DISEÑO DE REDES DE TRANSPORTE CON CONGESTION

J. Enrique Fernández
Departamento Ingeniería de Transporte
Pontificia Universidad Católica de Chile

Resumen

El diseño de redes es un tema de fundamental importancia para la gestión de sistemas de transporte, La especificación tradicionalmente mas tratada consiste en escoger la mejor red de vías de infraestructura, para satisfacer un conjunto de demandas de transporte en el espacio, representadas por una matriz O-D de viajes en automóvil, pero también pueden utilizarse especificaciones que permitan determinar las características óptimas de operación de una red de servicios de transporte público, que utilizan una red de infraestructura dada, o ambos problemas simultáneamente. La formulación matemática de estos problemas conduce sin embargo a modelos cuyo tratamiento y solución acarrea importantes dificultades teóricas y prácticas.

En este trabajo se plantean las formulaciones matemáticas de los problemas de diseño de redes y se analizan sus características para el caso de variables de decisión continuas. A continuación se describen y analizan los métodos más importantes que han sido propuestos para resolver el problema en casos reales.

1. Introducción

El diseño de redes de transporte está relacionado con dos problemas de gran relevancia práctica: la determinación de las características operacionales óptimas de una red de infraestructura, para satisfacer las demandas planteadas a través de una matriz O-D de viajes en automóvil (PDI) y la determinación de las características operacionales óptimas de una red de rutas de transporte público, a fin de satisfacer las demandas por tal servicio planteadas a través de la correspondiente matriz O-D (PDR).

En el caso del PDI las variables de decisión son las capacidades de los arcos (calles o vías) de la red de infraestructura analizada y en el caso del PDR son las frecuencias de los servicios de transporte público (líneas de buses, aviones, etc.) que operan en cada ruta de la red. Estas variables, fundamentalmente en el caso del PDI, son de carácter discreto, ya que la unidad física relevante de definición de capacidad es la "pista" y una vía tiene siempre un número discreto de pistas. Sin embargo, la utilización de variables discretas en la especificación matemática del PDI, conduce a problemas de carácter combinatorio cuya solución es prácticamente imposible para la mayoría de los casos de tamaño real (ver: Fernández, 1982; Gutierrez, 1982), lo que es aun más válido para el caso del PDR que posee un espacio de soluciones de mayor dimensionalidad. Por este motivo, durante el último tiempo los esfuerzos de la mayoría de los investigadores se han concentrado en el uso de formulaciones que utilizan variables de decisión continuas (Steenbrink, 1974; Abdulaal y Leblanc, 1979 y 1984; Marcotte, 1983a y 1983b; Harker y Friesz, 1984; Fernández, 1984 y 1985).

La utilización de variables de decisión continuas constituye una aproximación absolutamente razonable en el caso del PDR, principalmente cuando se analizan sistemas que presentan servicios con altas frecuencias (Ej.: redes de servicios de buses urbanos en países en desarrollo). Para el caso del PDI, Elton (1983) ha demostrado que es posible utilizar una formulación continua como primera etapa en la obtención de una solución discreta óptima, lo que permite resolver problemas de tamaño real en solo una pequeña fracción del tiempo computacional necesario para resolver directamente una formulación discreta.

La bondad de las soluciones a los problemas de diseño de redes se establece generalmente evaluando una función objetivo, que considera los costos totales de viaje de los usuarios y los costos de construcción de la infraestructura, en el caso del PDI, y los costos totales de viaje de los usuarios y costos totales de operación de los servicios de transporte público ofrecidos, en el caso del PDR. En todos los casos considerados en este trabajo se supone que las vías de transporte están sujetas al fenómeno de congestión, lo que hace que tanto el costo de viaje de los usuarios del sistema, como el costo de operación de los oferentes de servicios de transporte público, aumente al aumentar el número de vehículos que utilizan la vía. Para ambos problemas la función objetivo está por lo tanto constituída por dos tipos de funciones, una creciente con las variables de decisión (costos de construcción u operación) y otra decreciente con dichas variables (costos de viaje de los usuarios del sistema). Una solución es óptima si tiene asociado el menor costo total posible, al mismo tiempo que se respetan las restricciones del problema.

