen al mismo punto, realizándolo el SAB-I en menos iteraciones. Esto pone de manifiesto que no siempre ambos esquemas son coincidentes. En el resto de pruebas presentadas en esta subsección el SAB codificado es el SAB-I.

Los resultados obtenidos en el test 1 se muestran en la tabla 3.9 y 3.10. Se observa que el CGA y SAB-I alcanzan el óptimo en un menor número de iteraciones que el IEA. Además, el número de iteraciones empleados por los algoritmos son similares para el DAP_1 y el DAP_2 , exceptuando el caso del IEA para el caso de existir target matrix. existir una matriz de referencia. Esto es debido a la elección de los parámetros τ_1 y τ_2 en el DAP_1 y el DAP_2 .

Los resultados obtenidos en el test 2 se muestran en la tabla 3.11 y 3.12. Las observaciones efectuadas en el test 1 parecen seguir cumpliéndose, excepto el comportamiento del IEA que viene a ser similar al SAB-I. Cabe destacar que en el caso de la estimación mínimo cuadrática y para el DAP tipo 2 se llegan a óptimos locales del problema,

cuestión que no ocurre con la estimación mínimo error absoluto.

3.4 Conclusiones

En este trabajo se desarrolla un marco para elaborar algoritmos heurísticos, denominado CGA truncados, para la resolución del DAP. Se demuestra que los algoritmos SAB e IEA son ejemplos de esta clase. El hecho de ver estos dos algoritmos como algoritmos CGAs heurísticos tiene tres ventajas: i) permite una mejor implementación computacional en el espacio de flujo en los caminos en lugar de flujos en los arcos, ii) generaliza la definición del SAB y iii) permite obtener una condición suficiente para la convergencia del SAB a un punto KKT.

En las pruebas numéricas se ha dado un indicio que muestra que el DAP_1 constituye una alternativa al uso del DAP_2 . En la experiencia de simulación desarrollada el DAP_1 posee una capacidad predictiva similar al DAP_2 , pero el uso del DAP_1 en conjunción con ciertos tipos de CGA truncados tiene ventajas computacionales, en concreto, origina problemas *DAP restringidos* que son lineales en contraposición a los problemas cuadráticos originados por el DAP_2 , y todo esto no incrementa el número total de iteraciones.

Tabla 3.1: Notación empleada

\mathcal{A} :	Conjunto de arcos de la red
I :	Conjunto de pares O-D en la red
K :	Conjunto de caminos en la red
K_i :	Conjunto de caminos que unen en el par i
$i \in I$:	Un par O-D de la red
$k \in K$:	Un camino de la red
$a \in \mathcal{A}$:	Un arco de la red
g_i :	Demanda de viajes para el par O-D i
g :	Matriz de demanda de viajes O-D (\dots, g_i, \dots)
h_k :	Flujo en el camino k
h :	Vector de flujo en los caminos (\dots, h_k, \dots)
v_a :	Flujo en el arco a
v :	Vector de flujo en los arcos (\dots, v_a, \dots)
$\Delta = [\delta_{ak}]$:	Matriz de incidencia arco-camino, en la que el elemento δ_{ak} vale 1 si el camino k emplea el arco a y 0 en otro caso
$\Lambda = [\lambda_{ik}]$:	Matriz de incidencia par-camino, en la que el elemento λ_{ik} vale 1 si el camino k une el par i y 0 en otro caso
$c_a(v)$:	Coste de viaje en el arco a en función del flujo de todos los arcos de la red
$c(v) = [c_a(v)]$:	Vector de funciones de coste en los arcos
$C_k(h)$:	Coste de viaje en el camino k en función del flujo de todos los caminos de la
$C(h) = [C_k(h)]$:	Vector de funciones de coste en los caminos

Tabla 3.2: Algoritmo de generación de columnas (CGA)

-
0. (*Inicialización*). Determinar una matriz inicial de viajes O-D y denotarla por g^1 , inicializar el conjunto $K^0 = \emptyset$ y tomar $\ell = 1$. Sea $\epsilon > 0$ un parámetro de tolerancia.
 1. (*Problema de equilibrio*). Resolver el VIP($S, \Omega(g)$) para la matriz de viajes origen-destino $g = g^\ell$. Sea $K(g^\ell)$ el conjunto de caminos de mínimo coste en el equilibrio para la matriz O-D g^ℓ .
 2. (*Chequeo de la optimalidad*). Si $K(g^\ell) \subset K^{\ell-1}$, entonces parar g^ℓ es un mínimo local del DAP. En caso contrario tomar $K^\ell = K^{\ell-1} \cup K(g^\ell)$.
 3. (*Problema DAP restringido*). Encontrar un mínimo (local o global) del problema RDAP(K^ℓ) y denotarlo por $g^{\ell+1}$.
 4. (*Criterio de paro*). Si se satisface el criterio de convergencia, por ejemplo $\frac{\sum_{i \in I} |g_i^\ell - g_i^{\ell+1}|}{\sum_{i \in I} |g_i^\ell|} < \epsilon$, parar, en caso contrario tomar $\ell = \ell + 1$ y regresar el paso 1.
-

Tabla 3.3: Coeficientes de escalamiento para los DAPs

DAP	α_1	α_2
DAP_1	900	300
DAP_2	200000	10000
DAP_∞	400	50

Tabla 3.4: Resultados agregados

	DAP_1			DAP_2			DAP_∞		
Min	69.10	92.99	61.18	61.04	90.90	66.60	46.06	61.49	62.16
Max	98.38	99.99	98.98	98.44	99.98	98.68	95.29	99.87	96.91
Media	90.52	98.70	90.43	89.96	98.65	90.62	82.56	95.89	83.32
Desv. Est.	6.18	1.74	7.74	6.15	1.90	7.52	11.09	6.59	12.36

Tabla 3.5: Rangos y significatividad para el test de Friedman para la comparación de los DAPs

	% red ₁	% red ₂	% red _∞
DAP_1	2.44	2.44	2.30
DAP_2	2.30	2.26	2.36
$DAP_∞$	1.26	1.30	1.34
χ^2	41.56	37.56	32.76
gl	2.00	2.00	2.00
Sig. asint.	0.00	0.00	0.00

Tabla 3.6: Prueba de los rangos con signo de Wilcoxon

	% red ₁	% red ₂	% red _∞
Z	-1.487	-0.893	-0.275
Sit. Asint. (Bilateral)	0.137	0.372	0.783

Tabla 3.7: Descripción de los DAPs empleados en el experimento 2

TEST	DAP	MÁTRIZ REFERENCIA (\hat{g})	ARCOS AFORADOS ($\hat{\psi}$)
1	0	(360,640,540,160)	$a_1, a_2, a_4, a_{18}, a_{14}, a_{19}$
	1	(360,640,540,160)	TODOS
	2	NO	$a_1, a_2, a_4, a_{18}, a_{14}, a_{19}$
2	3	NO	TODOS
	1	(8,16)	TODOS
	2	NO	a_4, a_7, a_{10}, a_{18}
	3	NO	TODOS

Tabla 3.8: Diferencia entre las sucesiones generadas por los algoritmos SAB y el SAB-I

	ℓ	i_1	i_2	i_3	i_4	F^d
SAB	0.00	1000.00	1000.00	1000.00	1000.00	4455432.68
	1.00	799.05	773.29	613.99	165.58	300504.95
	2.00	545.42	622.85	544.22	205.45	61932.08
	3.00	444.24	686.86	581.77	210.46	22713.71
	4.00	404.27	721.35	580.72	211.03	19271.58
	5.00	404.27	721.35	580.72	211.03	19271.58
SAB-I	0.00	1000.00	1000.00	1000.00	1000.00	4455432.68
	1.00	435.05	740.52	594.60	206.14	22780.57
	2.00	404.27	721.35	580.72	211.03	19271.58
	3.00	404.27	721.35	580.72	211.03	19271.58

Tabla 3.9: Resultados test 1 para la estimación mínimo error absoluto

DAP	Algoritmo	Punto Inicial				F^1	ℓ	Punto Inicial				F^1	ℓ
		i_1	i_2	i_3	i_4			i_1	i_2	i_3	i_4		
0	IEA	400.00	799.93	599.94	200.09	300.44	14	400.00	799.86	600.00	200.16	300.76	15
	SAB-I	400.00	800.00	600.00	200.00	300.00	4	400.00	800.00	600.00	200.00	300.00	3
	CGA	400.00	800.00	600.00	200.00	300.00	2	400.00	800.00	600.00	200.00	300.00	3
1	IEA	400.00	799.93	599.94	200.09	300.44	14	400.00	799.86	600.00	200.16	300.76	15
	SAB-I	400.00	800.00	600.00	200.00	300.00	4	400.00	800.00	600.00	200.00	300.00	3
	CGA	400.00	800.00	600.00	200.00	300.00	2	400.00	800.00	600.00	200.00	300.00	3
2	IEA	400.00	799.99	600.00	200.05	0.17	11	400.00	799.99	599.96	200.07	0.32	13
	SAB-I	400.00	800.00	600.00	200.00	0.00	4	400.00	800.00	600.00	200.00	0.00	3
3	SAB-I	400.00	800.00	600.00	200.00	0.00	4	400.00	800.00	600.00	200.00	0.00	3
	CGA	400.00	800.00	600.00	200.00	0.00	2	400.00	800.00	600.00	200.00	0.00	3

Tabla 3.10: Resultados test 1 para estimación mínimo cuadrática

DAP	Algoritmo	Punto Inicial						F^2	ℓ	Punto Inicial						F^2	ℓ
		1000				1000				100				100			
		i_1	i_2	i_3	i_4					i_1	i_2	i_3	i_4				
0	IEA	401.76	715.55	580.73	215.05	19376.30	7	401.76	715.55	580.72	215.06	19376.35	7				
	SAB-I	404.27	721.35	580.72	211.03	19271.58	3	404.27	721.35	580.72	211.03	19271.58	3				
	CGA	404.27	721.35	580.72	211.03	19271.58	2	404.27	721.35	580.72	211.03	19271.58	3				
1	IEA	397.62	741.21	589.19	206.35	23623.83	7	397.60	741.23	589.20	206.33	23622.79	9				
	SAB-I	395.40	747.98	588.64	206.31	23470.81	3	395.44	748.12	588.34	206.32	23470.46	3				
	CGA	395.44	748.13	588.33	206.32	23470.46	2	395.44	748.13	588.33	206.32	23470.46	3				
2	IEA	400.11	799.92	599.93	200.05	0.01	11	400.50	799.64	599.86	200.12	0.08	14				
	SAB-I	400.00	800.01	600.01	199.99	0.00	3	400.00	800.00	600.00	200.00	0.00	3				
	CGA	400.00	800.01	600.01	199.99	0.00	2	400.00	800.00	600.00	200.00	0.00	3				
3	IEA	400.03	799.96	599.97	200.04	0.01	10	400.03	799.96	599.97	200.03	0.01	12				
	SAB-I	400.00	800.00	600.00	200.00	0.00	4	400.00	800.00	600.00	200.00	0.00	3				
	CGA	400.00	800.00	600.00	200.00	0.00	2	400.00	800.00	600.00	200.00	0.00	3				

Tabla 3.11: Resultados test 2 para estimación mínimo error absoluto

DAP	Algoritmo	Punto Inicial				Punto Inicial			
		6		9		18		27	
		i_1	i_2	F^1	ℓ	i_1	i_2	F^1	ℓ
0	IEA	10.00	20.00	6.00	3	10.00	20.00	6.00	3
	SAB-I	10.00	20.00	6.00	3	10.00	20.00	6.00	3
	CGA	10.00	20.00	6.00	3	10.00	20.00	6.00	2
1	IEA	10.00	20.00	6.00	3	10.00	20.00	6.00	3
	SAB-I	10.00	20.00	6.00	3	10.00	20.00	6.00	3
	CGA	10.00	20.00	6.00	3	10.00	20.00	6.00	2
2	IEA	10.00	20.00	0.00	3	10.00	20.00	0.00	3
	SAB-I	10.00	20.00	0.00	3	10.00	20.00	0.00	3
	CGA	10.00	20.00	0.00	2	10.00	20.00	0.00	2
3	IEA	10.00	20.00	0.00	3	10.00	20.00	0.00	3
	SAB-I	10.00	20.00	0.00	3	10.00	20.00	0.00	3
	CGA	10.00	20.00	0.00	2	10.00	20.00	0.00	2

Tabla 3.12: Resultados del test 2 para estimación mínimo cuadrática

		Punto Inicial				Punto Inicial			
		6	9			18	27		
DAP	Algoritmo	i_1	i_2	F^2	ℓ	i_1	i_2	F^2	ℓ
0	IEA	8.00	17.41	6.93	3.00	8.00	17.41	6.93	3.00
	SAB-I	8.00	17.47	6.92	3.00	8.00	17.47	6.92	3.00
	CGA	8.00	17.47	6.92	2.00	8.92	17.47	7.53	2.00
1	IEA	9.40	18.97	15.06	4.00	9.40	18.97	15.06	5.00
	SAB-I	9.53	18.99	14.97	4.00	9.53	18.99	14.97	4.00
	CGA	9.53	18.99	14.97	3.00	9.53	18.99	14.97	3.00
2	IEA	0.00	20.00	1.07	4.00	8.12	20.00	1.07	3.00
	SAB-I	0.00	20.00	1.07	3.00	10.00	20.00	0.00	4.00
	CGA	6.00	20.00	1.07	2.00	14.95	20.00	0.10	2.00
3	IEA	10.00	20.00	0.00	4.00	10.00	20.00	0.00	4.00
	SAB-I	10.00	20.00	0.00	3.00	10.00	20.00	0.00	3.00
	CGA	10.00	20.00	0.00	3.00	10.00	20.00	0.00	2.00

Capítulo 4

Ajuste de las funciones de coste en los arcos en los modelos de asignación de tráfico en equilibrio

Resumen

En este capítulo se aborda el problema de ajustar (o estimar) los parámetros de las funciones de coste en los arcos (PCPCA) de una red de tráfico congestionada, mediante el uso de observaciones de flujos en los arcos y tiempos/costes de viaje en equilibrio en un subconjunto de pares de demanda origen-destino. Se proponen dos programas matemáticos con restricciones de equilibrio (MPEC) para modelizar este problema. El primer modelo asume que la información disponible se ha obtenido en un único periodo (definido por una única matriz de demanda origen-destino). Esta información permite estimar los tiempos/costes de congestión en un subconjunto de arcos pero no la parametrización de las funciones de coste. El segundo modelo asume que los inputs empleados en el primer modelo son obtenidos en dos o más periodos diferentes, esto es, correspondientes a matrices de demanda origen-destino distintas. Este mayor nivel de información posibilita el ajuste de la parametrización de las funciones de coste.

Se establece la existencia de soluciones de los dos MPECs planteados bajo condiciones débiles en las funciones de coste en los arcos y se realiza un análisis de la sensibilidad de estas soluciones.

MPECs. En un paso previo el MPEC es transformado en un programa matemático equivalente de un sólo nivel, en el espacio de flujo en los caminos, mediante la formulación del problema de asignación de tráfico como un problema de complementariedad no lineal; y posteriormente se aplica, sobre el modelo equivalente, un esquema modificado de descomposición simplicial desagregada en el que la fase de generación de columnas está definida mediante un problema no lineal de asignación de tráfico. Se

CAPÍTULO 4. AJUSTE DE LAS FUNCIONES DE COSTE EN LOS ARCOS EN LOS MODELOS DE ASIGNACIÓN DE TRÁFICO EN EQUILIBRIO

ha demostrado, haciendo uso del análisis de sensibilidad, que el algoritmo propuesto converge a óptimos locales del problema PCPCA.

Se ha realizado un estudio numérico sobre redes de pequeño y mediano tamaño, con los objetivos de estudiar los modelos PCPA propuestos y, por otro lado, aspectos computacionales del algoritmo de resolución. Finalmente, se ha aplicado el método a la red de Ciudad Real (España) para evaluar aplicabilidad de esta metodología a ciudades de pequeño-mediano tamaño.

Palabras claves:

Ajuste de parámetros, asignación de tráfico en equilibrio, programación matemática con restricciones de equilibrio, métodos de generación de columnas, problemas de flujos en redes multiproducto

4.1 Introducción

En la planificación, gestión y control de redes de transporte se emplean multitud de modelos matemáticos. Todos ellos usan una gran cantidad de datos de entrada que son, a menudo, muy costosos de obtener. Esta circunstancia ha motivado un considerable esfuerzo de investigación dedicado tanto a los problemas teóricos como a los problemas prácticos derivados de las calibraciones de los modelos. Esto es así, porque la fiabilidad de las predicciones depende significativamente, pero no exclusivamente, de la precisión de los parámetros utilizados en ellos.

Una de las tareas más frecuentemente requerida, en la práctica de la planificación y gestión del tráfico, es la asignación de una matriz de demanda de viajes origen-destino a una red dada. Con este fin se han propuesto multitud de modelos matemáticos para simular cómo los usuarios de una red de tráfico eligen su ruta para satisfacer su viaje. En este trabajo nos centramos en los llamados *modelos estáticos de asignación en equilibrio*, aunque lo aquí expuesto se puede extender a otros modelos de equilibrio en redes, como los empleados para la asignación de pasajeros en redes de transporte público.

Los modelos estáticos de asignación de tráfico asumen una situación estacionaria, bajo la cual, el sistema permanece invariable a lo largo del periodo de estudio. La situación de equilibrio queda definida mediante valores medios en los flujos, tiempos y demandas. El *primer principio de Wardrop* ((Wardrop 1952b)), también denominado DUE (*deterministic user equilibrium*), constituye un marco para la elaboración de estos modelos de asignación en equilibrio. Este principio asume que todos los usuarios perciben el coste de la misma manera y, además, conocen los costes de todas las rutas (tienen información perfecta). Bajo este principio, en la situación de equilibrio, ningún usuario puede reducir su tiempo de viaje cambiando unilateralmente de ruta. En la realidad las percepciones de los costes están sujetas a variaciones y los usuarios eligen la ruta de acuerdo con su percepción. Para modelar esta situación, se han elaborado modelos de equilibrio en los que los costes de viajes son la suma de una parte fija más una componente aleatoria. En este caso los usuarios eligen una u otra ruta dependiendo de la distribución de probabilidad de los costes aleatorios. Esta forma de asignación recibe el nombre *asignación estocástica* (SUE).

Los modelos de asignación de tráfico requieren de expresiones funcionales que proporcionen el tiempo/coste de viaje en cada arco en función del vector de flujo en los arcos de la red. El llamado *problema asimétrico de asignación de tráfico* permite recoger. En muchas ocasiones, debido a la dificultad de formalizar estas relaciones, se recurre a considerar una aproximación mediante un modelo simétrico de asignación en el que el coste en cada arco sólo depende del flujo en el mismo. Las funciones de coste dependen de un conjunto de parámetros. Algunos de ellos son estimados basándose en consideraciones físicas de los propios arcos, como su longitud, su capacidad o la velocidad media en el arco sin flujo; pero para otros existe un cierto grado de libertad.

Estas y otras simplificaciones introducen errores que pueden afectar significativamente a la capacidad predictiva de los modelos de asignación en equilibrio. Una tarea importante, para poder minimizar el efecto de estos errores y poder obtener un buen modelo de asignación que pueda reproducir los flujos observados en la realidad, es la calibración de estas funciones de coste. Esta tarea se vuelve crucial cuando se desea estimar los tiempos de viajes (en escenarios futuros) con una cierta fiabilidad. Ahondando en esta cuestión, se puede dar la situación de que se tuviesen dos parametrizaciones de la misma red que reprodujesen fielmente un cierto conjunto de observaciones actuales de flujo y, sin embargo, las predicciones de tiempos de viajes, para un hipotético escenario futuro, fueran radicalmente diferentes.

En los últimos años, los investigadores han realizado muchos esfuerzos para proponer métodos de estimación que combinaran información agregada de la demanda de transporte, como la contenida en las observaciones de flujo en los arcos, con otras fuentes de información como datos desagregados de la demanda obtenida de la realización de encuestas o el uso de un conjunto de estimaciones iniciales. La ventaja de estos métodos es que mejoran las estimaciones, al emplear toda la información disponible, y reducen los costes de realización y los tiempos de ejecución. Estas metodologías han sido principalmente aplicadas en tres campos: i) la actualización y/o estimación de matrices origen - destino, ii) la calibración de modelos de demanda de viajes agregados y iii) la estimación de los parámetros de modelos combinados de asignación en equilibrio. En Willumsem (1981), Tamin and Willumsen (1989), Ortúzar and Willumsen (1994), Bell and Iida (1997) y en Cascetta (2001) se presenta una revisión de estos modelos.

Durante décadas se ha realizado un gran esfuerzo investigador para desarrollar modelos combinados de asignación en equilibrio, que recogiesen el mayor número de aspectos relevantes en el comportamiento de los usuarios y de la red de transporte (elección de modo de transporte, ruta, destino, realización del viaje, múltiple tipos de usuarios, percepción aleatoria de los costes de transportes, etc.) Este esfuerzo ha ido acompañado del desarrollo de algoritmos de resolución que permitían estimar ciertas predicciones (outputs) en función de la parametrización (input) del modelo. En la actualidad el problema que está adquiriendo importancia es el problema inverso, en el que se desea calcular una parte de la parametrización del modelo a partir de un conjunto de observaciones de los distintos items capaces de predecir el modelo. Un ejemplo de estos tipos de problemas se puede consultar en García and Marín (1998).

El objetivo de este trabajo es desarrollar una metodología que permita calibrar los parámetros de las funciones de coste en los arcos (PCPCA), con procedimientos automáticos y de bajo coste monetario y temporal. La información básica de los tiempo de viaje en equilibrio para un subconjunto de pares origen-destino o, en su defecto, el valor de alguna función de estos valores, como el tiempo medio en la red. Toda esta información será recogida en uno o en dos periodos diferentes de tiempo, esto es, periodos asociados a diferentes matrices de viaje origen-destino.

El PCPCA, que está relacionado con el *problema inverso de flujo en redes multipro-*

ducto, consistente en encontrar unos costes en los arcos tales que un patrón de flujo dado sea una solución del problema de flujo en redes para dichos costes y además estos costes satisfaga algún criterio de optimalidad (Rydergren (2001), Larsson et al. (2003)). Las dos dificultades para adaptar el problema aquí tratado a este marco son: i) los flujos observados en los arcos pueden ser inconsistentes, esto es, que no satisfacen las condiciones de conservación de flujo y/o ii) la naturaleza del problema de calibración no requiere que los flujos observados en los arcos sean reproducidos exactamente.

Una herramienta adecuada para formular el PCPCA es la *programación matemática con restricciones de equilibrio (MPEC)* (Luo et al. (1996)). Este modelo matemático se formula como sigue:

$$\begin{aligned} \min_x F(x, y) \\ x \in X \end{aligned}$$

donde y es una solución óptima del problema de desigualdades variacionales siguiente: Encontrar $y \in Y(x)$ tal que

$$c(y, x)^T (y - y') \geq 0, \quad \forall y' \in Y(x) \quad \text{[VIP]}$$

donde $Y(x)$ es la región factible para la variable y para una x dada.

El problema $\min_{x \in X} F(x, y)$ se denomina *problema de nivel superior* o *problema exterior*, y *problema interior* o *problema de nivel inferior* a VIP. La variable x recibe el nombre de variable del nivel superior e y variable del nivel inferior.

Actualmente, varios MPEC se han aplicado al problema de ajustar (o estimar) matrices origen destino en redes congestionadas (DAP), o al diseño de redes de transporte (NDP). El NDP determina una expansión de la capacidad de los arcos (NDP continuo, CDNP) o una modificación de la topología de la red existente (NDP discreto, DNNDP) de tal forma que se obtenga el máximo beneficio social.

Los problemas del nivel superior del DAP y del NDP son respectivamente un problema convencional de estimación y uno de diseño de redes. El problema del nivel inferior está definido por un modelo de asignación en equilibrio, parametrizado por las variables del nivel superior, para ambos problemas. La principal diferencia desde un punto de vista matemático entre el DAP y el NDP es que las variables del nivel superior solo parametrizan las restricciones del problema del nivel inferior para el DAP y únicamente la función objetivo del nivel inferior para el NDP.

problema híbrido entre un DAP y un NDP, en el cual el problema exterior es similar al del DAP y el problema interior es como el del NDP.

Suh et al. (1990) presentaron un modelo binivel para el problema PCPCA con el objeto de estimar los parámetros de las funciones de coste en los arcos empleando como información el flujo en los arcos y asumiendo las condiciones de equilibrio DUE. El método que ellos presentaron no encuentra soluciones adecuadas al problema.

Xu et al. (2002) desarrollan un procedimiento secuencial para estimar la matriz de viajes O-D y calibrar los parámetros bajo el principio SUE.

Meng et al. (2002) proponen un modelo MPEC para estimar simultáneamente la matriz de viajes O-D y calibrar los parámetros de las funciones de coste de viaje en el arco, bajo las condiciones de equilibrio DUE, a partir de flujos y tiempos de viaje observados en los arcos. Estos autores transforman el anterior problema binivel en uno no lineal y diferenciable a través de la función *gap* y propone la aplicación del método del Lagrangiano aumentado para resolverlo.

El MPEC es potencialmente un problema de optimización no convexo y no diferenciable. Consecuentemente, las técnicas tradicionales para resolver este tipo de problemas, tales como el método del gradiente reducido (Tan et al. (1979)), no funcionan adecuadamente y pueden detenerse en puntos que no son estacionarios para MPEC.

Se han propuesto un buen número de algoritmos heurísticos para encontrar una solución aproximada al NDP. El conocido como *algoritmo de diseño/asignación* itera entre un problema de equilibrio de usuario, con las variables de diseño fijadas a un cierto valor, y un problema de diseño con los flujos fijos. Este algoritmo es prometedor en términos de eficiencia computacional, pero no siempre obtiene el óptimo global del problema, como ha sido ilustrado numéricamente por Tan et al. (1979) y demostrado teóricamente por Marcotte (1983). En Abdulaal and LeBlanc (1979), se aplica el algoritmo de Hooke-Jeeves al NDP. Este método es computacionalmente intensivo, porque se requieren frecuentes evaluaciones del modelo de asignación en equilibrio.

El algoritmo conocido como EDO (*Equilibrium Decomposed Optimization*), desarrollado por Suwansirikul et al. (1987), optimiza secuencialmente los parámetros de diseño, considerando en cada iteración un arco determinado y tratando el resto de parámetros como constantes. En cada iteración se resuelve un problema de asignación en equilibrio.

En Friesz et al. (1992) y García and Marín (2002), se propone un *algoritmo de recristalización simulada* (SAA) para resolver respectivamente un problema de expansión de la capacidad y un modelo de tarifación y dimensionamiento de la capacidad de los aparcamientos empleados en los viajes de tipo *park'n ride*. Este algoritmo usa un método de búsqueda probabilística. En Meng et al. (2001) se realizan pruebas numéricas para el NDP, con los métodos computacionales antes mencionados, y se demuestra que la mejor solución la alcanza el SAA.

Para el problema DAP, formulado como un MPEC, la mayoría de los algoritmos propuestos tienen una marcada naturaleza heurística. A *grosso modo*, dos son las estrategias seguidas en la elaboración de algoritmos para el DAP. Una primera vía censo (*steepest descent method*). El método de Spiess (1990b), Spiess (1990a), Chen (1994), Florian and Chen (1995) y Codina and Montero (2003) pertenece a esta categoría. Una segunda estrategia consiste en aproximar (linealmente) el conjunto de restricciones de equilibrio. El denominado IEA (Yang et al. (1992)), SAB (Yang (1995)), o EAM (Codina et al. (2001)) son ejemplos de esta vía.

En Codina and Barceló (2000) se ha abordado la naturaleza no diferenciable del problema, habiendo sugerido el uso del método de direcciones conjugadas de Wolfe para la optimización no diferenciable.

Son escasos los métodos descritos en la literatura para resolver el MPEC que tengan ciertas propiedades de convergencia y que, a la vez, puedan ser aplicados a problemas reales de gran escala.

Recientemente, Patriksson and Rockafellar (2002) proporcionaron un modelo matemático para la gestión estratégica del tráfico formulada como un MPEC. Presentaron una reformulación equivalente, a través de un problema de optimización no diferenciable con restricciones lineales, y establecieron la convergencia de una clase de algoritmos para esta nueva formulación. En este trabajo no se han desarrollado pruebas numéricas que muestren la aplicabilidad de estos algoritmos a problemas reales.

Yang et al. (2001) estudian el problema DAP en un contexto de asignación estocástica tipo logit. Estos autores proponen un modelo para la estimación simultánea de una matriz origen-destino y del parámetro del modelo logit. El algoritmo de *programación cuadrática secuencial* (SQP) se aplica al problema y se muestra que converge a un punto de Karush-Kuhn-Tucker del problema bajo ciertas condiciones.

En este capítulo proponemos un algoritmo de generación de columnas (camino) para resolver el PCPCA, que es aplicable a problemas de gran escala, y probamos que este método converge a un mínimo local bajo ciertas condiciones débiles. Este algoritmo se relaciona con los algoritmos desarrollados en Lim (2002) y García and Verastegui (2003) los cuáles se aplican, respectivamente, al NDP y al DAP.

En la sección 2 presentamos la formulación de dos MPEC para modelizar el PCPCA que denominaremos $PCPCA(\rho)$ y $PCPCA(P)$, se introduce la notación e hipótesis empleadas, se describen ambas formulaciones y se realiza un análisis de las soluciones del modelo. En la sección 3 se procede a reformular el $PCPCA(\cdot)$ mediante un problema de optimización de un solo nivel y, posteriormente, se aplica un esquema de generación de columnas (CGA) a esta reformulación, se realiza un análisis de la convergencia de este algoritmo y se concluye que converge a mínimos locales. En la sección 4 se presenta una serie de experimentos numéricos, sobre redes de diferente tamaño, que apoyan y desarrollan los planteamientos teóricos previos. Las pruebas se dividen en dos grupos: un primer grupo en el que se estudian los modelos binivel y un segundo grupo donde se aborda el estudio del comportamiento del algoritmo CGA, testeando su aplicación a problemas reales. Finalmente, en la sección 5 se presenta el resumen y las conclusiones del capítulo.

4.2 Modelos para el PCPCA

En esta sección formulamos dos modelos de programación matemática con restricciones de equilibrio para modelizar el problema de calibración de los parámetros de las funciones de coste en los arcos de una red de tráfico (PCPCA). El primero de estos modelos

emplea como variables de optimización los tiempos/costes de viaje en los arcos en la situación de equilibrio. Esto se realiza así, pues si se empleara como variables de optimización los parámetros originales de las funciones de coste, el resultado sería un modelo sobreespecificado para la información de entrada obtenida en solamente un período, esto es, poseería infinitas soluciones y cada una de ellas daría lugar a predicciones diferentes. Este modelo lo hemos denominado PCPCA(ρ). Una solución del PCPCA(ρ) impone restricciones al valor de los parámetros del coste en los arcos pero no determina sus valores. Es necesario introducir, a priori, unas reglas para fijarlos.

Como alternativa al PCPCA(ρ), hemos propuesto un segundo modelo con el fin de poder estimar directamente los parámetros originales del coste en los arcos. Este modelo, denominado PCPCA(P), requiere mayor cantidad de información que el PCPCA(ρ). Emplea los mismo datos de entrada, pero evaluados en dos instantes de tiempo diferentes. Estas dos mediciones se pueden realizar simultáneamente, lo que correspondería a periodos diferentes del día, o, alternativamente, se podrían obtener en dos fechas diferentes, lo que permitiría considerar el mismo periodo del día.

4.2.1 Notación e hipótesis empleadas

En esta subsección revisamos brevemente el problema de asignación de tráfico con el objeto de definir la notación usada y enunciar las hipótesis sobre el modelo de asignación de tráfico empleado.

El problema de asignación de tráfico (TAP) modela matemáticamente la forma en que los usuarios de una red de tráfico eligen su ruta para satisfacer su viaje (ver Patriksson (1994)). Los principios más empleados para la modelización de la elección de la ruta son: (i) equilibrio determinista bajo el punto de vista del usuario (DUE), también conocido como *primer principio de Wardrop* (Wardrop (1952b)), en donde ningún usuario de la red puede reducir unilateralmente su tiempo de viaje (o coste generalizado) cambiando su ruta, (ii) equilibrio bajo el punto de vista del sistema (SE) o *segundo principio de Wardrop*, en donde el tiempo total de todos los viajes del sistema es mínimo y (iii) equilibrio estocástico bajo el punto de vista del usuario (SUE), en el que ningún usuario puede reducir unilateralmente su tiempo (o coste generalizado) percibido de viaje cambiando su ruta.

Por conveniencia, recogeremos en la tabla 4.1 la notación que emplearemos en la definición del TAP.

El TAP es un problema no lineal de flujos en redes multinodales en el cual se

Definición 4.2.1 (Vector de flujo en equilibrio) *Un vector de flujo en los caminos h^* está en equilibrio si se cumple que existe un vector $U \in \mathbb{R}^{|W|}$ tal que si $h_k^* > 0$ para $k \in K_i$, entonces $C_k(h) = U_i$ y si $h_k^* = 0$ para $k \in K_i$, entonces $C_k(h) \geq U_i$. Esto es, todos los caminos usados para un par origen-destino tienen el mismo coste U_i , y los caminos no utilizados para el par origen-destino i tiene un coste mayor.*

El TAP asume una expresión funcional, $C(h, \rho)$, para los tiempos de viaje en los caminos de la red en función del patrón de flujo h , de un vector de parámetros ρ y del criterio adoptado en la elección de la ruta (DUE, SE o SUE). Aquí no asumiremos ninguna expresión funcional determinada sino que supondremos que $C(h, \rho) : \Omega \times \mathcal{K} \rightarrow \mathbb{R}^N$ satisface que:

- i) es una función continua en el conjunto factible de flujos en los caminos $\Omega = \{h \mid \Lambda h = g, h \geq 0\}$ y en el espacio de parámetros \mathcal{K} .
- ii) es una función aditiva, esto es, que el coste en el camino es la suma del coste en los arcos de los que se compone y por tanto $C(h) = \Delta^T c(v) = \Delta^T c(\Delta h)$, donde T denota la transpuesta de una matriz (o vector).

Notar que las suposiciones i) y ii) permiten considerar un problema asimétrico de asignación de tráfico, en el que puedan ser tenidas en cuenta las interacciones entre los arcos de la red, la existencia de varios modos de transporte o la existencia de múltiples clases de usuario. Este problema puede ser formulado en un marco de complementariedad no lineal, desigualdades variacionales o punto fijo. Inicialmente emplearemos un marco de desigualdades variacionales, en el espacio de flujos en los caminos, para la formulación del PCPCA y, posteriormente, recurriremos a un marco de complementariedad no lineal para la elaboración de un algoritmo de resolución.

El TAP puede ser formulado como un problema de desigualdades variacionales (Smith (1979), Dafermos (1980)) en el espacio de flujo en los caminos. En su forma más general, el problema es: encontrar $h^* \in \Omega$ tal que

$$C(h^*)^T(h - h^*) \geq 0, \quad \forall h \in \Omega \quad [\text{VIP}(C, \Omega)]$$

donde $\Omega = \{h \mid \Lambda h = g, h \geq 0\}$.

El problema de encontrar un flujo en equilibrio en una red de tráfico, cuyos costes en los caminos es $C(h)$ y cuya matriz de viajes origen-destino es g , tiene solución si, y sólo si, el $\text{VIP}(C, \Omega)$ tiene una solución. Ver Smith (1979) y Dafermos (1980).

La formulación $\text{VIP}(C, \Omega)$ corresponde a un TAP bajo el criterio DUE. En el Apéndice se ilustra que esta formulación también recoge los criterios SE y SUE-logit.

4.2.2 PCPCA(ρ)

Supondremos que la función de coste en los arcos puede ser escrita como

$$c_a(v) := t_a^0 + \phi_a(v, P_a)$$

donde t_a^0 es el tiempo de viaje en el arco vacío y ϕ_a es el incremento del tiempo empleado en atravesar el arco debido al efecto de la congestión. Este tiempo depende del nivel

de servicio de la red de tráfico, que es dado por el vector de flujo en los arcos v , y un conjunto de parámetros para el arco $a \in \mathcal{A}$ denotado P_a .

Lo primero que se debe tener en cuenta es que pueden existir infinitos vectores de parámetros que reproducen un patrón de flujos en equilibrio dado. Matemáticamente, todos los conjuntos de parámetros P_a que resuelven el siguiente sistema de ecuaciones

$$c_a^* := t_a^0 + \phi_a(v^*, P_a), \forall a \in \mathcal{A}$$

donde c_a^* y v^* son, respectivamente, el coste en el arco a en el equilibrio y un flujo en equilibrio, reproducen el patrón de flujos en equilibrio v^* . Una implicación de este resultado es que los modelos para el PCPCA, basados únicamente en observaciones de flujo en los arcos, podrían tener infinitas soluciones. Por esta razón, parece más conveniente considerar, como parámetros a ajustar en el PCPCA, el incremento de tiempo por congestión en los arcos $\rho_a = \phi_a(v^*, P_a)$ más que el vector de parámetros P_a , y, después de ajustar estos valores, ya sea introduciendo información adicional o mediante una regla establecida a priori, se ajustarían los parámetros originales.

De la hipótesis ii) (aditividad), la función de coste en los caminos, en función del vector ρ , se puede expresar como:

$$C_k(h, \rho) := \sum_{a \in \hat{A}} \delta_{ak} c_a(v) + \sum_{a \in \bar{A}} \delta_{ak} (t_a^0 + \rho_a), \quad k \in K \quad (4.1)$$

donde \hat{A} denota el subconjunto de arcos que deben ser ajustados y \bar{A} denota el resto de los arcos de la red de tráfico. Los subconjuntos $\{\hat{A}, \bar{A}\}$ constituyen una partición del conjunto de arcos \mathcal{A} .

Si únicamente dispusiésemos como fuente de información para realizar la calibración de ρ los aforos en cierto subconjunto de arcos en la red, una formulación inicial del problema PCPCA, es:

$$\begin{aligned} &\text{Minimizar } Z = F(v, \hat{v}) \\ &\text{sujeto a: } C'(h, \rho)^T (h' - h) \geq 0, \quad \forall h' \in \Omega \\ &\quad v = \Delta h \\ &\quad h \in \Omega, \end{aligned}$$

donde $F(v, \hat{v})$ puede ser cualquier métrica entre el vector de observaciones de flujo en de tiempos de congestión ρ .

Hearn and Ramana (1998) han discutido el problema inverso no lineal de flujos en redes multiproducto en un contexto de tarifación de arcos (peajes). Estos autores probaron que si las observaciones de flujo \hat{v} tienen asociadas un vector de flujos en los caminos no negativos, que satisfacen la demanda de viajes, entonces el conjunto de peajes óptimos es un poliedro. Por otra parte, si el flujo \hat{v} corresponde al flujo en

equilibrio bajo el criterio SE, entonces es posible mostrar que el conjunto de peajes óptimos no está acotado. Larsson and Patriksson (1998) mostraron numéricamente que los peajes óptimos pueden no estar acotados para otros tipos de flujos \hat{v} .

Estos resultados tienen una aplicación directa al PCPCA(ρ), basta con observar que la variable ρ juega el papel de los peajes en el problema de tarifación de arcos. Si consideramos que \hat{v} es un vector de flujo en los arcos en equilibrio bajo el criterio SE, tendríamos que existe un conjunto no acotado de valores de ρ para el cual $v(\rho) = \hat{v}$, donde $v(\rho)$ es un vector de flujo en equilibrio para un vector ρ dado. La función $F(v, \hat{v})$, por ser una métrica, tomaría el valor mínimo de 0 en dicho conjunto. Estos resultados teóricos demuestran que el PCPCA(ρ) puede admitir infinitas soluciones, algunas de ellas ($\|\rho^*\| \rightarrow +\infty$) poco realistas porque introducen tiempos de congestión potencialmente infinitos. Una manera de evitar este tipo de soluciones consiste en añadir, a la formulación de PCPCA(ρ), el siguiente conjunto de restricciones laterales lineales:

$$\sum_{a \in \hat{A}} w_{ia} \rho_a \leq b_i, \quad i \in \mathcal{I} \quad (4.2)$$

donde \mathcal{I} es un conjunto finito de índices y w_{ia} un conjunto de coeficientes. Estas restricciones se utilizan para poner límites a los valores del tiempo de congestión en arcos individuales o en una suma de arcos (por ejemplo los arcos de un camino).