Las restricciones normalmente consideradas incluyen: continuidad y consistencia de flujos, comportamiento espontáneo de los usuarios en su utilización del sistema y no negatividad de las variables de decisión.

La formulación matemática del problema de diseño de redes cae dentro de la categoría de problemas de programación multinivel, cuyo tratamiento ha suscitado gran interés últimamente. Ella consiste en una rama de la programación matemática que puede ser vista, ya sea como una generalización de problemas del tipo mini-max, o como una clase particular de juegos tipo Stakelberg (ver: Marcotte, 1983b Fisk, 1984).

Lo que resta de este trabajo está estructurado como sigue: a continuación planteamos la formulación matemática del problema y analizamos sus características teóricas; finalmente presentamos y analizamos las características de los principales métodos de solución propuestos.

2. Planteamiento del Problema

2.1. Problema de diseño de redes de infraestructura (PDI)

En el planteamiento del PDI usaremos la siguiente notación:

$G(N, A)$ = Grafo dirigido que representa a la red de transporte a analizar y en la que N y A corresponden a los conjuntos de nodos y arcos respectivamente.

W = Conjunto de todos los pares origen-destino (O-D) que sirve la red.

w = Indice general correspondiente a un par O-D ($w \in W$)

P = Conjunto de todas las rutas que existen sobre la red para unir pares O-D.

P_w = Conjunto de las rutas que unen el par w.

p = Indice general correspondiente a una ruta ($p \in P$).

a = Indice general correspondiente a un arco de la red ($a \in A$).

f_a = Flujo sobre el arco a.

h_p = Flujo sobre la ruta p.

T_w = Demanda por viajes entre el par w.

u_a = Inversión o mejoramiento propuesto para el arco a.

β_a = Costo constante unitario de inversión, por unidad de capacidad, sobre el arco a.

$c_a(f_a, u_a)$ = Costo medio de viaje sobre el arco a, que se supone es una función diferenciable en segundo orden con respecto a las variables f_a y u_a.

C_p = Costo medio de viaje sobre la ruta p.

Definiremos también los siguientes vectores:

$$f = \{f_a \mid a \varepsilon A\}\ ,\quad c = \{c_a \mid a \varepsilon A\}\ ,\qquad u = \{u_a \mid a \varepsilon A\}$$

y la matriz arco - ruta $\{\delta_{ap}\}$ que especifica la topología de la red y cuyos componentes se definen como:

$$\delta_{ap} = \begin{cases} 1, & \text{si el arco } a \ \varepsilon\ A, \text{ es parte de la ruta } p \ \varepsilon\ P \\ 0, & \text{en el resto de los casos} \end{cases}$$

Hacemos notar además que se cumplen las siguientes relaciones entre algunas de las variables definidas:

- Restricciones de Demanda : $T_w = \sum_{p \varepsilon P_w} h_p,\ \forall\ w\ \varepsilon\ W,$ (1)

- Consistencia de Flujos : $f_a = \sum_{p \varepsilon P} \delta_{ap}\ h_p,\ \forall\ a\ \varepsilon\ A,$ (2)

- Consistencia de Costos : $C_p = \sum_{a \varepsilon A} \delta_{ap}\ c_a,\ \forall\ p\ \varepsilon\ P,$ (3)

En el presente trabajo nos concentraremos en el caso en que todas las variables u_a representan modificaciones de la capacidad de arcos existente (no se consideran arcos nuevos) y supondremos que las funciones de costos de inversión son lineales en la variable capacidad.

Supondremos también, que las funciones de costo c_a son de la siguiente forma:

$$c_a = A_a + B_a\ [f_a \mid u_a]^n \tag{4}$$

en que A_a es el costo de operación de flujo libre (sin congestión).

Dada la notación anterior, podemos formular como sigue el problema de diseño de redes de infraestructura:

$$\left.\begin{array}{l} \underset{u}{\text{Min}}\ :\ \sum_{a \varepsilon A} [c_a(f^*_a(u),\ u_a)\ f^*_a(u) + \lambda\ \beta_a\ u_a] \\ \\ \text{s.a.}\ :\ u_a \geq 0,\ \forall\ a\ \varepsilon\ A^{(1)} \end{array}\right\}\ \text{(PDI)}$$

en que $f^*(u) = \{f^*_a(u) \mid a\ \varepsilon\ A\}$ corresponde a un equilibrio de usuarios sobre la red, (Fernández y Friesz, 1983) y por lo tanto se obtiene como solución del siguiente problema:

(1) La restricción $u_a \geq 0$, puede cambiarse sin inconvenientes por $u_a \geq k_a$, si existe una capacidad original que no puede ser disminuída.

$$\left.\begin{aligned} &\text{Min} : \sum_a \int_o^{f_a} c_a(x, u_a)\, dx \\ &\text{s.a.}: \sum_{p \varepsilon P_w} h_p = T_w, \quad \forall\, w \,\varepsilon\, W \\ &\qquad f_a = \sum_{p \varepsilon P} \delta_{ap}\, h_p, \quad \forall\, a \,\varepsilon\, A; \\ &\qquad h_p \geq 0, \qquad \forall\, p \,\varepsilon\, P \end{aligned}\right\} \quad \text{(PEU)}$$

El parámetro λ utilizado en la formulación del PDI, corresponde al inverso del valor del tiempo de viaje, y se utiliza cuando c_a representa el tiempo de viaje sobre el arco a. Si c_a representa costos de operación en unidades monetarias, λ debe tomar un valor igual a la unidad.

El problema así planteado es un problema de programación de dos niveles. Los agentes decisores a cada nivel actúan en forma jerárquica: en el nivel superior, la autoridad planificadora toma decisiones que involucran especificaciones del vector, u, de tal forma de minimizar el costo social total asociado al sistema, pero tomando en cuenta las reacciones de los usuarios que se encuentran en el segundo nivel jerárquico (f*(u)). A su vez, los usarios toman decisiones de uso del sistema, que determinan el valor del vector f, de tal forma de minimizar sus costos de operación individuales, pero restringidos por las decisiones tomadas en el nivel superior respecto del vector u.

Recientemente, algunos autores han investigado problemas de este tipo en que están involucrados varios niveles jerárquicos; en cada nivel los agentes están restringidos por las decisiones del nivel superior y maximizan sus beneficios tomando en consideración las reacciones de los niveles inferiores (ver Bard y Falk, 1982).

2.2. PDI con comportamiento óptimo de los usuarios del sistema

La función f*(u) representa el comportamiento espontáneo de los usuarios de una red de vías, obtenido como resultado de la superposición de decisiones individuales en que cada usuario trata de minimizar independientemente su costo de operación. Dado que la operación de las vías de la red están sujetas a externalidades de congestión, tal comportamiento no conduce a un óptimo social del sistema, en que el costo total social de operación sea mínimo, para un valor del vector u, ya que el equilibrio f*(u) se obtiene como consecuencia de la consideración de los costos medios individuales de operación sobre las posibles rutas que unan pares O-D sobre la red y no los costos marginales sociales (ver, Fernández y Friesz, 1983).