La existencia de infinitas soluciones para el PCPCA(ρ) no se debe a un fallo en su formulación matemática, sino que está originada por una deficiencia en los datos de entrada del modelo. Para corregirla, se requiere una información adicional sobre los costes de viaje en equilibrio. Por ejemplo, el coste observado en la situación de equilibrio, denotado \hat{U} , para un cierto conjunto de pares origen-destino (denotado \hat{I}), o de un valor agregado de él, como el coste medio de viaje en toda la red de tráfico o sobre un subconjunto de pares origen-destino. Esta información se puede utilizar para derivar un nuevo objetivo para el nivel superior y el PCPCA propuesto se convierte

$$\begin{aligned} \text{Minimizar} \quad & Z = \eta_1 F_1(v, \hat{v}) + \eta_2 F_2(U, \hat{U}) \\ \text{sujeto a:} \quad & \rho \in \mathcal{K} \\ & C(h, \rho)^T (h' - h) \geq 0, \quad \forall h' \in \Omega \\ & u \in \Delta^h \end{aligned} \quad [\text{PCPCA}(\rho)]$$

donde F_1 y F_2 representan dos métricas entre el vector de predicciones y el de observaciones, y los pesos $\eta_1, \eta_2 \geq 0$ reflejan la confianza relativa en la información contenida en \hat{v} y \hat{U} . La estimaciones derivadas de este modelo dependen significativamente de los pesos η_1 y η_2 elegidos. La calibración de estos parámetros no siempre resultará una tarea fácil. Los dos parámetros η_1 y η_2 se pueden reducir a uno solo, ya que si

multiplicamos la función objetivo por $\frac{1}{\eta_1 + \eta_2} \geq 0$ las soluciones óptimas del problema no varían y en este caso los coeficientes $\frac{\eta_1}{\eta_1 + \eta_2}$ y $\frac{\eta_2}{\eta_1 + \eta_2}$ corresponden a una combinación convexa. La función objetivo del PCPCA(ρ) se puede escribir equivalentemente como $\eta F_1(v, \hat{v}) + (1 - \eta) F_2(U, \hat{U})$ con $\eta \in [0, 1]$.

La restricción $\rho \in \mathcal{K}$ representa cualquier otra información que dispongamos acerca del valor del vector de parámetros ρ . Esta restricción puede incluir restricciones del tipo (4.2), cualquier otra, o simplemente no existir ($\mathcal{K} = \mathbf{R}^{|\hat{A}|}$). Independientemente de la restricción $\rho \in \mathcal{K}$ se puede demostrar que el conjunto de soluciones óptimas de PCPCA(ρ) es un conjunto acotado bajo la hipótesis de que cualquier arco del conjunto de arcos \hat{A} es empleado por un par observado \hat{I} .

Como ya se ha mencionado anteriormente una solución del PCPCA(ρ) no determina unívocamente el vector de parámetros originales de las funciones de coste en los arcos (P_a , $a \in \hat{A}$) sino que impone el siguiente conjunto de restricciones que debe cumplir:

$$\rho_a^* = \phi_a(v^*, P_a), \forall a \in \hat{A} \quad (4.3)$$

donde ρ_a^* es la solución obtenida con el PCPCA(ρ) y v^* es su vector de equilibrio en los arcos.

4.2.3 PCPCA(P)

En esta subsección asumimos que los vectores \hat{U} y \hat{v} se conocen para varios periodos diferentes, en los cuales las matrices origen-destino son distintas. Esto no quiere decir que necesariamente estos datos hayan sido recogidos en dos fechas diferentes, sino que pueden estar obtenidos en dos periodos diferentes de un mismo día. Denotaremos con el subíndice i cada uno de los dos periodos. Hemos elegido dos periodos por simplicidad, pero podría existir más de dos periodos.

Una forma de determinar el valor de los parámetros originales $(P_a)_{a \in \hat{A}}$ podría consistir en la aplicación del PCPCA(ρ) a cada uno de los periodos y obtener dos conjuntos de parámetros P . Este camino tiene dos inconvenientes claros: i) puede ocurrir que finalmente el sistema obtenidos no tuviera solución y ii) ambos periodos se consideran igualmente relevantes. Un proceso de estimación que evita estos dos inconvenientes, consiste en convertir los dos PCPCA(ρ) en un solo modelo, que lo denotaremos como PCPCA(P), para enfatizar que en este modelo los parámetros de estimación son el vector P . PCPCA(P) se formula:

$$\begin{aligned}
\text{Minimizar } & Z = \sum_{i \in \{1,2\}} \gamma_i \left[\eta_1 F_1(u_i, \hat{u}_i) + \eta_2 F_2(U_i, \hat{U}_i) \right] \\
\text{sujeto a: } & P \in \mathcal{K} \\
& C(h_i, P)^T (h'_i - h_i) \geq 0, \quad \forall h'_i \in \Omega_i, \quad \forall i \in \{1,2\} \\
& u_i = \Delta h_i, \quad \forall i \in \{1,2\} \\
& h_i \in \Omega_i, \quad \forall i \in \{1,2\}.
\end{aligned} \tag{PCPCA(P)}$$

donde los coeficientes γ_i reflejan la confianza relativa en la información usada como entrada en cada periodo.

4.2.4 Análisis de los Modelos

En esta subsección, la discusión se enfoca al establecimiento de una condición suficiente para la existencia de soluciones para los modelos PCPCA(ρ) y PCPCA(P) presentados anteriormente. Matemáticamente ambos modelos son equivalentes. Por un lado el PCPCA(ρ) se puede considerar como un PCPCA(P) con un solo periodo y por otro el PCPCA(P) puede ser visto como un PCPCA(ρ) aplicado a una red que ha sido duplicada. Por esto nos hemos centrado en el estudio de PCPCA(ρ). La existencia de soluciones del PCPCA (ρ) ha sido demostrado empleando argumentos similares a los usados en Chen and Florian (1996).

El siguiente lema puede verse como un análisis de la sensibilidad para la aplicación multievaluada que a cada vector ρ le asocia el conjunto de soluciones del VIP($C(\cdot, \rho), \Omega$). Este resultado solo requiere la continuidad de C . Denotemos por $\Omega^*(\rho)$ el conjunto de flujos en los caminos en equilibrio correspondiente al conjunto de soluciones del VIP($C(\cdot, \rho), \Omega$) para un ρ dado.

LEMMA 4.2.2 (Análisis de la sensibilidad). *Si el mapa de costes $C(h, \rho) : \Omega \times \mathcal{K} \rightarrow \mathbb{R}^N$ es continuo, entonces $\Omega^*(\rho)$ es semicontinuo superiormente en \mathcal{K} .*

DEMOSTRACIÓN.

Previamente observaremos que $C(h, \rho)$ es continuo, entonces VIP($C(\cdot, \rho), \Omega$) tiene al menos una solución para cada ρ dado y por tanto $\Omega^*(\rho)$ es no vacío para todo $\rho \in \mathcal{K}$.

Este lema lo probaremos por contradicción. Supongamos que existe un $\tilde{\rho} \in \mathcal{K}$ en el cual la aplicación multievaluada $\Omega^*(\rho)$ no es semicontinua superiormente. Entonces sucesión de flujos en equilibrio n tal que $n \in \Omega^*(\rho) \subset \Omega$ y $n \rightarrow \tilde{\rho}$, cumpliéndose

$$\tilde{h} \notin \Omega^*(\tilde{\rho}).$$

Esto es, que existe $h' \in \Omega$ tal que

$$C(\tilde{h}, \tilde{\rho})^T (h' - \tilde{h}) < 0 \tag{4.4}$$

Como $h^j \in \Omega^*(\rho^j)$, se tiene que

$$C(h^j, \rho^j)^T (h' - h^j) \geq 0, \quad \forall j,$$

Tomando límites en ambos lados y usando la continuidad de $C(h, \rho)$, se tiene que

$$C(\tilde{h}, \tilde{\rho})^T (h' - \tilde{h}) \geq 0,$$

que contradice a (4.4). Así, esto prueba que $\Omega^*(\rho)$ es semicontinuo superiormente en su dominio. \square

TEOREMA 4.2.3 (Existencia de soluciones). *Si las métricas F_i para todo i y $C(h, \rho) : \Omega \times \mathcal{K} \rightarrow \mathbb{R}^N$ son funciones continuas en sus dominios y \mathcal{K} es un conjunto compacto, entonces PCPCA(ρ) tiene al menos una solución.*

DEMOSTRACIÓN. Mediante el lema 4.2.2 y el resultado obtenido en Hogan (1973), la gráfica de $\Omega^*(\rho)$ es cerrada, lo cual implica que la región factible del PCPCA(ρ) es cerrada en el conjunto compacto $\mathcal{K} \times \Omega$. Por tanto PCPCA(ρ) es la minimización de una función continua en un conjunto compacto, y tiene al menos una solución. \square

NOTA 4.2.4 (UNA OBSERVACIÓN ACERCA DE LAS SOLUCIONES DEL PCPCA(ρ)) El propósito de esta nota es mostrar que existen elementos del conjunto $\Omega^*(\rho^*) \times \{\rho^*\}$ que pudieran no ser óptimos para PCPCA(ρ). Esto indica que su conjunto de soluciones óptimas se deben expresar tanto en función de ρ como de h . Esta afirmación la ilustraremos con el siguiente ejemplo.

Considérese una red de tráfico con dos arcos $\mathcal{A} = \{a, b\}$ y un par origen-destino. Cada arco de la red representa dos caminos disjuntos para satisfacer el único par. En este ejemplo, las formulaciones en el espacio de flujo en los arcos y en el espacio de flujo en los caminos son idénticas, y las variables v y h son intercambiables. Las verdaderas funciones de coste en los arcos son $c_a(v_a) := 1 + 2v_a$ y $c_b(v_b) = 2 + v_b$ y la demanda es $g = 5$. El flujo en equilibrio se obtiene para los valores de $v_a^* = 2, v_b^* = 3$ y el coste de equilibrio es $U^* = 5$. Si definimos el PCPCA(ρ) para estos datos de entrada, esto es, $\hat{v} = v^*$ y $\hat{U} = U^*$, entonces el vector de parámetros óptimos ρ^* es $\rho_a^* = 4$ y $\rho_b^* = 3$.

En este ejemplo, $\Omega^*(\rho^*) = \Omega$, y $F_2(U, \hat{U}) = 0$ para todo $(h, \rho^*) \in \Omega \times \{\rho^*\}$, pero $F_1(v, \hat{v}) = 0$ se satisface únicamente en $v = v^*$, por tanto la solución óptima es (v^*, ρ^*) . usando solamente atoros en los arcos, el problema tiene infinitas soluciones, que serian $\rho^* = (4 + \delta, 3 + \delta)$. Otra característica de este supuesto es que la región factible para el PCPCA(ρ) es no convexa.

Una situación en la que el conjunto $\Omega^*(\rho^*) \times \{\rho^*\}$ es un conjunto de soluciones óptimas del PCPCA(ρ), es cuando el equilibrio en término de flujos en los arcos es único para ρ^* . El conjunto de flujos en los caminos en equilibrio puede no ser único, pero

ellos tienen que satisfacer la relación $v^* = \Delta h$, y entonces la métrica F_1 es constante en $\Omega^*(\rho^*)$. Por otro lado, la igualdad $C(h) = \Delta^T c(\Delta h) = \Delta^T c(v^*)$ muestra que $C(h)$ es también constante en $\Omega^*(\rho^*)$ y, por tanto, el coste de equilibrio U es constante en $\Omega^*(\rho^*)$. Además, como F_2 depende de U , se tiene que esta función es constante en $\Omega^*(\rho^*)$. Esto muestra que la función objetivo es constante en $\Omega^*(\rho^*)$, lo que implica que todas las soluciones son óptimas.

4.3 Un algoritmo de generación de columnas para el PCPCA

En esta sección supondremos que $C(h, \rho) : \Omega \times \mathcal{K} \rightarrow \mathbb{R}^N$ es continua en su dominio, y que \mathcal{K} es un conjunto compacto. Estas suposiciones se realizan para asegurar la existencia de soluciones para los PCPCAs (Teorema 4.2.3).

4.3.1 El algoritmo CGA

En esta subsección reformulamos el PCPCA(\cdot) mediante un problema de optimización de un solo nivel y, posteriormente, aplicamos un esquema de generación de columnas a esta reformulación. Para mayor claridad, y sin perder generalidad, nos concentraremos en la solución del PCPCA(ρ). El PCPCA(P) puede ser resuelto mediante la adaptación del algoritmo propuesto y el análisis de su convergencia se puede derivar empleando los mismos argumentos.

Para establecer un algoritmo de resolución, emplearemos la formulación del TAP como un problema de complementariedad en lugar de una desigualdad variacional. El TAP se puede reescribir como el siguiente sistema de igualdades y desigualdades.

$$C(h, \rho) = \Lambda^T U + \pi \tag{4.5}$$

$$\Lambda h = g \tag{4.6}$$

$$\pi^T h = 0 \tag{4.7}$$

$$\pi, h \geq 0. \tag{4.8}$$

mas un coste no negativo. La ecuación (4.6) es la satisfacción de la demanda. La ecuación (4.7) es la condición de holgura complementaria, e indica que si el coste extra es positivo para un camino k (esto es, $\pi_k > 0$) entonces el flujo en el camino es cero. La restricción (4.8) impone la no negatividad de las variables.

Con la anterior reformulación del TAP, el PCPCA(ρ) se puede reescribir como el siguiente problema de optimización de un solo nivel

$$\begin{aligned}
 \text{Minimizar } & Z = \eta_1 F_1(v, \hat{v}) + \eta_2 F_2(U, \hat{U}) \\
 \text{sujeto a: } & \rho \in \mathcal{K} \\
 & \Delta h = v \\
 & C(h, \rho) = \Lambda^T U + \pi \\
 & \Lambda h = g \quad [\text{PCPCA}(\rho)] \\
 & \pi^T h = 0 \\
 & \pi, h \geq 0.
 \end{aligned}$$

Esta formulación es muy interesante pero tiene una desventaja seria. Requiere la generación de todos los caminos. Esta tarea no es trivial, ni siquiera en redes pequeñas, y es prohibitiva para grandes redes, pues el número de caminos crece exponencialmente con el tamaño de la red. Esta desventaja causa que ningún algoritmo pueda ser aplicado directamente a la anterior formulación del PCPCA. En esta sección proponemos un algoritmo basado en la generación de columnas (caminos), el cual obtiene los caminos según van siendo necesitados.

Para poder describir el algoritmo propuesto son necesarias las siguientes definiciones.

Definition 4.3.1 (*Flujo restringido a un conjunto de caminos*) Definimos el espacio de flujo restringido en el conjunto de caminos \hat{K} por el conjunto

$$\Omega(\hat{K}) := \left\{ h \mid \sum_{k \in \hat{K}_i} h_k = g_i, \forall i \in I; h_k \geq 0, \forall k \in \hat{K} \text{ y } h_k = 0, \forall k \in K - \hat{K} \right\}$$

Definition 4.3.2 (*Flujos en equilibrio restringido a un conjunto de caminos*) Definiremos el conjunto de flujos en equilibrio restringido al conjunto de caminos \hat{K} por

$$\Omega^*(\hat{K}, \rho) := \left\{ h \in \Omega(\hat{K}) \mid C_k(h, \rho) = U_i + \pi_k, \forall k \in \hat{K}_i, \forall i \in I; \sum_{k \in \hat{K}} \pi_k h_k = 0; \pi_k \geq 0, \forall k \in \hat{K} \right\}.$$

Obsérvese que si el conjunto \hat{K}_i es no vacío para todo $i \in I$, entonces $\Omega^*(\hat{K}, \rho) \neq \{\emptyset\}$ porque $C(h, \rho)$ es una función continua y $\Omega(\hat{K})$ es un conjunto convexo.

Definition 4.3.3 (*PCPCA restringido*) Denominaremos PCPCA restringido a \hat{K} a:

$$\begin{aligned}
 \text{Minimizar } & Z = \eta_1 F_1(v, \hat{v}) + \eta_2 F_2(U, \hat{U}) \\
 \text{sujeto a: } & \rho \in \mathcal{K} \\
 & h \in \Omega^*(\hat{K}, \rho) \quad [\text{RPCPCA}(\hat{K})]
 \end{aligned}$$

El algoritmo propuesto para resolver el PCPCA es un algoritmo generador de columnas, denominado CGA, que resuelve en cada iteración principal dos problemas de optimización. En una primera fase, encuentra una solución del TAP para el vector actual de parámetros ρ y obtiene el conjunto de caminos en equilibrio para dicho vector. Estos caminos se añaden a los generados en las iteraciones previas y posteriormente se resuelve el PCPCA restringido a este conjunto de caminos. La solución (aproximada) de este problema proporciona un nuevo vector de parámetros ρ con el que iterar el proceso. El algoritmo CGA se resume en la tabla 4.2. En el paso 2, el CGA chequea si el TAP ha conseguido ampliar el conjunto de caminos actuales, si esto no es así, el vector ρ actual es un mínimo local del PCPCA y el CGA finaliza, en otro caso plantea un nuevo PCPCA restringido. Obsérvese que no se ha asumido ninguna hipótesis que nos asegure que los TAP resueltos sean no degenerados y, por tanto, puede ocurrir que debamos almacenar caminos mínimos sin flujo en el equilibrio.

4.3.2 Análisis de la convergencia del CGA

La idea clave, usada en la demostración de la convergencia del CGA, se basa en el hecho que si la aplicación $C(h, \rho)$ es perturbada por un ε suficientemente pequeño, entonces las soluciones de $\text{VIP}(C(h, \rho + \varepsilon), \Omega(K(\rho)))$, donde $K(\rho)$ es el conjunto de caminos de mínimo coste en el equilibrio derivados de la aplicación $C(\cdot, \rho)$, también resuelven el problema original $\text{VIP}(C(h, \rho + \varepsilon), \Omega)$. Esto nos muestra que existe un entorno de ρ tal que si restringiéramos el RPCPCA(ρ) y el PCPCA(ρ) a dicho entorno, estos problemas serían equivalentes.

TEOREMA 4.3.4 *Sea $\rho' \in \mathcal{K}$. Supongamos que existe un entorno de ρ' , denotado por $\tilde{\mathcal{B}}$, tal que para todo $\rho \in \tilde{\mathcal{B}}$ verifica las dos siguientes propiedades:*

- i) La aplicación $C(h, \rho)$ es constante con respecto de h en $\Omega^*(\rho)$,*
- ii) La aplicación $C(h, \rho)$ es constante respecto de h en $\Omega^*(K(\rho'), \rho)$.*

Entonces existe un entorno abierto $\tilde{\mathcal{B}}$ de ρ' tal que

$$\Omega^*(\rho) = \Omega^*(K(\rho'), \rho), \quad \forall \rho \in \tilde{\mathcal{B}}$$

donde $K(\rho')$ es el conjunto de caminos de coste mínimo en el equilibrio derivado de ρ' .

DEMOSTRACIÓN. Primero, probaremos la inclusión

Sea $h_{\rho'}^* \in \Omega^*(\rho')$. Por la hipótesis i) del teorema, sea cual sea la elección de $h_{\rho'}^*$, se tiene que existe solamente un vector de costes de equilibrio, el cual se denota por $U(\rho')$, y que satisface $C(h_{\rho'}^*, \rho') = U(\rho')$.

El conjunto $K(\rho')$ contiene todos los caminos de coste mínimo para todo $i \in I$, incluso alguno de los caminos pudiera no tener flujo en el equilibrio. Por eso, y por la

Por contradicción. Para todo entorno de ρ' contenido en B , existe un elemento ρ tal que $h_\rho^* \notin \Omega^*(K(\rho'), \rho)$ para algún $h_\rho^* \in \Omega^*(\rho)$. Entonces, podemos construir dos sucesiones $\{\rho^j\}$ y $\{h^j\}$ que satisfacen: a) $\{\rho^j\} \rightarrow \rho'$ y b) $h^j \in \Omega^*(\rho^j) - \Omega^*(K(\rho'), \rho^j)$.

Como $\{h^j\} \in \Omega$ y el conjunto factible de flujos en los caminos es compacto, existe una subsucesión convergente $\{h^j\}_{j \in J}$. El límite de esta subsucesión lo denotamos por \bar{h} .

El mapa solución $\Omega^*(\rho)$ es semicontinuo superiormente en ρ' por el lema 4.2.2, y se tiene que

$$\text{Si } \begin{array}{l} h^j \in \Omega^*(\rho^j), \quad \forall j \in J \\ \{\rho^j\}_{j \in J} \rightarrow \rho' \\ \{h^j\}_{j \in J} \rightarrow \bar{h} \end{array} \quad \Rightarrow \quad \bar{h} \in \Omega^*(\rho')$$

Porque $(h^j, \rho^j)_{j \in J} \rightarrow (\bar{h}, \rho') \in V \times B$, donde $V \times B$ es un conjunto abierto, existe un número entero j_0 tal que $(h^j, \rho^j) \in V \times B$ para todo $j > j_0$.

Usando la relación (4.11) para (h^j, ρ^j) con $j > j_0$, obtenemos que el conjunto de caminos mínimos es un subconjunto de $K(\rho')$. Por otra parte, $h^j \in \Omega^*(\rho^j)$ (es un vector de flujos en los caminos en equilibrio) y ello implica que $h_{k'}^j = 0$ para todo $k' \in K_i - K_i(\rho') \subset K_i - K_i(\rho^j)$ y para todo $i \in I$ y ello muestra que $h^j \in \Omega^*(K(\rho'), \rho^j)$ para todo $j > j_0$, y esto es una contradicción.

Ahora probaremos la inclusión contraria. Sea $\bar{h} \in \Omega^*(K(\rho'), \rho)$ donde $\rho \in \bar{B} \cap \mathcal{B}$. Por lo probado anteriormente se tiene que $\Omega^*(\rho) \subset \Omega^*(K(\rho'), \rho)$ y por la hipótesis *ii*) $C(h, \rho)$ es constante en $\Omega^*(K(\rho'), \rho)$, entonces existe un flujo de camino en equilibrio $h^* \in \Omega^*(\rho)$, que cumple

$$C(\bar{h}, \rho) = C(h^*, \rho). \quad (4.12)$$

Sea $h \in \Omega$, probaremos que $C(\bar{h}, \rho)^T(h - \bar{h}) \geq 0$. De la optimalidad de h^* se tiene que

$$C(h^*, \rho)^T(h - h^*) \geq 0. \quad (4.13)$$

Por la optimalidad de \bar{h} en el espacio de flujos restringidos $\Omega(K(\rho'))$ y como $h^* \in \Omega(K(\rho'))$, entonces se verifica

$$\text{Empleando (4.12), (4.13) y (4.14), se obtiene } C(\bar{h}, \rho)^T(h^* - \bar{h}) > 0. \quad (4.14)$$

$$\begin{aligned} C(\bar{h}, \rho)^T(h - \bar{h}) &= C(\bar{h}, \rho)^T[(h - h^*) + (h^* - \bar{h})] \\ &= C(h^*, \rho)^T(h - h^*) + C(\bar{h}, \rho)^T(h^* - \bar{h}) \geq 0. \end{aligned}$$

□

COROLARIO 4.3.5 (CONDICIÓN SUFICIENTE PARA UN OPTIMO LOCAL DEL PCPCA(ρ')) *Supongamos que las condiciones del teorema 4.3.4 son satisfechas. Supongamos además que $(h_{\rho'}^*, \rho')$ es un mínimo local para el RPCPCA(\hat{K}) y $K(\rho') \subset \hat{K}$, entonces $(h_{\rho'}^*, \rho')$ es un mínimo local para el PCPCA(ρ).*

DEMOSTRACIÓN. Del hecho que $\Omega(K(\rho')) \subset \Omega(\hat{K}) \subset \Omega$ y el teorema 4.3.4, existe un entorno \tilde{B} de ρ' tal que

$$\Omega^*(K(\rho'), \rho) = \Omega^*(\hat{K}, \rho) = \Omega^*(\rho), \quad \forall \rho \in \tilde{B}.$$

Por la suposición de que $(h_{\rho'}^*, \rho')$ sea un óptimo local de

$$\begin{aligned} \text{Minimizar } & Z = \eta_1 F_1(v, \hat{v}) + \eta_2 F_2(U, \hat{U}) \\ \text{sujeto a: } & \rho \in \mathcal{K} \cap \tilde{B} \\ & h \in \Omega^*(\hat{K}, \rho). \end{aligned}$$

se deriva el resultado del corolario. □

El CGA genera nuevos caminos en cada iteración, y como el número de esto caminos en la red es finito, el CGA tiene que terminar en un número finito de iteraciones. El siguiente teorema muestra una condición para establecer que la última iteración es un mínimo local para el DAP.

TEOREMA 4.3.6 (CONVERGENCIA DEL CGA) *Supongamos que ρ' es la última iteración generada por el CGA. Supongamos que ρ' satisface las condiciones del teorema 4.3.4, entonces ρ' es un mínimo local para el PCPCA(ρ).*

DEMOSTRACIÓN. El CGA genera una sucesión de conjuntos de caminos $\{K^\ell\}$ que satisfacen $K^\ell \subset K^{\ell+1}$. Por la finitud de caminos de la red, el CGA tiene que terminar en un número finito de iteraciones. La última iteración ρ' tiene que satisfacer el criterio de parada

$$K(\rho') \subset K^{s-1} \text{ para un cierto número entero } s.$$

Más aún, ρ' es un mínimo local para el RPCPCA(K^{s-1}) y usando el corolario 4.3.5 para $\hat{K} = K^{s-1}$, se deduce el teorema. □

4.4 Experimentos numéricos

En esta sección completamos el estudio teórico realizado anteriormente con un conjunto de pruebas numéricas. Las pruebas las hemos dividido en dos experimentos: en el primer experimento se estudia los modelos binivel propuestos y el segundo se aborda el estudio del algoritmo CGA, testeando su aplicación a problemas reales.

4.4.1 Experimento 1: Estudio de los modelos PCPCA

Los objetivos generales de este primer experimento son: i) motivar la necesidad de desarrollar herramientas como las aquí expuestas para la calibración de las funciones de coste en los arcos y ii) comparar la capacidad predictiva del modelo PCPCA(ρ) frente al PCPCA(P).

El primer objetivo lo abordaremos mostrando que dos parametrizaciones que representen la situación de equilibrio actual pueden conducir a predicciones significativamente diferentes en un escenario futuro.

En el estudio teórico se muestra que el PCPCA(ρ) no puede determinar unívocamente la parametrización original P y este hecho motiva la elección de la variable tiempo de congestión en los arcos ρ como parámetro a optimizar, requiriendo un criterio adicional para poder estimar el vector P . La segunda cuestión que nos planteamos con el experimento 1 es que si conociésemos el "verdadero" criterio adicional las predicciones derivadas de los modelos PCPCA(ρ) y PCPCA(P) son similares, o por contra, si el hecho que el PCPCA(P) emplee información en dos periodos mejora significativamente su fiabilidad. Otro aspecto importante que se va a evaluar es el uso del tiempo medio de viaje en la red, en lugar de una muestra de tiempos en un conjunto de pares en el equilibrio, en relación a la calidad de las estimaciones.

Todos estos aspectos dependen sustancialmente del comportamiento real de los usuarios en la red y del modelo elegido para modelarlo. Para controlar este hecho hemos simulado varios comportamientos de usuarios a través de varios modelos de asignación en equilibrio y hemos considerado que deseamos calibrar un TAP-DUE.

Finalizaremos estudiando la existencia de óptimos locales en los modelos PCPCA(ρ), ya que este hecho podría conducir a que el algoritmo CGA no encontrase el valor óptimo de ρ .

En el gráfico 4.2 se describe este primer experimento, que ha sido realizado sobre la red descrita en Nguyen and Dupuis (1984), en adelante ND, cuya topología se muestra en la figura 4.1. Se trata de una red que tiene 13 nodos, 19 arcos y cuatro pares O-D.

Las funciones de coste en los arcos son funciones lineales, y en este experimento nos planteamos calibrar los parámetros de las funciones BPR:

$$\int_0^{u_a} c_a(x) dx = t_a^0 u_a + d_a u_a^{m_a}$$

para los arcos a . El parámetro t_a^0 se refiere al tiempo de viaje libre y d_a y m_a son parámetros de congestión.

El experimento 1 (ver el gráfico 4.2) considera dos periodos actuales definidos por la matriz O-D original g , empleada en la parametrización de ND, y una segunda matriz O-D definida por $1.5g$. Se considera un escenario futuro definido por la matriz O-D $2g$.

Hemos simulado cinco comportamientos de los usuarios definidos por: (i) el principio del óptimo para el usuario en un contexto determinista (TAP-DUE), (ii) el principio de

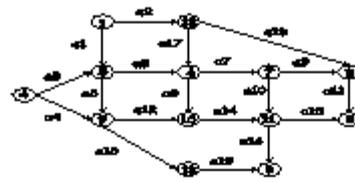
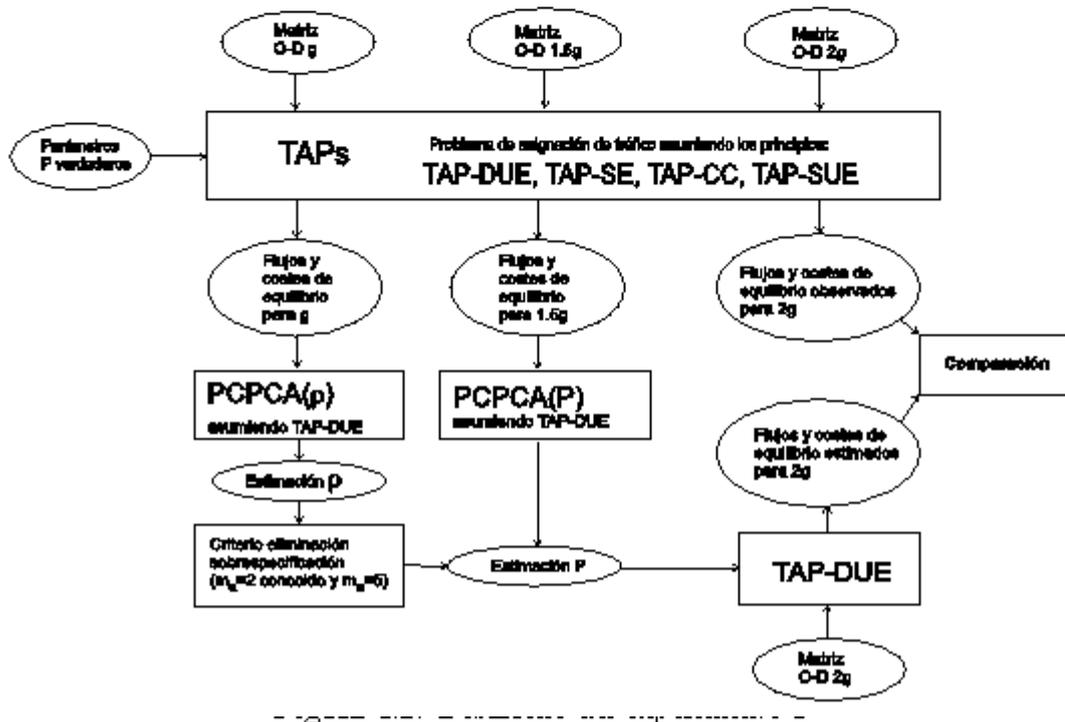


Figura 4.1: Topología de la red de Nguyen-Dupuis



óptimo para el sistema (TAP-SE), (iii) un modelo TAP con restricciones de capacidad (TAP-CC) en los arcos a_{14} , con capacidad máxima de 400, y a_{18} , con capacidad máxima de 300, (iv) el principio de óptimo para el usuario en contexto aleatorio tipo logit con $\frac{1}{\theta} = 25$ (TAP-SUE($\frac{1}{\theta} = 25$)) y (v) idem que (iv) pero para el valor del parámetro

$\frac{1}{\theta} = 50$ (TAP-SUE($\frac{1}{\theta} = 50$)).

Para los dos escenarios presentes y para todos los tipos de usuarios considerados, resolviendo los correspondientes problemas de asignación de tráfico, se han calculado los flujos en equilibrio, los costes/tiempos medios en los pares O-D y el coste medio en la red para cada una de las situaciones. Todos estos datos han sido utilizados como datos de partida de los modelos PCPCA(\cdot). Los parámetros de "peso" en la función objetivo (η) se han fijado mediante tanteos, eligiendo los que se consideraban más idóneos para cada situación.

La resolución del modelo PCPCA(ρ) proporciona una estimación del coste/tiempo de congestión para los arcos a_2 y a_{14} . Se ha utilizado como información de entrada el flujo en todos los arcos y considerándose dos informaciones adicionales diferentes: el tiempo medio en los pares origen-destino en equilibrio y el tiempo medio en la red.

Para el PCPCA(ρ) el valor del vector de parámetros $K_a = (d_a, m_a)$ se ha obtenido de la relación

$$\rho_a^* = m_a d_a (v_a^*)^{m_a - 1}$$

y asumiendo que m_a es conocido, donde ρ_a^* es el tiempo de congestión estimado por PCPCA(ρ). Se ha considerado dos supuestos: (i) $m_a = 2$ (valor verdadero) y (ii) $m_a = 5$ (valor usual en muchos estudios prácticos).

Como se ha indicado anteriormente, el PCPCA(P) estima directamente los parámetros originales P . Para ello, emplea los mismos datos de entrada que el PCPCA(ρ), pero evaluados en dos instantes de tiempo diferentes definidos por las matrices g y $1.5g$.

Finalmente con la estimación del vector P (obtenidas usando tanto el PCPCA(ρ) o PCPCA(P)) y asumiendo un modelo TAP-DUE se ha efectuado las predicciones para el escenario futuro definido por la matriz O-D $2g$. Estos resultados se han comparado con los verdaderos valores obtenidos al emplear el modelo de asignación que describe el comportamiento del usuario y los verdaderos parámetros. Se han observado tres items: i) la máxima diferencia relativa de flujo en los arcos (en porcentaje), ii) máxima diferencia de tiempo en equilibrio y iii) diferencia relativa en el tiempo medio en toda la red.

Los resultados obtenidos con el modelo PCPCA(ρ) se muestran en las tablas 4.3 y 4.4, y para el modelo PCPCA(P) en la tabla 4.5.

La primera observación es que en el escenario TAP-DUE (el comportamiento de los usuarios coincide con el modelo elegido) el PCPCA(P) es capaz de estimar la como una función de éstos (tiempo medio en la red). Este hecho también ocurre para el PCPCA(ρ) con el criterio correcto. También se observa que el PCPCA(P) y el PCPCA(ρ) para el criterio correcto $m_a = 2$ son dos modelos comparables.

Las parametrizaciones obtenidas por el PCPCA(ρ) con $m_a = 2$ y $m_a = 5$ tienen la misma capacidad para reproducir las observaciones en el primer periodo (definido por g) pero el error de las predicciones derivadas para el escenario futuro (ver las

tablas 4.3 y 4.4) es aproximadamente el doble. Este hecho muestra la importancia de una adecuada parametrización en las predicciones de los costes de equilibrio, teniendo presente que únicamente son dos de los 19 arcos de la red los que introducen errores.

Se observa que la relación entre comportamiento real del usuario y modelo de asignación en equilibrio para modelarlo tiene un efecto muy significativo en los errores.

De los resultados se puede deducir que, en general, se obtienen mejores resultados cuando se utiliza como información adicional el coste medio en los pares O-D que el coste medio en la red.

Los resultados del modelo, para los parámetros de las funciones de coste en los arcos a_2 y a_{14} se pueden observar en las tablas 4.6 y 4.7 donde se presentan todos los casos considerados. Se observa que reproducen los verdaderos parámetros perturbados por el efecto del error cometido al modelar el comportamiento del usuario.

Finalmente, se ha investigado en la naturaleza de las soluciones obtenidas, para ello se ha calculado numéricamente las curvas de nivel de la función objetivo para el PCPCA (ρ) para el coste medio en los pares O-D. Los resultados se muestran en las figuras 4.3, 4.4, 4.5, 4.6 y 4.7. Se observa que algunos de estos problemas son no convexos ya que poseen curvas de nivel que no encierran a regiones convexas, y podrían existir (como parece en los TAP-SUE) mínimos locales que no son globales.

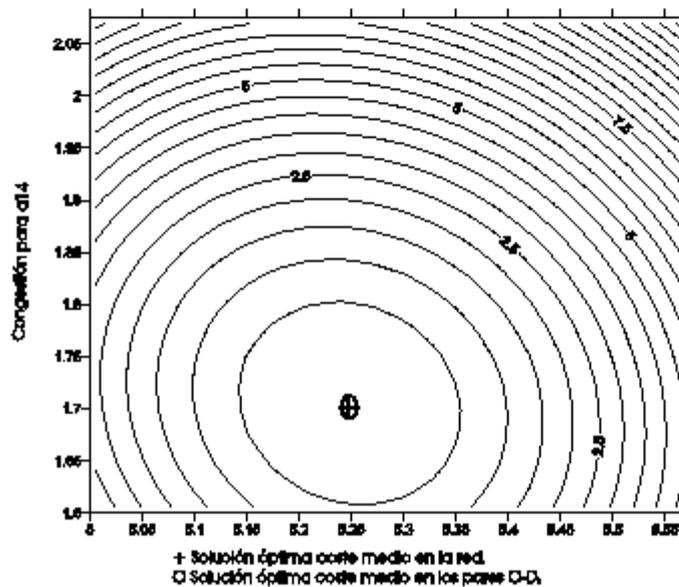
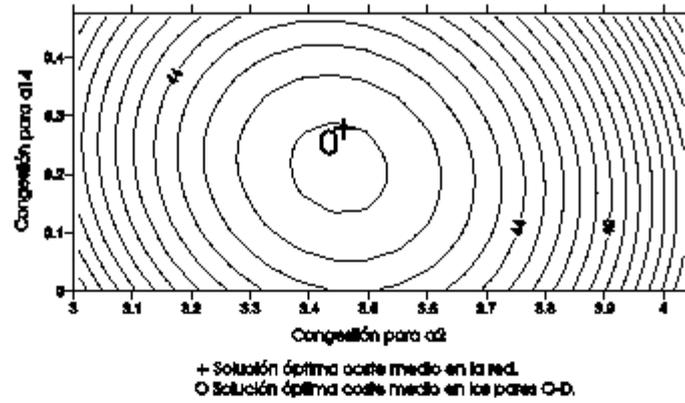
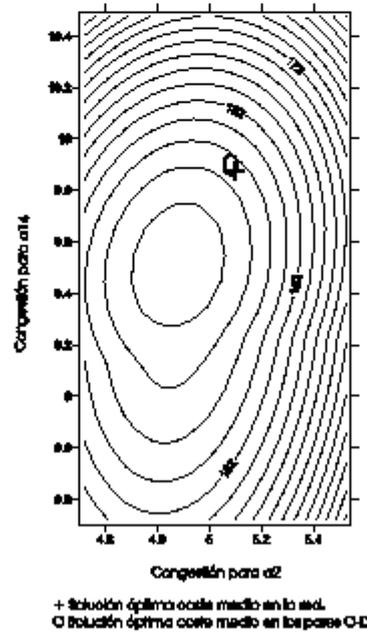


Figura 4.3: PCPCA(ρ) curvas nivel TAP-DUE

Figura 4.4: PCPCA(ρ) curvas nivel TAP-SEFigura 4.5: PCPCA(ρ) curvas nivel TAP-CC

4.4.2 Experimento 2: estudio del algoritmo CGA

El tipo de información empleado por el PCPCA(ρ) se ajusta en muchas ocasiones al disponible en muchas aplicaciones. Las conclusiones derivadas en el experimento anterior resultaron similares a los obtenidos al PCPCA(F). Estas consideraciones hacen que en este experimento empleemos el modelo PCPCA(ρ). El objetivo del experimento 2 es evaluar la aplicabilidad del algoritmo propuesto, evaluando su comportamiento, coste computacional y las posibilidades que tiene para ser aplicado a problemas reales.