Sin embargo, si suponemos que los usuarios se comportan en forma socialmente óptima, en su elección de rutas de viaje sobre la red, podemos considerar el vector f, además del vector u, como variables de decisión del PDI. En tal caso desaparece la restricción f*(u), representada por el problema PEU, y obtenemos la la siguiente formulación notablemente más sencilla:

$$\left.\begin{aligned} &\underset{(u,f)}{\text{Min}} : \sum_{a\varepsilon A} [c_a(f_a, u_a)\, f_a + \lambda\, \beta_a\, u_a] \\ &\text{s.a.: } (f_a, u_a) \geq 0, \quad \forall\, a \in A \;;\quad f \in \Phi \end{aligned}\right\} \quad (\text{PDI}')$$

en que Φ es el espacio definido por el conjunto de restricciones del PEU.

Este problema puede ser simplificado aun más si obtenemos la relación óptima entre las variables f_a y u_a, $u_a(f_a)$ para cada arco de la red y la reemplazamos en la función objetivo (Marcotte, 1983b). $u_a(f_a)$ se obtiene resolviendo:

$$\underset{u_a}{\text{Min}} : \; c_a(f_a, u_a) \cdot f_a + \lambda\, \beta_a\, u_a \tag{5}$$

$$\text{s.a.: } u_a \geq 0, \quad \forall\, a \in A$$

que será un problema convexo con solución única siempre que c_a sea convexa. En nuestro caso y dado que estamos suponiendo funciones c_a de la forma expresada en (4), la solución de (5) conduce a:

$$u_a(f_a) = \left(\frac{n\, B_a}{\lambda\, \beta_a}\right)^{1/(n+1)} f_a \tag{6}$$

Introduciendo la relación $u_a(f_a)$ en la función objetivo del PDI' este se transforma en general en:

$$\underset{f}{\text{Min}} : \sum_{a\varepsilon A} [c_a(f_a, u_a(f_a)) \cdot f_a + \beta_a\, u_a(f_a)] \tag{7}$$

$$\text{s.a.: } f_a \geq 0, \quad \forall\, a \in A \;;\quad f \in \Phi$$

y utilizando la expresión de $u_a(f_a)$ dada en (6) para reemplazar en (7) obtenemos:

$$\underset{f}{\text{Min}} : \sum_{a\varepsilon A} (A_a + \gamma_a)\, f_a \tag{8}$$

$$\text{s.a.: } f_a \geq 0 \,, \ \forall\, a\, \varepsilon\, A \,;\ f\, \varepsilon\, \Phi$$

$$\text{con: } \gamma_a = (\lambda\, \beta_a\, \alpha_a^{-1} + B_a\, \alpha_a^{n})$$

$$\alpha_a = (n\, B_a / \lambda\, \beta_a)^{-(1/n)}$$

y esta formulación es equivalente a un problema de asignación con costos de operación constante (sin congestión). Para su solución solo se requiere calcular las rutas mínimas sobre la red original suponiendo costos de operación constantes iguales a $(A_a + \gamma_a)$.

Por lo tanto, en este caso, el PDI se transforma en un problema de solución prácticamente trivial. El problema (8) es aun más fácil de resolver que un problema estandard de asignación de flujos sobre la red original. Es necesario aclarar que esta notable característica es una consecuencia de haber supuesto que las funciones de costo de capacidad son lineales. Si en vez de ello suponemos que estas funciones son de la forma:

$$g_a(u_a) = \beta_a\, u_a^{m},\ m \neq 1 \tag{9}$$

la expresión de $u_a(f_a)$ resultante de resolver el problema (5) resulta ser en nuestro caso:

$$u_a(f_a) = \left(\frac{n\, B_a}{m\, \lambda\, \beta_a} \right)^{1/(n+m)} f_a^{(n+1)/(n+m)} \tag{10}$$

e introduciendo (10) en (7) obtenemos:

$$\text{Min} : \sum_{a\varepsilon A} [A_a + \gamma_a\, f_a^{n(m-1)/(n+m)}]\, f_a \tag{11}$$

$$\text{s.a.: } f_a \geq 0 \,, \ \forall\, a\, \varepsilon\, A \,;\ f\, \varepsilon\, \Phi$$

$$\text{con : } \gamma_a = (\lambda\, \beta_a\, \alpha^{-m} + B_a\, \alpha^{n})$$

$$\alpha_a = \left(\frac{n\, B_a}{m\, \lambda\, \beta_a} \right)^{-1/(n+m)}$$

Esta última formulación es equivalente a un problema estandard de asignación de flujos sobre la red original, con funciones de costo modificadas, el que puede ser resuelto eficientemente utilizando algoritmos del tipo Frank-Wolfe (Leblanc et al, 1975) o Partan (Florian et al, 1985). Es fácil ver que (11) se transforma en (8) para m = 1.