Otra conclusión del experimento 1 es la posibilidad de presentar el PCPCA(ρ) mínimos locales no globales. Este hecho provocaría que el algoritmo CGA obtuviese

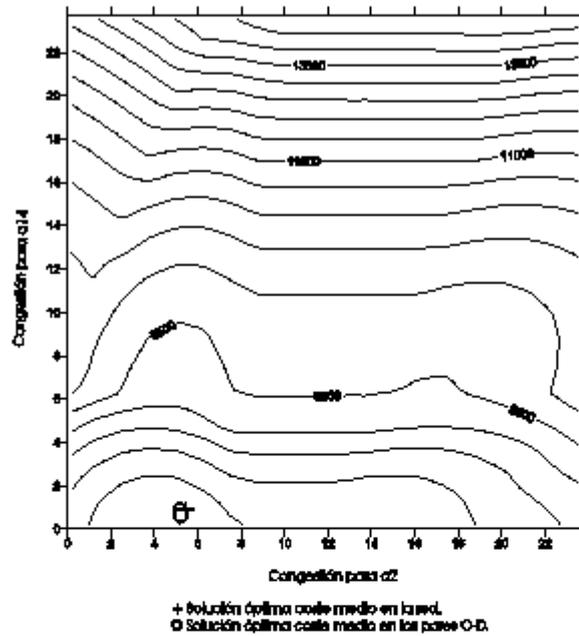


Figura 4.6: PCPCA(ρ) curvas nivel TAP-SUE $\frac{1}{\theta} = 25$

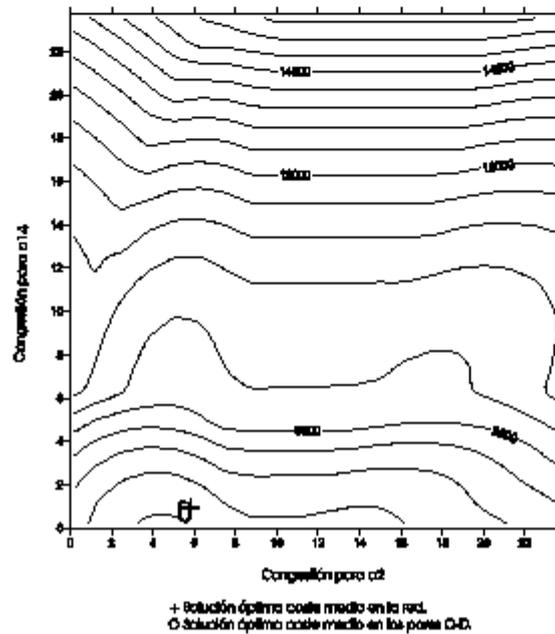


Figura 4.7: PCPCA(ρ) curvas nivel TAP-SUE $\frac{1}{\theta} = 50$

costes de congestión no reales. Esta cuestión se estudiará en este experimento. Para evitar respuestas condicionadas al tipo de información disponible, se asumirá el escenario más favorable, en el que se dispone de la máxima información posible. Distinguiremos dos casos: en el primero se asume que disponemos de los aforos en todos los arcos de la red cuyo flujo es positivo, y en el segundo consideramos que además de esta información

disponemos del coste en equilibrio para todos los pares de la red.

En el experimento 2 se ha considerado la red Sioux Falls (LeBlanc et al. (1975)) que se muestra en la figura 4.8 y que consta de 528 pares origen-destino, 76 arcos y 24 nodos y la de Hull (Florian (1990)) con 142 pares, 798 arcos y 501 nodos.

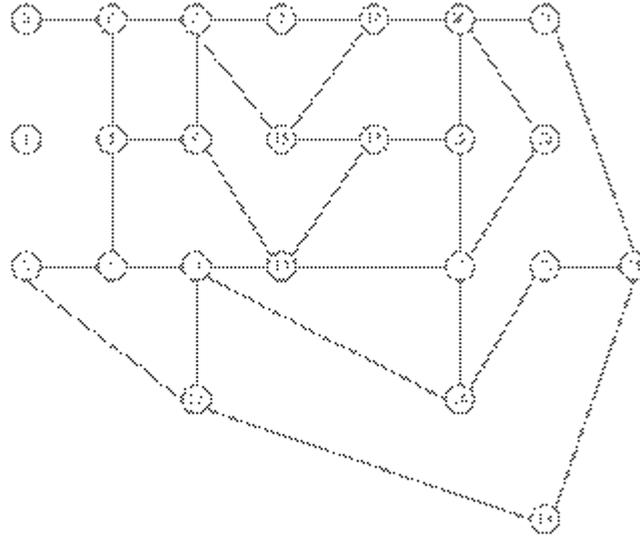


Figura 4.8: Red de Sioux Falls

La dificultad de calcular el equilibrio en el primer caso deriva (posiblemente) de que el flujo en 59 de los 76 arcos excede la capacidad del arco, considerada ésta como el parámetro de capacidad empleada en las funciones BPR utilizada. Sin embargo, la segunda red tiene un menor nivel de congestión.

El algoritmo CGA ha sido codificado en GAMS. A su vez este algoritmo emplea el algoritmo de descomposición simplicial desagregada que también ha sido codificado en GAMS. El DSD itera entre una fase de generación de columnas (resolución de un conjunto de problemas de caminos mínimos) con el llamado problema maestro restringido (RMP). Se ha implementado el método de D'Esopo (Pape (1974)) para calcular los caminos mínimos y se ha utilizado como solver del RMP y del RPCPCA(ρ) MINOS. Se han resuelto los cuatro problemas anteriores, empleando como criterio de paro $|Z^{\ell+1} - Z^\ell| < 0.01$. Se ha tomado, como punto de inicialización del algoritmo CGA, $\rho_a = 30, a \in \hat{A}$ para los problemas 1 y 2, tomándose $\rho_a = 10, a \in \hat{A}$ para Fujitsu Siemens Computers mobile con 128 Mb de RAM y Pentium IV (Celeron).

El algoritmo CGA descrito en la tabla 4.2 almacena todos los caminos generados en las iteraciones anteriores en el PCPCA, con el objeto de tener garantizada su convergencia en un número finito de iteraciones. Esta regla puede conducir a un gran número de caminos (variables) y en consecuencia a un gran coste computacional. En la experiencia numérica hemos introducido un nuevo criterio de almacenamiento de

columnas. En cada iteración el conjunto de caminos empleados en el PCPCA es el conjunto de caminos con flujo positivo en la solución del último problema de asignación de tráfico resuelto. Esto es, el paso 1 del citado algoritmo sería ahora: 1. (*Problema de asignación de tráfico*). Sea $C(h, \rho)$ la función de costes en los caminos, entonces resuélvase $VIP(C(\cdot, \rho), \Omega)$. Sea $K(\rho)$ el conjunto de caminos de mínimo coste con flujo positivo en el equilibrio para el vector de parámetros ρ . Esta elección podría conducir a que el algoritmo perdiese la convergencia, debido a algún proceso de ciclaje, pero en el caso de converger, y bajo la condición de que el punto límite satisface la condición de holgura complementaria estricta, este punto es un mínimo local.

Los inputs para los PCPCA(ρ) se obtienen resolviendo mediante el DSD (Larsson and Patriksson (1992)) la red Sioux Falls con un error relativo menor que 10^{-14} y la de Hull con un error relativo menor que 10^{-10} . No se ha introducido ninguna perturbación, ni en los aforos, ni en los tiempos/costes de equilibrio. Esto hace que el valor óptimo para los PCPCA(ρ) planteados sea $Z^* = 0$, y por tanto podemos saber si lo obtenido corresponde a un óptimo local o global.

Se ha elegido aleatoriamente los arcos en los que deseamos estimar el tiempo de congestión en la red de Sioux Falls, mientras que para la red de Hull se ha elegido los siete arcos más congestionados.

En la tabla 4.8 se definen los cuatro modelos considerados. En esta tabla se muestran las funciones objetivo empleadas en cada problema, la red de tráfico empleada y los verdaderos valores de congestión en los arcos (para la parametrización original de las redes). Esto es, se muestra el valor de ρ^* .

La tabla 4.9 muestra las soluciones obtenidas en cada modelo.

La primera observación es que los modelos basados únicamente en mediciones de aforo no obtienen los verdaderos valores de la congestión. Esto es una característica del modelo PCPCA(ρ) y no del algoritmo empleado. Para clarificar esta afirmación considérese la red dada en la figura 4.9 y supóngase que existe una única demanda entre los nodos a y c . Supóngase además que se quiere estimar el coste de congestión en el arco (a, b) y que las parametrización de los dos arcos que unen los nodos b y c son conocidas. El flujo en el arco (a, b) está *cautivo* y cualquier valor de la congestión en el arco (a, b) no afectaría a los aforos observados, lo que implica que a partir únicamente de los aforos no podemos recuperar los tiempos de congestión.



Figura 4.9: Red de ejemplo

Este mismo ejemplo permite ilustrar una deficiencia que puede tener emplear el tiempo medio en la red en lugar de observaciones de tiempo de viajes en ciertos pares.

Supóngase que una red dispone de algún arco con *flujo cautivo* o fijo como, lo es el arco (a, b) . Para cualquier estimación de la congestión en el subconjunto de arcos dado, si el tiempo medio asociado es menor que el observado (el de referencia) se puede incrementar la congestión en el arco hasta hacerla coincidir exactamente con el tiempo medio observado. Este incremento de la congestión en el arco cautivo no haría cambiar su flujo. La conclusión es que ese dato solo aportaría información en la estimación de la congestión del arco cautivo, probablemente dando un valor extremadamente grande. Este es el motivo de no emplear en estas pruebas numéricas el tiempo medio en la red, aunque los resultados obtenidos en el experimento 1 son buenos.

Para entender la solución hallada es necesario que nos preguntemos ¿Qué tipo de información suministran los aforos en el cálculo de tiempo de congestión? Para responderla supongamos el caso que tenemos un modelo de asignación de tráfico sin congestión (tipo todo-o-nada) y lo que se está haciendo es afinar en el cálculo del tiempo de viaje por el arco vacío. Los aforos suministrarían no un valor puntual sino un intervalo de tiempo para el cual los usuarios no cambiaría de ruta. Cualquier valor de esos intervalos serían igualmente verosímiles (o en otros palabras, mínimos locales).

Una situación parecida también ocurre en el caso con congestión. Supóngase que cuando se esta resolviendo el PCPCA(ρ) se tiene un arco cautivo, esto ocurre en este contexto del problema restringido cuando todos los pares que emplean dicho arco usan un único camino, esto hace que cualquier valor de la congestión para dicho arco posea el mismo valor de la función objetivo. En este tipo de arcos los aforos suministran intervalos para el valor del parámetro. Únicamente cuando el flujo de un determinado arco tiene la posibilidad de ser enrutado por arcos alternativos, y esta variación de flujo es continua respecto al valor del tiempo de congestión, es cuando se produce una estimación puntual.

En este sentido, la solución obtenida en el problema 2 muestra que solamente en algunos de los arcos se obtiene el verdadero tiempo de congestión y en otros no. Pequeñas variaciones de los valores de los parámetros en esos arcos no consiguen cambiar el flujo en los arcos, y por tanto mejorar la solución, haciendo que se trate de un óptimo local. Al introducir la información del tiempo de viaje en el equilibrio se tiene información del coste de congestión en todos los arcos que están siendo usados y este hecho proporciona una estimación puntual. En el problema 2 al no contener los datos errores la solución es capaz de reproducir los verdaderos valores.

En la red de Hull, en la solución de equilibrio para la parametrización original, solamente 11 de los 142 pares emplean más de un camino. Esto provoca que muchos de los problemas de optimización presentados al resolver los problemas restringidos de congestión se lleguen a diferentes óptimos locales. Otra observación es que con los tiempos de viaje en el equilibrio (problema 3) se consigue estimar todos los arcos, exceptuando uno.

Para visualizar mejor la solución obtenida se representará para los cuatro problemas los tiempos de viaje y los flujos (predichos) en el último PCPCA(ρ) restringido frente a los verdaderos valores y se mostrará el coeficiente de determinación. Se observa

que aunque en algún caso las soluciones verdaderas no son las óptimas, las soluciones obtenidas son muy alta calidad (Se realiza un estudio de correlación lineal).

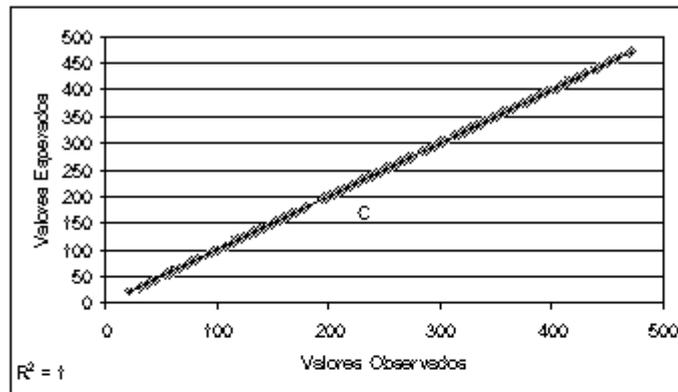


Figura 4.10: Problema 1: Tiempo en los pares O-D

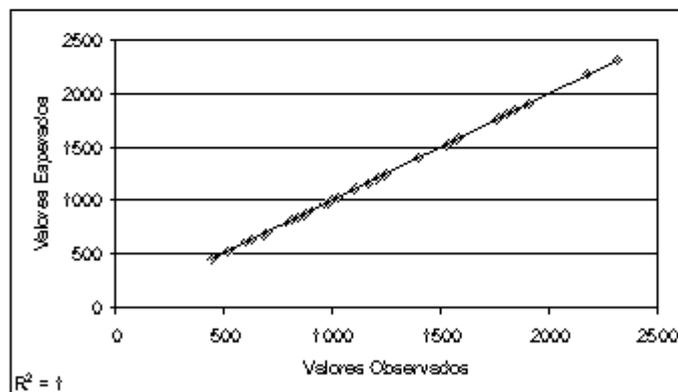


Figura 4.11: Problema 1: Flujo en los arcos

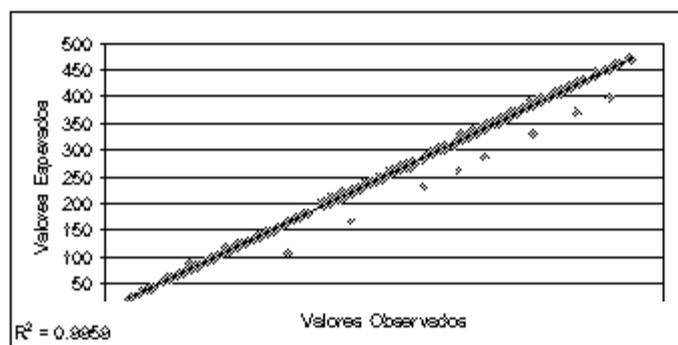


Figura 4.12: Problema 2: Tiempo en los pares O-D

La tabla 4.10 muestra los resultados computacionales obtenidos. En la primera columna se indica el problema resuelto, la segunda muestra el valor alcanzado de la

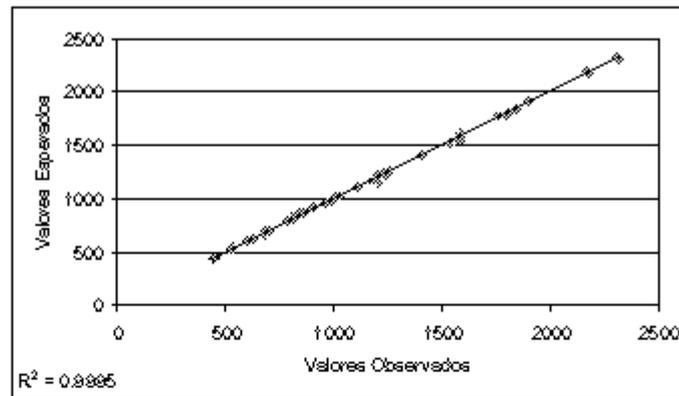


Figura 4.13: Problema 2: Flujo en los arcos

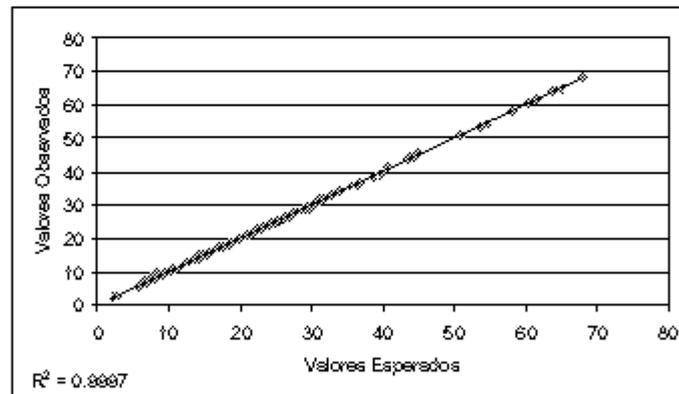
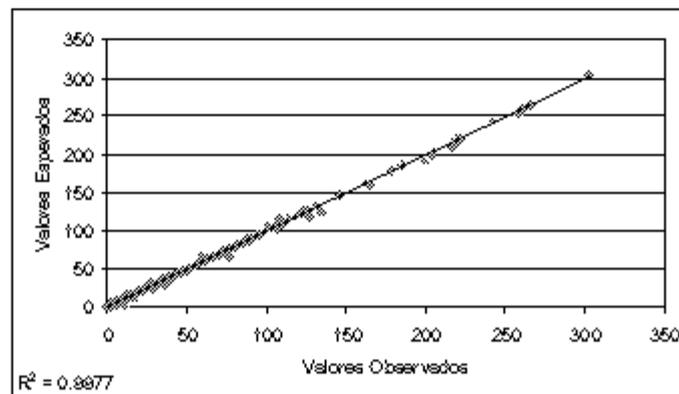


Figura 4.14: Problema 3: Tiempo en los pares O-D



función objetivo, en la tercera incluye el número de segundos de CPU (incluyendo todas las operaciones de entrada y salida de datos), la cuarta columna recoge el tiempo dedicado exclusivamente a la resolución de los PCPCAs y en la siguiente a los RMPs. En las dos últimas columnas se recoge el número de iteraciones principales y el número de iteraciones totales del algoritmo DSD para resolver todos los modelos de asignación

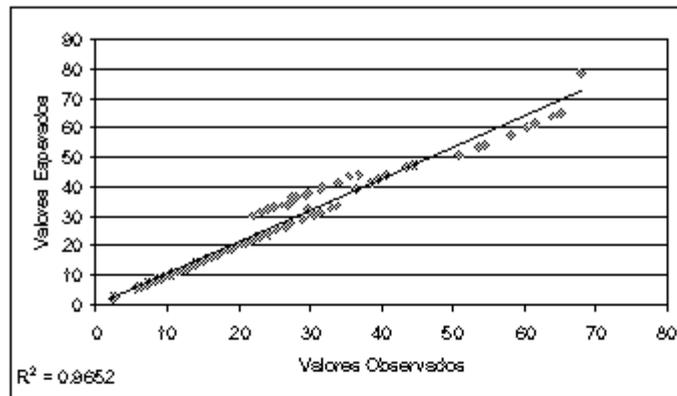


Figura 4.16: Problema 4: Tiempo en los pares O-D

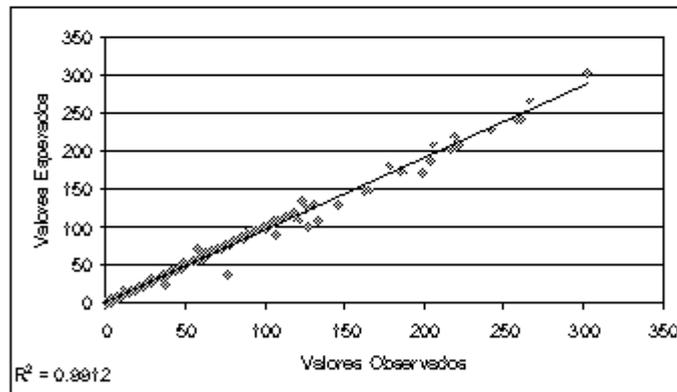


Figura 4.17: Problema 4: Flujo en los arcos

de tráfico.

La observación más importante es que el número de iteraciones principales es relativamente pequeño, 5 iteraciones para los problemas donde se incluye el coste en equilibrio. Lo que unido a que el algoritmo DSD es capaz de resolver redes de mediano tamaño, sugiere que este algoritmo puede ser aplicable a problemas reales de tamaño mediano, siempre y cuando los problemas PCPCA se pudieran resolver. También se observa que el coste computacional de estos problemas es muy superior a la resolución de los problemas de RMPs, lo que hace que esta última cuestión no sea tan evidente.

Otra observación es que disponer del coste de equilibrio no solamente conduce a los modelos. Esto es, en general mejora la estimaciones y simplifica su resolución.

Para terminar el experimento 2, vamos a aplicar el $PCPCA(\rho)$ a Ciudad Real. Se trata de una población de unos 75.000 habitantes situada en el centro-sur-oeste español. Esta red consta de 230 nodos, 456 pares origen-destino, y 456 arcos. Esta red está poco congestionada, una prueba de ello es que solamente 10 de los 456 arcos originales emplean más de dos caminos en la situación de equilibrio, calculando ésta

con un error relativo de $3E. - 05$. En la citada situación de equilibrio hay 469 caminos para un total de 456 pares origen-destino.

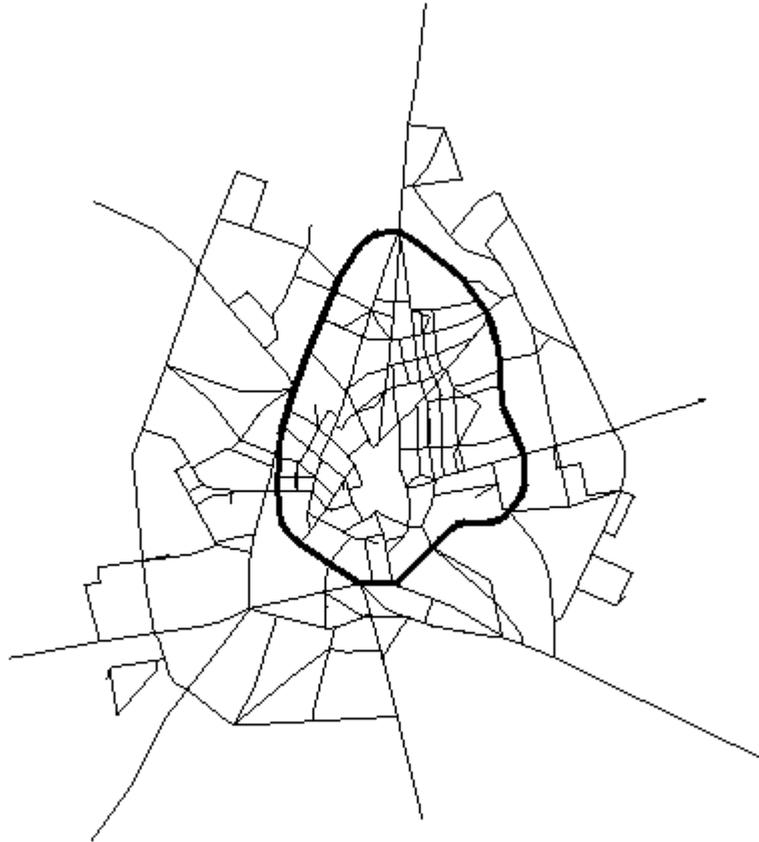


Figura 4.18: Red de Ciudad Real

El objetivo de este experimento es evaluar la aplicabilidad de esta metodología (PCPCA(ρ)) a ciudades de pequeño o mediano tamaño. Nos planteamos estimar los tiempos de congestión de la vía más importante de la ciudad, la denominada ronda de circunvalación. Los datos disponibles son los aforos de 11 arcos durante el periodo de estudio, la hora pico de 8 : 30 – 9 : 30, y una estimación del tiempo de viaje en los pares que emplean la ronda de circunvalación. Mediante la técnica del *coche flotante* obtuvimos las velocidades medias en los arcos de la ronda de circunvalación y en el resto de la ciudad para el periodo de estudio. Con estas velocidades, y a partir de las longitudes de los arcos, ajustamos los tiempos de viaje para los distintos pares.

cuadrados ponderados. Se ha planteado los tres siguientes modelos en función de la información disponible:

Caso 1: Ponderación de aforos y tiempos de viajes con un valor $\eta = 0.001$.

Caso 2: Uso de solamente los aforos ($\eta = 1$).

Caso 3: Uso de únicamente los tiempos de viajes ($\gamma = 0$).

Hemos empleado como criterio de paro $|Z^{\ell+1} - Z^{\ell}| < 0.01$. En la tabla 4.11 se presentan los resultados computacionales para la red de Ciudad Real. Aquí no podemos comparar con los verdaderos valores ya que es un caso real y estos valores no son conocidos. El objetivo perseguido, más que analizar la calidad de la solución, es evaluar el rendimiento del algoritmo. El tiempo de ejecución en las tres situaciones es de unos 5 minutos, lo que pone de manifiesto que es aplicable en este tipo de estudios de movilidad en ciudades de tamaño pequeño o mediano.

Esta red es poco congestionada y los usuarios emplean, por lo general, un único camino para satisfacer su demanda. En este contexto, la metodología desarrollada puede ser vista como un procedimiento para estimar las rutas empleadas por los usuarios de modo que el flujo en la red y los tiempos de viaje se ajusten lo más posible al conjunto de observaciones.

4.5 Conclusiones

En este trabajo se ha presentado una metodología para ajustar/estimar la parametrización de las funciones de coste en los arcos empleadas en el problema (estático) de asignación de tráfico.

Tras discutir la información necesaria para este objetivo, se han desarrollado dos modelos binivel: PCPCA(ρ) y PCPCA(P). El PCPCA(ρ) emplea información sobre aforos y el coste de equilibrio en algunos pares de la red, obtenidos en un único periodo y tiene por objetivo estimar los costes/tiempos de congestión en los arcos y, a partir de estos tiempos, empleando una regla adicional, estima la parametrización original. El PCPCA(P) usa información de dos o más periodos y su objetivo es estimar directamente la parametrización original.

Se ha desarrollado un algoritmo de resolución de estos modelos binivel procediendo, previamente, a su reformulación mediante un problema de optimización de un solo nivel y, posteriormente, se ha aplicado un esquema de generación de columnas (CGA) a esta reformulación. Se ha probado la convergencia de este algoritmo a óptimos locales.

Este estudio teórico se ha completado con un estudio computacional en el que se para estimar los tiempos de viaje en el futuro, motivando este hecho la necesidad de mejorar los métodos para obtener estas parametrizaciones. En el estudio computacional se pone de manifiesto i) lo insuficiente de la información recogida en los aforos, ii) la convergencia a mínimos locales del algoritmo propuesto y iii) la aplicabilidad de la metodología a redes reales de pequeño o mediano tamaño, como se ilustrado en la red de Ciudad Real.

Apéndice

En este apéndice se justifica que la formulación $VIP(C, \Omega)$ correspondiente a un TAP bajo el criterio DUE puede recoger los criterios SE y SUE-logit. Esto es, existe una cierta función de coste, en función de la función de coste original, de tal modo que si es usada en un TAP bajo el punto de vista DUE, el resultado es una asignación bajo el punto de vista SE (o SUE-logit) para las funciones de coste originales.

Para ilustrar la idea, consideremos el criterio SE y el caso donde las funciones de coste son separables (el coste en cada arco depende únicamente del flujo en ese arco). Deseamos encontrar unas funciones $\hat{c}_a(v_a)$ de modo que si la empleásemos en un TAP-DUE obtuviésemos la solución del TAP-SE para las funciones originales $c_a(v_a)$. Esto se puede conseguir si la función objetivo del problema TAP-DUE coincidiese con la del TAP-SE. Entonces se debe cumplir

$$\int_0^{v_a} \hat{c}_a(x) dx = c_a(v_a)v_a$$

Derivando a ambos lados de la igualdad obtenemos:

$$\hat{c}_a(x) dx = c_a(v_a) + c'_a(v_a)v_a$$

Por ejemplo, si aplicásemos la anterior fórmula a las funciones de coste BPR $c_a(v_a) = t_a^0 + \alpha_a \left(\frac{v_a}{L_a}\right)^{n_a}$ obtendríamos $\hat{c}_a(v_a) = t_a^0 + (n_a + 1)\alpha_a \left(\frac{v_a}{L_a}\right)^{n_a}$.

A continuación repetiremos lo hecho anteriormente, pero para el caso de emplear funciones de coste asimétricas (aquellas cuya matriz jacobiana es asimétrica). El TAP-SE puede formularse como el siguiente problema de optimización:

$$\begin{aligned} \text{Minimizar } Z(h) &= C(h)^T h && \text{TAP-SE} \\ \text{sujeto a: } & h \in \Omega \end{aligned}$$

Bajo las hipótesis adecuadas, de modo que garanticen que los mínimos locales del anterior problema son mínimos globales, las condiciones de optimalidad conducen a que las soluciones deben satisfacer:

$$\nabla Z(h^*)^T (h - h^*) \geq 0, \quad \forall h \in \Omega \quad [VIP-SE(C, \Omega)]$$

Haciendo operaciones

donde $\nabla_h C$ es la matriz Jacobiana de $C(h)$. Bajo la hipótesis ii) (aditividad de las funciones de coste) podemos reescribir $\nabla Z(h^*)$ en términos de los flujos en los arcos

$$\begin{aligned} \nabla Z_h(h^*) &= \left[\Delta^T \nabla_v c(v^*)^T \Delta \right]^T h^* + \Delta c(v^*) \\ &= \Delta^T \nabla_v c(v^*) \Delta h^* + \Delta^T c(v^*) \\ &= \Delta^T [\Delta_v c(v)v + c(v)] \end{aligned}$$

donde $v^* = \Delta h^*$.

Finalmente, la expresión de $\nabla Z(h^*)$ en el caso de funciones de coste separables, se simplifica a la expresión obtenida inicialmente de:

$$\nabla Z(h^*) = \sum_{a \in \mathcal{A}} \delta_{ak} [c'_a(v_a)v_a + c_a(v_a)]$$

En el TAP-SE deberíamos exigir, para poderse satisfacer la condición i), que la función $c(v)$ fuese continuamente diferenciable.

En el caso TAP-SUE tipo logit, las funciones a utilizar son:

$$\hat{C}_k(h) := C_k(h) + \frac{1}{\theta} \ln h_k$$

En la situación de equilibrio tendríamos que, para los caminos del par i con flujo positivo ($k \in K_i^*$)

$$\hat{C}_k(h^*) := C_k(h^*) + \frac{1}{\theta} \ln h_k^* = U_i,$$

de donde se obtiene:

$$h_k^* = \exp[-\theta(U_i - C_k(h^*))]$$

y sumando todos los caminos

$$g_i = \sum_{k \in K_i^*} h_k^* = \sum_{k \in K_i^*} \exp[\theta(U_i - C_k(h^*))]$$

y por tanto se verifica que las proporciones de uso de los caminos vienen dadas por:

$$\frac{h_k^*}{g_i} = \frac{\exp[\theta(U_i - C_k(h^*))]}{\sum_{k \in K_i^*} \exp[\theta(U_i - C_k(h^*))]} = \frac{\exp[-\theta C_k(h^*)]}{\sum_{k \in K_i^*} \exp[-\theta C_k(h^*)]}$$

Para garantizar la continuidad exigida en i), si suponemos que el conjunto de parámetro \mathcal{K} es un conjunto compacto, podremos restringir el conjunto de flujos factibles Ω a un cierto conjunto

$$\hat{\Omega} := \{h \in \Omega / h_k > \epsilon\}$$

sin pérdida de generalidad. Esto es, el conjunto de soluciones de $\text{VIP}(\hat{C}(\cdot, \rho), \Omega)$ coincide con el de $\text{VIP}(\hat{C}(\cdot, \rho), \hat{\Omega})$ para todo $\rho \in \mathcal{K}$.

ta hipótesis viene derivada del objetivo de este trabajo, que es la calibración de los parámetros en el arco, y por tanto parece natural considerar que las funciones de coste en los caminos sean aditivas. Todo este trabajo se puede plantear relajando esta hipótesis, pero por buscar la claridad no se ha considerado. El marco más general es plantearse la calibración de los parámetros de las funciones de coste en los caminos y en este contexto se puede plantear la propia calibración del parámetro θ .

Tabla 4.1: Notación utilizada para formular el TAP

\mathcal{N} :	conjunto de nodos de la red de tráfico
\mathcal{A} :	conjunto de arcos de la red de tráfico
$a \in \mathcal{A}$:	un arco de la red de tráfico
I :	Conjunto de pares origen-destino
$i \in I$:	un par origen-destino
K :	conjunto de caminos en la red
$k \in K$:	un camino en la red
K_i :	conjunto de caminos posibles para el par origen-destino i
g_i :	demanda de tráfico para el par origen-destino i
g :	vector con todas las demandas de tráfico (\dots, g_i, \dots) . Esta es la denominada matriz de viajes origen-destino
$\Delta = [\delta_{ak}]$:	matriz de incidencia arco/camino, donde $\delta_{ak} = 1$ si el arco a pertenece al camino k , 0 en caso contrario
$\Lambda = [\lambda_{ik}]$:	matriz de incidencia camino/par origen-destino, donde $\lambda_{ik} = 1$ si el camino $k \in K_i$, 0 en caso contrario
h_k :	flujo en el camino k
h :	vector de flujo en los caminos (\dots, h_k, \dots)
v_a :	flujo en el arco a
v :	vector de flujo en los arcos (\dots, v_a, \dots) ; observese que $v = \Delta h$
$c_a(v)$:	el coste en el arco a como una función del flujo en todos los arcos de la red
$c(v)$:	vector de funciones de coste en los arcos $(\dots, c_a(v), \dots)$
$C_k(h)$:	el coste en el camino k como una función del flujo en todos los caminos de la
$C(h)$:	vector de funciones de coste en los caminos; se asume que $C(h) = \Delta^T c(v) = \Delta^T c(\Delta h)$
U_i :	Coste de equilibrio para el par origen-destino i
U :	Vector de costes de equilibrio

Tabla 4.2: El algoritmo de generación de columnas (CGA)

0. (*Inicialización*). Sea ρ^1 un vector inicial de parámetros, sea $K^0 = \emptyset$ y sea $\ell = 1$.
1. (*Problema de asignación de tráfico*). Sea $C(h, \rho)$ la función de costes en los caminos, entonces resuélvase $VIP(C(\cdot, \rho), \Omega)$. Sea $K(\rho)$ el conjunto de caminos de mínimo coste en el equilibrio para el vector de parámetros ρ .
2. (*Chequeo de la optimalidad*). Si $K(\rho) \subset K^{\ell-1}$ entonces Stop, ρ es un mínimo local para el PCPCA. En otro caso, sea $K^\ell = K(\rho) \cup K^{\ell-1}$.
3. (*PCPCA restringido*). Resuélvase el $RPCPCA(K^\ell)$. Sea $\rho^{\ell+1}$ un mínimo local.
4. (*Criterio de paro*). Si $\rho^{\ell+1}$ es una solución aceptable, entonces parar, en caso contrario sea $\ell = \ell + 1$ y vuelvase al paso 1.

Tabla 4.3: Predicciones 2g Modelo PCPCA(ρ) con coste en los pares O-D

Escenarios	$m_a = 2$			$m_a = 5$		
	(1)	(2)	(3)	(1)	(2)	(3)
TAP-UE	0.00	0.00	0.00	-100.00	25.91	15.25
				a_{17}	\dot{e}_1	
TAP-SE	29.10	-5.10	-3.99	-100.00	14.20	6.88
	a_{17}	\dot{e}_1		a_{17}	\dot{e}_2	
TAP-CC	-9.75	2.18	1.50	-28.31	7.70	3.81
	a_{18}	\dot{e}_1		a_{17}	\dot{e}_1	
TAP-SUE	-31.74	27.07	16.97	53.93	46.49	29.49
$\frac{1}{\delta} = 25$	a_{17}	\dot{e}_1		a_{18}, a_{19}	\dot{e}_1	
TAP-SUE	-23.09	28.98	19.10	52.96	56.91	38.10
$\frac{1}{\delta} = 50$	a_{17}	\dot{e}_1		a_{18}, a_{19}	\dot{e}_1	

(2) Máxima diferencia relativa del coste medio en los pares en %

(3) Máxima diferencia relativa del coste medio en la red en %

Tabla 4.4: Predicciones 2g Modelo PCPCA(ρ) con coste en la red

Escenarios	$m_a = 2$			$m_a = 5$		
	(1)	(2)	(3)	(1)	(2)	(3)
TAP-UE	0.00	0.00	0.00	-100.00	25.91	15.25
				a_{17}	i_1	
TAP-SE	29.07	-5.07	-3.87	-100.00	14.43	7.06
	a_{17}	i_1		a_{17}	i_2	
TAP-CC	5.88	3.79	0.43	-44.59	10.86	5.87
	a_{18}	i_3		a_{17}	i_1	
TAP-SUE	-34.23	23.93	10.44	91.57	57.91	39.47
$\frac{1}{\rho} = 25$	a_{17}	i_1		a_{18}, a_{19}	i_1	
TAP-SUE	-25.49	32.14	21.24	55.23	58.67	39.23
$\frac{1}{\rho} = 50$	a_{17}	i_1		a_{18}, a_{19}	i_1	

(1) Máxima diferencia relativa del flujo en los arcos en %

(2) Máxima diferencia relativa del coste medio en los pares en %

(3) Máxima diferencia relativa del coste medio en la red en %

Tabla 4.5: Predicciones 2g Modelo PCPCA(P) con coste en los pares y en la red

Escenarios	PCPCA(P) Pares			PCPCA(P) Red		
	(1)	(2)	(3)	(1)	(2)	(3)
TAP-UE	0.00	0.00	0.00	0.00	0.00	0.00
TAP-SE	-25.88	-3.33	-2.78	7.88	-2.14	-0.89
	a_{19}	i_1		a_9	i_1	
TAP-CC	-100.00	13.12	4.98	3.60	7.73	1.79
	a_{18}	i_2		a_{18}	i_3	
TAP-SUE	-100.00	33.45	16.08	-46.59	39.43	23.63
$\frac{1}{\rho} = 25$	a_{18}	i_1		a_{17}	i_1	
TAP-SUE	-20.22	25.65	17.11	-29.39	36.83	24.14

(1) Máxima diferencia relativa del flujo en los arcos en %

(2) Máxima diferencia relativa del coste medio en los pares en %

(3) Máxima diferencia relativa del coste medio en la red en %

Tabla 4.6: PCPCA(P): Predicción de los parámetros coste arcos con tiempos medios en los pares O-D

Escenarios	α_2		α_{14}	
	d	m	d	m
Reales	0.0050000000	2.0000000000	0.0012500000	2.0000000000
TAP-DUE	0.0050000000	2.0000000000	0.0012500000	2.0000000000
TAP-SE	0.0024192889	2.0667230280	0.0057927021	1.7263100771
TAP-CC	0.0006434864	2.3310516806	0.0000337622	2.9618901911
TAP-SUE $\frac{1}{\theta} = 25$	0.0669086130	1.8084368295	0.0117048202	1.8795675370
TAP-SUE $\frac{1}{\theta} = 50$	0.0789455653	1.8026750221	0.0066802134	1.9757652145

Tabla 4.7: PCPCA(P): Predicción de los parámetros coste arcos con tiempos medios en la red

Escenarios	α_2		α_{14}	
	d	m	d	m
Reales	0.0050000000	2.0000000000	0.0012500000	2.0000000000
TAP-DUE	0.0049999856	2.0000003809	0.0012499907	2.0000010909
TAP-SE	0.0003626880	2.3309288861	0.0000042539	2.7101712371
TAP-CC	0.0034336142	2.0386681816	0.0008797941	2.5005849504
TAP-SUE $\frac{1}{\theta} = 25$	0.4653752146	1.6218319798	0.0015486215	2.2057607278
TAP-SUE $\frac{1}{\theta} = 50$	0.0092755977	2.1561952826	0.0015486215	2.2057607278

Tabla 4.8: Definición de los problemas

PCPCA(ρ)	Red	F
1	Sioux Falls	$\frac{1}{2} \sum_{i \in I} \frac{(U_i^* - \bar{U}_i)^2}{\bar{U}_i} + \frac{1}{2} \sum_{a \in A} \frac{(f_a - \bar{f}_a)^2}{\bar{f}_a}$
2	Sioux Falls	$\sum_{a \in A} \frac{(f_a - \bar{f}_a)^2}{\bar{f}_a}$
3	Hull	$\frac{1}{2} \sum_{i \in I} \frac{(U_i^* - \bar{U}_i)^2}{\bar{U}_i} + \frac{1}{2} \sum_{a \in A} \frac{(f_a - \bar{f}_a)^2}{\bar{f}_a}$
4	Hull	$\sum_{a \in A} \frac{(f_a - \bar{f}_a)^2}{\bar{f}_a}$
PCPCA(ρ)	Congestión en los arcos	
1, 2	a10 a20 a30 a40 a50 a60 a70	
	11.3330 25.0141 83.0802 98.4265 1.6346 2.6023 83.6581	
3, 4	a5 a9 a12 a323 a378 a623 a633	
	13.4664 6.5016 19.4199 5.3609 9.8621 16.8265 6.4979	

Tabla 4.9: Resultados obtenidos

PCPCA(ρ)		Congestión en los arcos						
Z		a10	a20	a30	a40	a50	a60	a70
Verdadera		11.3330	25.0141	83.0802	98.4265	1.6346	2.6023	83.6581
1	0.0000	11.3330	25.0141	83.0802	98.4265	1.6346	2.6023	83.6581
2	6.8576	12.9789	0.0000	28.0177	98.4023	1.9024	0.8028	83.3753
Verdadera		a5	a9	a12	a323	a378	a623	a633
		13.4664	6.5016	19.4199	5.3609	9.8621	16.8265	6.4979
3	54.6812	13.1008	6.5016	19.4199	5.1879	9.8621	17.5872	6.9277
4	109.80	16.1740	15.0000	19.0339	12.8199	9.8624	17.1218	11.1339

Tabla 4.10: Resultados computacionales

PCPCA(ρ)	Z	CPU	CPU ^a	CPU ^b	ℓ	ℓ_{RMP}
1	0.0000	253.1	15.3	2.2	5	25
2	6.8771	724.4	36.6	4.4	9	50
3	54.6812	59.5	3.2	1.9	3	10
4	109.8087	65.8	2.1	1.9	3	11

^a Segundos de CPU empleados en los problemas PCPCA.