Por lo tanto, podemos conlcuir que si suponemos que los usuarios utilizan la red de transporte en forma socialmente óptima, el PDI se transforma en un problema que puede ser resuelto en forma eficiente utilizando algoritmos ampliamente conocidos.

Más aun, Marcotte (1983b) demostró que si las funciones c_a y g_a son tales que:

i) las g_a son lineales y su valor proporcional al largo del arco y su capacidad:

$$g_a(u_a) = \ell\, d_a\, u_a$$

en que ℓ es una constante y d_a representa la longitud física del arco a.

ii) las c_a son homogéneas a través de la red y proporcionales a la longitud del arco:

$$c_a(f_a, u_a) = d_a\, c(f_a, u_a)$$

entonces las soluciones del PDI y el PDI' son coincidentes

2.3.Problema de diseño de rutas de transporte público (PDR)

En el caso del PDR consideraremos una red G(N, A) que sirve como infraestructura común para la operación de transporte privado y transporte público. Si tomamos como ejemplo el sistema de rutas de transporte público de una ciudad, tendremos que la mayoría de los arcos de la red serán simultáneamente utilizados por buses y automóviles y por lo tanto los costos de operación de ambos medios estarán interrelacionados: un aumento del flujo de automóviles afectará al costo de operación de los buses y viceversa. Por lo tanto, tendremos dos funciones de costo de operación para cada arco, cada una de ellas dependiente del flujo total:

$$c_a(f_a, d) = A_a + B_a\,[f_a + \mu \sum_{r \in R} \delta_{ar}\, d_r]^n \qquad (12)$$

$$\bar{c}_a(f_a, d) = \bar{A}_a + \bar{B}_a\,[f_a + \mu \sum_{r \in R} \delta_{ar}\, d_r]^n \qquad (13)$$

en que $\bar{c}_a$ es la función de tiempo de viaje de los vehículos de transporte público sobre el arco a y c_a la función de costos de operación de transporte privado, además:

R = Conjunto de rutas de transporte público en la red G.

d_r = Frecuencia de operación de la línea r de transporte público (que opera la ruta $r \in R$).

d = $\{d_r / r \in R\}$, vector de frecuencias

$\{\delta_{ar}\}$ = Matriz arco-ruta de transporte público cuyos componentes se definen como:

$$\delta_{ar} = \begin{cases} 1, & \text{si la ruta } r \text{ de transporte público pasa por el arco } a \\ 0, & \text{en el resto de los casos.} \end{cases}$$

μ = Factor de equivalencia de un vehículo de transporte público (en vehículos equivalentes).

Tendremos también que los costos totales de operación sobre un arco, para cada medio serán:

$$C_a(f_a, d) = c_a(f_a, d)\, f_a \tag{14}$$

$$\bar{C}_{ar}(f_a, f_{ar}, d) = \bar{c}_a(f_a, d)\, \bar{f}_{ar} \tag{15}$$

en que:

C_a = Costo total de operación para el transporte privado en el arco a.

$\bar{C}_{ar}$ = Tiempo total de viaje de los usuarios de la línea r de transporte público, en el arco a.

$\bar{f}_{ar}$ = Número de pasajeros de transporte público que viajan en la línea r a través del arco a.