^b Segundos de CPU empleados en los RMPs del DSD.

Tabla 4.11: Resultados computacionales para la red de Ciudad Real

η	Z	CPU	CPU ^a	CPU ^b	ℓ	ℓ_{RMP}
0	20.5	302.6	25.4	4.7	3	11

^a Segundos de CPU empleados en los problemas PCPCA.

^b Segundos de CPU empleados en los RMPs del DSD.

Capítulo 5

Conclusiones y sugerencias para futuras líneas de investigación

5.1 Conclusiones

El objetivo principal de esta tesis es la realización de aportaciones importantes en el campo de la estimación de matrices origen-destino y en la calibración de parámetros en el problema de asignación de tráfico en redes congestionadas. Tras su desarrollo, se han alcanzado tres grandes grupos de objetivos que desarrollan el principal:

Objetivo 1: Desarrollo de una metodología para la resolución del DAP

El problema de la estimación de la matriz origen-destino en redes de tráfico congestionadas (DAP) en problemas reales, de grandes dimensiones, requiere de algoritmos altamente especializados que empleen la estructura del problema a tratar, y en cuya implementación se tenga especial atención a los aspectos informáticos tales como las estructuras de los datos y la estructuración de los códigos, de forma que sea factible su integración en entornos de editores gráficos de redes especializados. En este contexto, con esta tesis, se ha pretendido desarrollar una metodología para la resolución del DAP que pudiera ser aplicada a redes de tráfico urbano reales.

Para alcanzar este objetivo, la propuesta algorítmica inicial ha arrancado de un *esquema de generación de columnas* que ha sido exitosamente aplicado al problema de el espacio de flujo en los caminos, frente a la tradicional formulación sobre el espacio de flujo en los arcos. Sobre esta nueva formulación, se ha aplicado un esquema de generación de columnas/variables (Algoritmo CGA), en el contexto de programación matemática con restricciones de equilibrio. Esta técnica puede ser vista como una introducción de un esquema de generación de columnas en la denominada programación implícita que permita hacerla viable para problemas de grandes dimensiones.

El algoritmo CGA propuesto resuelve iterativamente un problema de asignación para una matriz OD dada, cuya solución permite obtener nuevos caminos, y un problema de estimación, en el que se resuelve el DAP original restringido a los caminos generados. En el DAP restringido se reemplaza el problema inferior restringido por sus condiciones de optimalidad de primer orden, las llamadas condiciones de Karush-Kuhn-Tucker. Esta operación conduce a un problema no lineal y no convexo de un solo nivel.

Bajo la hipótesis de continuidad de las funciones de coste en los caminos y de la norma usada en el problema de nivel superior, se demuestra la convergencia del algoritmo a un óptimo local, asumiendo cierta condición de regularidad similar a las condiciones de cualificación de restricciones. Concretamente, se ha empleado la selección de todos los caminos de mínimo coste, en lugar de los caminos con flujo en la situación de equilibrio. Esto ha permitido demostrar la convergencia a un óptimo local del DAP, sin necesidad de que la sucesión de patrones de equilibrio cumpla la condición de holgura complementaria estricta.

Se han realizado experiencias numéricas donde se ha contrastado la citada convergencia del algoritmo CGA cuando se cumplen las hipótesis mencionadas e, incluso, se han presentado evidencias de que se podrían relajar, en cierta forma, dichas hipótesis. También se ha comprobado que el número de iteraciones necesarias para alcanzar la convergencia es pequeño y que depende de la identificación del conjunto de caminos en equilibrio para los mínimos locales, esto, unido a que dos de las estrategias heurísticas que se presentan funcionan bien, hace pensar que estos métodos se pueden aplicar a problemas de gran escala. Se han comprobado, también, que la convergencia se da a óptimos locales, esto justifica que se deba realizar, en el futuro, un esfuerzo de investigación sobre cómo abordar problemas de gran escala con múltiples óptimos locales.

Objetivo 2: Elaboración de metodologías heurísticas de resolución del DAP

La metodología desarrollada en el objetivo 1 puede requerir de altas prestaciones computacionales que hagan necesario algún tipo de relajación o aproximación en su aplicación a problemas reales. En este sentido se han desarrollado estrategias heurísticas que conducen a una implementación satisfactoria de la metodología desarrollada en el objetivo 1, buscando un compromiso entre la exactitud de la solución obtenida, para el problema de estimación, y el esfuerzo computacional necesario para obtenerla.

Así, se ha desarrollado un marco para elaborar algoritmos heurísticos, denominados *CGA heurísticos* para la resolución del DAP donde se utiliza como estrategia en este contexto, se muestra que los algoritmos IEA (*Iterative estimation assignment*, Yang et al. (1992)) y SAB (*Sensitive analysis based*, Yang (1995)) son ejemplos de esta clase de algoritmos, los cuales, analizados bajo esta nueva óptica, presentan ciertas ventajas: i) permiten una mejor implementación computacional en el espacio de flujo en los caminos, en lugar de flujo en los arcos, ii) generaliza la definición del SAB y iii) permite obtener una condición suficiente para la convergencia del SAB a un punto de

condiciones débiles en las funciones de coste en los arcos (continuidad) y se ha realizado un análisis de la sensibilidad de esas soluciones.

Por otra parte, se ha propuesto un algoritmo de generación de columnas para resolver los dos modelos que, previamente, se han transformado en un programa matemático equivalente de un sólo nivel, en el espacio de flujo en los caminos. Mediante el análisis de la sensibilidad, se ha concluido que el algoritmo propuesto converge a óptimos locales del PCPCA. Este algoritmo es una adaptación exitosa, del algoritmo CGA desarrollado en el capítulo 2 para el DAP, a los modelos PCPCA. Esto permite el estudio unificado de ambos problemas.

Asimismo, se ha completado un estudio numérico, sobre redes de pequeño y mediano tamaño, con los objetivos de estudiar los modelos PCPCA propuestos y los aspectos computacionales del algoritmo de resolución.

Respecto de los modelos propuestos se deduce que la relación entre el comportamiento real del usuario y el modelo de asignación en equilibrio, utilizado para modelarlo, tiene un efecto muy significativo en los errores. Además, en general, se obtienen mejores resultados cuando se utiliza como información adicional el coste medio en los pares que el coste medio en la red. Otra observación importante es que se pueden obtener mínimos locales no globales y que si se dispusiera de un buen criterio adicional para el primer modelo propuesto, los resultados obtenidos serían equivalentes a los obtenidos por el segundo modelo. También se concluye que los modelos basados únicamente en mediciones de aforo no obtienen los verdaderos valores de la congestión (Esto es una característica del PCPCA (ρ) y no del algoritmo empelado).

Respecto del algoritmo de resolución propuesto, se ha demostrado que el número de iteraciones empleado es relativamente pequeño, lo que unido a que el algoritmo DSD es capaz de resolver redes de mediano tamaño, sugiere que este algoritmo puede ser aplicable a problemas reales de tamaño medio, siempre y cuando los problemas PCPCA(.) se puedan resolver. Por otra parte, se ha puesto de manifiesto que disponer del coste de equilibrio no solamente conduce a mejoras en las estimaciones, sino que reduce el coste computacional de la resolución de los modelos.

Finalmente se aplica el método PCPCA(ρ) a la red de tráfico de Ciudad Real, para evaluar la aplicabilidad de esta metodología a ciudades de pequeño o mediano tamaño. En este caso, los resultados no se han podido comparar con los verdaderos valores, ya que es un caso real y estos valores no son conocidos. Sin embargo, si que ha servido para evaluar el rendimiento del algoritmo. El tiempo de ejecución, en las tres situaciones contempladas, ha sido muy bajo, lo que pone de manifiesto que es aplicable a estudios

5.2 Novedad y relevancia

En la actualidad existen muy pocos métodos para la estimación de matrices O-D a partir de volúmenes de tráfico. El más eficiente, que está implementado en un paquete

comercial, es el método de Spiess, implementado en forma de macro en el paquete de planificación de transporte EMME/2. Este método tiene tres deficiencias significativas. La primera es que no emplea matrices O-D desactualizadas u obtenidas mediante otros procedimientos (eg.: encuestas domiciliarias), la segunda es que sólo utiliza como métrica de ajuste la de mínimos cuadrados y la tercera es que la convergencia del método es lenta y no está garantizada hacia un óptimo global del problema. La posibilidad de paliar estas tres deficiencias es altamente sugerente puesto que apunta a afianzar la robustez de estas técnicas.

La utilización del DAP para la estimación de matrices origen destino en redes de tráfico congestionadas, a partir de aforos y matrices desactualizadas, , respecto del *estado-de-la-práctica*, supone importantes reducciones de costes monetarios y humanos en la realización de estudios de movilidad y pasa por disponer de algoritmos específicos. En este aspecto es donde esta tesis hace su contribución fundameteal. Esto es así, pues el método presentado, no solamente tiene un sustento teórico, sino que es eficiente y recoge un buen número de situaciones prácticas.

En este sentido, es interesante utilizar una estrategia complementaria para la aplicación de los algoritmos que se proponen a problemas de gran escala. Estrategia que consiste en considerar las métricas derivadas de las normas uno e infinito, las cuales conducen a una función objetivo lineal para el DAP(restringido), lo que unido a ciertas aproximaciones lineales convierte los problemas de estimación en problemas lineales, siendo este un aspecto novedoso en la modelización del DAP.

Finalmente, respecto a la metodología desarrollada para la estimación de parámetros de las funciones de coste en los arcos a partir de aforos y otras informaciones disponibles, hay que indicar que esta metodología es novedosa en la literatura científica. Existen muy pocas publicaciones al respecto. El proceso tradicional de calibrado de los costes en los arcos es manual (prueba/error), lo que consume muchos recursos materiales y humanos. Frente a esto, la metodología propuesta presenta las siguientes ventajas: (i) Automatización del procedimiento, lo que acelera y reduce los costes de realización del estudio, (ii) Mejora de las estimaciones, al evaluar, de forma eficiente, mayor conjunto de valores. Además, esta metodología es fácilmente extensible a la estimación simultánea del submodelo de oferta (parámetros de coste en los arcos) y de la matriz origen-destino de la demanda. Esta extensión, a priori, produce mejores parametrizaciones de los modelos de asignación de tráfico.

5.3 Futuras líneas de investigación

Tras la realización de esta tesis, y teniendo en cuenta *el-estado-del-arte* y *el-estado-de-la-práctica* en el momento de su terminación, se puede concluir que los distintos métodos de estimación desarrollados en la la literatura del problema de estimar/actualizar matrices de viajes O-D se pueden definir especificando los siguientes elementos:

- i. Un modelo de asignación

- ii. Un método estadístico
- iii. Un conjunto de datos

La manera unificada de formular el anterior problema es usar la programación matemática binivel y el DAP, como ya se ha visto, puede ser formulado de la siguiente forma:

$$\begin{aligned} \text{Minimizar } & F(g) = \eta F_1(g, \hat{g}) + (1 - \eta) F_2(v, \hat{v}) \\ \text{sujeto a: } & v \in \mathcal{A}(g) \\ & g \in G \end{aligned} \quad \text{[DAP]}$$

Donde \mathcal{A} representa un modelo general de asignación (en equilibrio) que predice los aforos en la red en función de una matriz de viajes O-D g . La aplicación de la asignación es una función multievaluada, esto es, que para cada valor de la matriz g , posiblemente, existe un conjunto de vectores de flujo en equilibrio. Los parámetros positivos η y $(1 - \eta)$ reflejan la confianza en los datos disponibles, y las restricción $g \in G$ representa la región factible para el conjunto de matrices O-D. Este conjunto tiene en cuenta información a priori sobre el valor de la matriz, como por ejemplo la no negatividad de sus entradas o la posible acotación superior. La función objetivo se interpreta como una ponderación de dos objetivos, uno derivado de reproducir las observaciones de flujo y otro en reproducir los resultados obtenidos en el otro procedimiento de estimación. Las funciones F_1 y F_2 son dos métricas empleadas por el procedimiento estadístico que comparan los valores observados \hat{v} y \hat{g} con los valores predichos de los flujos en los arcos y en la matriz O-D.

Este tema, por su importancia práctica (se realizan cientos de estudios de movilidad a nivel mundial cada año) ha generado multitud de esfuerzos de investigación que han cubierto numerosos aspectos. Sin embargo, existen otros muchos que requieren una mejor comprensión. En este sentido se considera necesario abundar en los siguientes aspectos para mejorar su entendimiento:

- *Validación de los métodos de estimación.*
- *Definición del plan de aforamiento de la red.*

Validación de los métodos de estimación

Como ya se ha indicado, existen multitud de modelos DAP en función de las tres elecciones básicas: i) modelo de asignación, ii) método de estimación y iii) información disponible. Sin embargo, un hecho que no ha sido estudiado es la influencia de estas especificaciones en la fiabilidad de las estimaciones de las matrices O-D.

Además de la escasez de trabajos, cabe resaltar que éstos se desarrollan no propiamente para el DAP, sino para el problema de calibrar modelos de demanda usando simultáneamente información agregada y desagregada sobre el patrón de viaje en la zona en estudio. Este problema tiene gran similitud al DAP, siendo la diferencia esencial que el papel de la matriz O-D lo desempeña un conjunto de parámetros que definen un modelo de demanda. En este contexto, en Cascetta and Russo (1992) y Cascetta and Russo (1997) se desarrollaron formas funcionales específicas para las funciones F_1 y F_2 del DAP, derivadas de los procedimientos de estimación máximo verosímil, mínimos cuadrados generalizados y estimación bayesiana bajo el supuesto de que los errores de las observaciones de flujo son estadísticamente independientes.

Ante esta situación, se considera interesante que una línea de investigación futura fuese la validación de las metodologías de estimación de las matrices O-D a partir de aforos de la red. En este sentido, se podría estudiar la fiabilidad de las estimaciones de la matriz O-D en función del modelo DAP elegido, definido por el tipo de método estadístico, el tipo de selección de los aforos, el método de asignación empleado y la ponderación de las métricas elegidas. En concreto, aspectos relevantes que se podrían conocer sería qué expresiones funcionales elegir para F_1 y F_2 para una zona de estudio, cómo ponderar la información derivada de los aforos respecto a la información aún contenida en la matriz de referencia O-D, qué importancia tiene el uso de modelos de asignación en función del tamaño de la zona de estudio, en especial qué relevancia tiene este hecho en ciudades de tamaño mediano o pequeño con escasa congestión, etc.

Para llevar a cabo esto, se debería abordar su validación analítica mediante estudios de simulación. Así, mediante la simulación de Monte Carlo se podrían estudiar diversas alternativas para la modelización del DAP. Este tipo de estudios permitiría simular situaciones prácticas, pero con la diferencia crucial de que la matriz O-D a estimar es conocida. Este hecho nos permitiría evaluar la fiabilidad de las estimaciones obtenidas en cada alternativa estudiada para el DAP. De este estudio se podrían obtener reglas y recomendaciones para la elección de las métricas, para la elección del método, para generar el plan de aforamientos de la red y para la obtención del parámetro η de la función objetivo.

Por otra parte, se podría llevar a cabo una validación experimental mediante el estudio de casos reales. El objetivo podría ser contestar preguntas como las siguientes: (i) ¿Las metodologías basadas en aforos constituyen una alternativa a las metodologías basadas en la realización de encuestas? y (ii) ¿Qué formulación del DAP es más recomendable para cada situación práctica?

La pregunta (i) es la verificación experimental de la motivación fundamental sobre la que se sustenta el desarrollo de métodos de estimación basados en aforos: la hipótesis de que los métodos basados en aforos reducen el coste económico y temporal del estudio, sin minorar significativamente la exactitud obtenida, frente a los procedimientos tradicionales basados en encuestas domiciliarias.

La pregunta (ii) es la que se podría contestar con la simulación de Monte Carlos antes mencionada. Estas técnicas son extremadamente poderosas para estudiar sistemas complejos. No obstante, poseen una debilidad importante que radica en que se basan en hipótesis sobre el modo en que se perturban los errores en las mediciones de los flujos en los arcos y en una estructura de la matriz O-D a estimar elegida por el investigador. Esto podría producir que ciertos modelos DAP fuesen adecuados para los escenarios considerados en el estudio de simulación, pero su eficacia fuese menor en las situaciones reales.

Este hecho hace que el estudio anterior deba ser completado mediante el estudio de casos reales. En esta situación, la dificultad principal es el conocimiento de la matriz O-D verdadera, la cual es imprescindible para la evaluación de la fiabilidad de las distintas estimaciones obtenidas. Esta dificultad podría ser paliada si hiciéramos desempeñar a la estimación derivada del uso de encuestas domiciliarias el rol de la verdadera matriz O-D, siendo entonces el objetivo determinar qué modelizaciones del DAP producen resultados más próximos a la estimación a partir de encuestas.

Selección de los arcos a aforar

A diferencia del interés despertado en el desarrollo de modelos para la estimación de matrices O-D empleando aforos y matrices O-D de referencia, el número de métodos desarrollados para el problema de determinar qué arcos de la red aforar (*Traffic Count Location*, TCL) es escaso, pese a la importancia que este hecho tiene en la calidad final de la estimación de la matriz O-D.

Medir los aforos en todos los arcos de la red provee la mayor cantidad de información para la estimación de la matriz O-D, pero es, desde el punto de vista económico ineficiente, y en estudios (la mayoría) donde existan restricciones presupuestarias, inviable. Una solución a este problema pasa por desarrollar modelos matemáticos para la localización de los sensores de tráfico en una red viaria bajo criterios de captación de máxima congestión y baja correlación con la calidad de la estimación inicial de la matriz O-D. La programación matemática es la herramienta que actualmente se emplea en el desarrollo de modelos matemáticos para la localización de sensores de tráfico. Estos modelos eligen los arcos que deben de ser aforados como aquel conjunto que maximiza cierto criterio. Se podrían desarrollar nuevos modelos cuya función objetivo se base explícitamente en la congestión total de la red de tráfico y, por otro lado, no dependan fuertemente de la calidad de la estimación inicial de la matriz O-D.

literatura son TCL-P1 y TCL-P2 y su deficiencia esencial es que el TCL-P1 considera todos los pares igualmente importantes. El TCL-P2 corrige este hecho, haciendo que la importancia de los pares venga determinada por el número de usuarios en cada par. Este modelo maximiza el flujo neto interceptado por los sensores, aunque asume implícitamente que cualquier par de viaje origen destino es igualmente importante. En este sentido, se podrían perseguir modelos con características radicalmente diferentes

como que la importancia de un determinado tipo de viaje dependa de su contribución al nivel de congestión de la red, pudiendo, en alguna ocasión, no muestrear pares irrelevantes, con el objeto de poder obtener una mejor fiabilidad en la estimación final de la matriz O-D mediante una mayor preponderación de la topología de la red en el modelo. Lo que se pretendería conseguir es asignar un cierto nivel de incertidumbre a la congestión total en la red, en función de qué subconjunto de arcos hayan sido aforados.

.

Almadén, invierno 2005-2006

Referencias

- Abdulaal, M. and L. LeBlanc (1979). Methods for combining modal split and equilibrium assignment models. *Transportation Science* 13, 292-314.
- Aboussoror, A. and A. Mansouri (2005). Weak linear bilevel programming problems: existence of solutions via a penalty method. *Journal of mathematical analysis and applications* 304, 399-408.
- Abrahmsson, T. (1998). Estimation of origin-destination matrices using traffic counts - a literature survey. Interim report, International Institute for Applied Systems Analysis, A-2361 Laxenburg - Austria.
- Allsop, R. E. (1974). Some possibilities for using traffic control to influence trip distribution and route choice. In D. J. Buckley (Ed.), *Transportation and traffic theory (Proceedings of the Sixth International Symposium on Transportation and Traffic Theory)*, pp. 345-375. New York: Elsevier.
- Bar-Gera, H. (2002a). Origin-based algorithm for the traffic assignment problem. *Transportation Science* 36.
- Bar-Gera, H. (2002b). *Origin-based network assignment in Transportation Planning - State of the Art*. The Netherlands: M. Patriksson and M. Labbe (eds.).
- Barceló, J. (1997). Some mathematical programming models in transportation. *TOP, Journey of the Spanish Statistical and Operation Research Society* 5, 1-40.
- Barceló, J., J. Casas, E. Codina, A. Fernández, J. Ferrer, and D. García (1998). Real time microscopic simulation to support dynamic traffic management decisions. *Proceedings - Conference of the Australian Road Research Board*, 171-185.
- Barceló, J., E. Codina, J. Casas, J. Ferrer, and D. García (2005). Microscopic traffic simulation: A tool for the design, analysis and evaluation of intelligent transport systems. *Journal of Intelligent and Robotic Systems* 41(2-3), 173-203. . . . 15-27.
- Bazaraa, M. S., H. D. Sherali, and C. M. Shetty (1993). *Nonlinear Programming: Theory and Algorithms* (second ed.). New York, NY: John Wiley & Sons.
- Beckman, M. J., C. B. McGuire, and C. B. Wisten (1956). *Studies in the Economics of Transportation*. New Haven: Yale University Press.

- Bell, M. and Y. Iida (1997). *Transportation network analysis*. Chichester(UK): John Wiley and Sons.
- Bell, M., C. Shield, F. Busch, and G. Kruse (1996). A stochastic user equilibrium path flow estimator. *Transportation Research C* 5, 197-210.
- Bell, M., C. Shield, F. Busch, and G. Kruse (1997). A stochastic user equilibrium path flow estimator. *Transportation Research Part C: Emerging Technologies* 5(3-4), 197-210.
- Bell, M. G. H. (1983a). The estimation of an origin-destination flows and their confidence intervals from measurement of link volumes: a computer program. *Traffic Engineering and Control* 24(4), 202-205.
- Bell, M. G. H. (1983b). The estimation of an origin-destination matrix from traffic counts. *Transportation Science* 17, 198-217.
- Bell, M. G. H. (1984). Log-linear models for the estimation of origin-destination matrices from traffic counts. *Proc. of the Ninth International Symposium on Transportation and Traffic Theory. Delft, The Netherlands*.
- Bell, M. G. H. (1991). The estimation of an origin-destination matrix by constrained generalized least squares. *Transportation Research* 25, 13-22.
- Ben-Akiva, M. and S. Lerman (1985). *Mathematical Programs with Equilibrium Constraints*. Transportation Study. MIT Press.
- Bi, Z., P. Calami, and A. Conn (1989). An exact penalty function approach for the linear bilevel programming problem. Technical Report 167-0-310789, Department of Systems Design and Engineering, University of Waterloo, Waterloo.
- Bialas, W. F. and M. H. Karwn (1982). On two-level linear optimization. *IEEE Transaction on Automatic Control AC-27*, 211-214.
- Bialas, W. F. and M. H. Karwn (1984). Two-level linear programming. *Management Science* 30, 1004-1020.
- Bierlaire, M. and F. Crittin (2004). An efficient algorithm for real-time estimation and prediction of dynamic od tables. *Operations Research* 52(1), 116-127.
- Bierlaire, M. and P. Toint (1994). An origin-destination matrix estimator that exploits structure. *Transportation Research B* 29.
- Boyce, D., B. Ran, and L. LeBlanc (1991). Dynamic user-optimal traffic assignment: A new model and solution techniques. Presented at First Triennial Symposium.
- Brenninger-Gothe, M., J. Jornsten, and J. Lundgren (1989). Estimation of origin-destination matrix from traffic counts using multiobjective programming formulations. *Transportation Research B* 23, 257-265.
- Brézis, H. (1968). Équations et inéquations non linéaires dans les espaces vectoriales en dualité. *Annales de l'Institut Fourier* 18, 115-175.

- Burrell, J. E. (1968). Multiple route assignment and its application to capacity restraint. In *Proceedings of the 4-th International Symposium on Transportation and Traffic Theory*, University of Karlsruhe.
- Carey, M. (1987). Optimal time-varying flows on congested networks. *Operations Research* 35, 58–69.
- Casas, J. (2004). *Estimation of the Transport Demand for Real-Time Applications*. Ph. D. thesis, UPC, Barcelona.
- Cascetta, E. (1984). Estimation of trip matrices from traffic counts and survey data: a generalised least square estimator. *Transportation Research* 18B, 289–299.
- Cascetta, E. (2001). *Transportation systems engineering: theory and methods*. Kluwer Academic Publishers.
- Cascetta, E. and S. Nguyen (1988). A unified framework for estimating or updating origin/destination matrices from traffic counts. *Transportation Research B* 22, 437–455.
- Cascetta, E. and M. Postorino (2001). Fixed point approaches to the estimation of O/D matrices using traffic counts on congested networks. *Transportation Science* 35, 134–147.
- Cascetta, E. and F. Russo (1992). Calibrating travel demand models from traffic counts: statistical performances and computational aspect. In *Select Proceedings of the Sixth World Conference on Transportation Research*, Lyon.
- Cascetta, E. and F. Russo (1997). Calibrating aggregate travel demand model with traffic counts: Estimators and statistical performance. *Transportation* 24, 271–293.
- Castillo, E., A. Conejo, P. Pedregal, R. García, and N. Alguacil (2002). *Building and solving mathematical programming models in engineering and science*. New York: Pure and Applied Series (PAMS), John Wiley & Sons, Inc.
- Ceylan, H. and M. Bell (2005). Genetic algorithm solution for the stochastic equilibrium transportation networks under congestion. *Transportation Research Part B: Methodological* 39(3), 169–1858.
- Chen, Y. (1994). Bilevel programming problems: analysis, algorithms and applications. Publication 984, Centre de Recherche sur les Transports, Université de Montréal.
- Chen, Y. and M. Florian (1990). O-D demand adjustment problem with congestion: Part I. model analysis and optimality conditions. In L. Bianco and P. Toth (Eds.), *Advanced Methods in Transportation Analysis*, pp. 1–22. Berlin: Springer-Verlag.
- Chin, S., H. Hwang, and T. Pei (1994). Using neural networks to synthesize origin destination flows in a traffic circle. *Transportation Research Board, Washington, D. C.*, Paper N 940353.

- Clark, S. and D. Watling (2005). Modelling network travel time reliability under stochastic demand. *Transportation Research Part B: Methodological* 39(2), 119-140.
- Clarke, F. H. (1990). *Optimization and Nonsmooth Analysis*. Philadelphia: SIAM.
- Codina, E. and J. Barceló (2000). Adjustment of O-D trip matrices from observed volumes: an algorithmic approach based on conjugate directions. In *Proceedings of 8th EURO Working Group on Transportation*, Roma, Italy.
- Codina, E. and J. Barceló (2004). Adjustment of o-d trip matrices from observed volumes: An algorithmic approach based on conjugate directions. *European Journal of Operational Research* 155(3), 535-557.
- Codina, E., R. García, and A. Marín (2001). New algorithmic alternatives for the O-D matrix adjustment problem on traffic networks. *European Journal of Operational Research*.
- Codina, E. and L. Montero (2001). Approximation of the steepest descent direction for the od matrix adjustment problem. In *TRISTAN IV*, A ores, Portugal.
- Codina, E. and L. Montero (2003). Approximation of the steepest descent direction for the OD matrix adjustment problem. *Annals of Operations Research*.
- Consultants, I. (1990). *EMME/2 User's manual*.
- Dafermos, S. C. (1980). The traffic equilibrium and variational inequalities. *Transportation Science* 14, 42-54.
- Daganzo, C. F. and Y. Sheffi (1977). On stochastic models of traffic assignment. *Transportation Science* 11, 253-274.
- Damberg, O., J. T. Lundgren, and M. Patriksson (1996). An algorithm for the stochastic user equilibrium problem. *Transportation Research* 30B, 115-131.
- Dantzig, G. B. and P. Wolfe (1960). Decomposition principle for linear programs. *Operations Research* 8, 101-111.
- Dempe, S. (2003). Bilevel programming-a survey. pp. 1-25.
- Denault, L. (1994). Étude de deux méthodes d'ajustement de matrices origine-destination à partir des flots des véhicules observés (in French) . Report crt-991, memoria d'étudiant, Centre de Recherche sur les Transports, Université de Montréal, Canada.
- Dial, R. B. (1971). A probabilistic multipath traffic assignment model which obviates
- Doblas, J. and F. Benítez (2005). An approach to estimating and updating origin-destination matrices based upon traffic counts preserving the prior structure of a survey matrix. *Transportation Research Part B: Methodological* 39, 565-591.
- Drissi-Kaitouni, O. and J. Lundgren (1992). Bilevel origin-destination matrix estimation using a descent approach. Technical report lith-mat-r92-49, Department of Mathematics, Linköping, Institute of Technology, Linköping, Sweden.

- Erlander, S., K. Jörnsten, and J. Lundgren (1984). On the estimation of trip matrices in the case of missing and uncertain data. Technical report lith.mat-r-84-20, Department of Mathematics, Linköping University. Sweden.
- Erlander, S., S. Nguyen, and N. F. Stewart (1979). On the calibration of combined distribution-assignment model. *Transportation Research* 13B, 259-267.
- Fiacco, A. V. (1983). Sensibility analysis for nonlinear programming using penalty methods. *MP* 10, 287-311.
- Fisk, C. (1980). Some developments in equilibrium traffic assignment. *Transportation Research* 14B, 243-255.
- Fisk, C. S. (1988). On combining maximum entropy trip estimation with user optimal assignment. *Transportation Research B* 22, 69-79.
- Fisk, C. S. (1989). Trip matrix estimation from link traffic counts: The congested network case. *Transportation Research B* 23, 331-336.
- Fisk, C. S. and D. E. Boyce (1983). A note on trip matrix estimation from link traffic count data. *Transportation Research B* 17, 245-250.
- Florian, M. (1986). Nonlinear cost network models in transportation analysis. *Mathematical programming study* 26.
- Florian, M. (1990). Private communication.
- Florian, M. and Y. Chen (1993). A coordinate descent method for the bilevel O-D matrix adjustment problem. Publication, Centre de Recherche sur les Transports, Université de Montréal.
- Florian, M. and Y. Chen (1995). A coordinate descent method for the bilevel OD matrix adjustment problem. *International Transactions on Operations Research* 2, 165-179.
- Frank, M. and P. Wolfe (1956). An algorithm for quadratic programming. *Naval Research Logistics Quarterly* 3, 95-110.
- Friesz, T., F. Luque, R. Tobin, and W. B.W. (1989). Dynamic network traffic assignment considered as a continuous time optimal control problem. *Operations Research* 37, 893-901.
- Friesz, T. L., H.-J. Cho, N. Tobin, and G. Anandalingam (1992). A simulated annealing approach to the network design problem with variational inequality constraints.
- Friesz, T. L. and P. Harker (1985). Properties of the iterative optimization-equilibrium algorithm. *Civil Engineering Systems* 2, 142-154.
- Friesz, T. L., R. Tobin, H.-J. Cho, and N. Mehta (1990). Sensibility analysis based heuristic algorithms for mathematical programs with variational inequality constraints. *Mathematical Programming* 48, 265-284.

- Gale, D., H. Kuhn, and A. W. Tucker (1951). Linear programming and the theory of games. *T. C. Koopmans (ed) Activity Analysis of Production and Allocation*, 317-329.
- Gan, L., H. Yang, and S. Wong (2005). Traffic counting location and error bound in origin-destination matrix estimation problems. *Journal of Urban Planning and Development* 131(7), 524-534.
- Gao, Z., H. Sun, and L. L. Shan (2004). A continuous equilibrium network design model and algorithm for transit systems. *Transportation Research, part B* (38), 235-250.
- García, R. (2001). *Metodología para el diseño de redes de transporte y para la elaboración de algoritmos en programación matemática convexa diferenciable*. Ph. D. thesis, Escuela Técnica Superior de Ingenieros Aeronáuticos, Universidad Politécnica de Madrid.
- García, R. and A. Marín (1998). A bi-level programming approach for estimation of origin-destination matrix and calibration of parameters of a network equilibrium model with combined modes. In *Proceedings of the Conference 6th EURO Working Group on Transportation*, Gothenburg, Sweden.
- García, R. and A. Marín (2002). Parking capacity and pricing in park'n ride trips: A continuous equilibrium network design problem. *Annals of Operations Research* 116, 153-178.
- García, R., A. Marín, and M. Patriksson (2003). Column generation algorithms for nonlinear optimization, I: Convergence analysis. *Optimization* 52, 171-200.
- García, R. and D. Verastegui (2003). A column generation method for dap. Technical report, Universidad de Castilla-La Mancha.
- Ghali, M. and M. Smith (1993). Traffic assignment, traffic control and road pricing. *Proceedings of 12th International Symposium on Transportation and Traffic Theory, Elsevier Science, Amsterdam*, 147-170.
- Giannessi, F., A. Maugeri, and P. M. Pardalos (2002). *Equilibrium problems: non-smooth optimization and variational inequality models*. Dordrecht: Kluwer Academic Publishers.
- Gur, Y., M. Turnquist, M. Scheneider, L. Leblanc, and D. Kurth (1980). Estimation of an origin-destination trip table on observed link volumes and turning.
- Hall, M. D., D. Van Vliet, and L. G. Willumsen (1980). SATURN: A simulation assignment model for the evaluation of traffic management schemes. *Traffic Engineering and Control* 21, 168-176.
- Hammerslag, R. and B. Immers (1988). Estimation of trip matrices: shortcomings and possibilities for improvement. *Transportation research Record* 1203, 27-39.

- Han, A. F., R. G. Dowling, E. C. Sullivan, and A. May (1981). Deriving origin-destination information from routinely collected traffic counts. *Trip Table Synthesis for Multipath Networks, RR81-9. Institute for Transportation Studies, University of California, Berkeley 2.*
- Hand, A. F. and E. Sullivan (1983). Trip table synthesis for cbd networks: evaluation of the linkod model. *Transportation Research Record 944*, 106 112.
- Hansen, P., B. Jaumard, and G. Savard (1992). New branch-and-bound rules for linear bilevel programming. *SIAM Journal on Scientific and Statistical Computing 13*, 1194 1217.
- Hartman, P. and G. Stampacchia (1966). On some non-linear elliptic differential-functional equations. *Acta Mathematica 115*, 271 310.
- Hearn, D. H. and M. Ramana (1998). Solving congestion toll pricing models. In P. Marcotte and S. Nguyen (Eds.), *Advanced Transportation Modelling*, pp. 109 124. Massachusetts: Kluwer Academic Publishers.
- Hearn, D. W., S. Lawphongpanich, and J. A. Ventura (1985). Finiteness in restricted simplicial decomposition. *Operations Research Letters 4*, 125 130.
- Hearn, D. W., S. Lawphongpanich, and J. A. Ventura (1987). Restricted simplicial decomposition: Computation and extensions. *Mathematical Programming Study 31*, 99 118.
- Ho, J. (1990). Solving the dynamic traffic assignment problem on a hypercube multicomputer. *Transportation Research 24B*, 443 451.
- Hogan, W. (1973). Point-to-set maps in mathematical programming. *SIAM Review 15*, 591 603.
- Hohenbalken, B. v. (1977). Simplicial decomposition in nonlinear programming algorithms. *Mathematical Programming 13*, 49 68.
- Holloway, C. A. (1974). An extension of the Frank and Wolfe method of feasible directions. *Mathematical Programming 6*, 14 27.
- Janson, B.Ñ. (1991). Dynamic traffic assignment for urban road networks. *Transportation Research 25B*, 143 161.
- Janson, B.Ñ. (1993). Dynamic traffic assignment with arrival time costs. *Proceedings of 12th International Symposium on Transportation and Traffic Theory, Elsevier Science, Amsterdam*, 127 146.
- Jaumard, B., G. Savard, and X. Xiong (1995). An exact algorithm for convex bilevel programming. *Mathematical Programming 68*, 137 152.
- Jeroslow, R. G. (1988). The polynomial hierarchy and a simple model for competitive analysis. *Mathematical Programming 32*, 273 284.
- Jörnsten, K. and S. Nguyen (1979). On the estimation of a trip matrix from network data (revisado en abril de 1983). Technical Report 153, Centre de Recherche sur les Transports de Montréal.

- Júdice, J. and A. M. Faustino (1992). A sequential LCP method for bilevel linear programming. *Annals of Operations Research* 34, 89 106.
- Kawakami, H., H. Lu, and Y. Hirobata (1992). Estimation of origin-destination matrices from link traffic counts considering the interaction of the traffic modes. *Papers in Regional Science* 71.
- Kim, H., S. Baek, and Y. Lim (2001). Origin-destination matrices estimated with a genetic algorithm from link traffic counts. *Transportation Research Record* (1771), 156 163.
- Kolstad, C. D. and L. Lasdon (1986). Derivative evaluation and computational experience with large bi-level mathematical programs. Faculty working paper, College of Commerce and Business Administration, University of Illinois at Urbana-Champaign.
- Larsson, T., M. Patriksson, and A.-B. Strömberg (1997). Ergodic convergence in subgradient optimization. Technical report, Department of Mathematics, Linköping Institute of Technology, Linköping Institute of Technology, Linköping, Sweden.
- Larsson, T. and M. Patriksson (1992). Simplicial decomposition with disaggregated representation for the traffic assignment problem. *Transportation Science* 26, 4 17.
- Larsson, T. and M. Patriksson (1998). Side constrained traffic equilibrium models traffic management through link tolls. In P. Marcotte and S. Nguyen (Eds.), *Advanced Transportation Modelling*, pp. 125 151. Massachusetts: Kluwer Academic Publishers.
- Larsson, T., M. Patriksson, and C. Rydergren (1997). Applications of simplicial decomposition with nonlinear column generation to nonlinear network flows. In P. M. Pardalos, W. W. Hager, and D. W. Hearn (Eds.), *Network Optimization*, Number 450 in Lecture Notes in Economics and Mathematical Systems, pp. 346 373. Berlin: Springer-Verlag.
- Larsson, T., M. Patriksson, and C. Rydergren (2003, October). Inverse nonlinear multicommodity flow optimization by column generation. *Optimization Methods and Software* 18(5), 601 613.
- Larsson, T., M. Patriksson, and A. B. Strömberg (1999). Ergodic, primal convergence in dual subgradient schemes for convex programming. *Mathematical Programming* 86, 283 312.
- LeBlanc, L. and K. Farhangian (1982). Selection of a trip table which reproduces observed link flows. *Transportation Research B* 16.
- LeBlanc, L., E. K. Morlok, and W. P. Pierskalla (1975). An efficient approach to solving the road network equilibrium traffic assignment problem. *Transportation Research B* 9, 309 318.