Adicionalmente definiremos los conjuntos : $\bar{L}$, que contiene todos los arcos de la red $G(N, A)$ que son utilizados por recorridos de transporte público; L' , que contienen el conjunto de arcos por los que no circula transporte público y L_r que es el conjunto de arcos utilizados por la línea r, con:

$$\bar{L} \cup L' = A \quad , \quad \bar{L} \cap L' = \phi$$

A diferencia de los usuarios de transporte privado, que solo incurren en un gasto de tiempo de viaje sobre el vehículo mismo, los usuarios de transporte público tienen que incurrir además en un tiempo de acceso, gastado en recorrer el espacio que separa su lugar de origen y el nodo de la red a través del cual acceden a los servicios de transporte más la distancia entre el nodo de egreso y su destino final, y un tiempo de espera hasta que llegue un vehículo que pueda utilizar para viajar a su destino. Además, en general, el conjunto de usuarios pertenecientes a una misma zona geográfica tendrá varios nodos alternativos a través de los cuales acceder a la red de servicios de transporte público. Por lo tanto denominaremos:

$\overline{O}_w$ = Conjunto de nodos a través de los cuales se puede acceder a la red de transporte público, para viajar entre el par w (par O-D).

$\overline{D}_w$ = Conjunto de nodos a través de los cuales se puede egresar de la red de transporte público, al viajar entre el par w.

V_{ij}^w = Proporción, del total de usuarios de transporte público que viajan entre el par w, que utilizan el nodo i como acceso a la red y el nodo j como egreso.

$\overline{t}_{ij}^w$ = Tiempo de acceso para los usuarios que viajan entre el par w utilizando i como acceso y j como egreso.

Por último, un fenómeno importante a considerar en la operación de transporte público(Ej.: servicios de buses en ciudades de países en desarrollo) es el de las líneas comunes (De Cea, 1984). Es así como distintas líneas utilizan un mismo conjunto de arcos o segmento común de ruta, haciendo que un usuario cuyos nodos de acceso y egreso a la red estén sobre dicho segmento, tenga varias líneas o servicios alternativos que puede utilizar. Por lo tanto, incluíremos la siguiente notación adicional:

$\overline{f}_{ij}^r$ = Número de usuarios de transporte público que acceden a la línea r en el nodo i y egresan en el nodo j.

El número de variables de este tipo a definir para cada línea r depende del número de sergmentos$(i, j)_s^r$ sobre los cuales existen distintos conjuntos de líneas comunes $\overline{\ell}_{ij}^r$.

Finalmente, definiremos como $G_r(d_r)$ la función de costo de operación de la línea r de transporte público.

Utilizando toda la notación definida podemos proceder a formular el PDR como sigue:

$$\begin{aligned}\min_{d} :\ & \sum_{a\varepsilon \overline{L}} C_a(f_a^*, d) + \sum_{a\varepsilon L'} C_a(f_a^*) + \\ & \sum_{r\varepsilon R} \sum_{a\varepsilon L_r} \overline{C}_{ar}(f_a^*, \overline{f}_{ar}^*, d) + \\ & \sum_{w} T_w \Big[\sum_{i\varepsilon \overline{O}_w} \sum_{j\varepsilon \overline{D}_w} V_{ij}^w\ \overline{t}_{ij}^w \Big] + \\ & \sum_{r\varepsilon R} \sum_{(i,j)_s^r} \Big[\overline{f}_{ij}^r /2 \sum_{\overline{\ell}_{ij}^r} d_r \Big] + \lambda \sum_{r\varepsilon R} G_r(d_r)\end{aligned} \tag{16}$$

$$\text{s.a.}: d_r \geq 0, \quad \forall r \in R, \quad f \in \Phi$$

$$F^* = F(d) \text{ (flujos de equilibrio multimodal)}$$

$$F = (f_a / a \in A;\ f_{ar} / a \in A,\ r \in R;\ V_{ij}^{w} / (i, j) \in (i, j)_s^r,\ w \in W)$$

$$V_{ij}^{w} = V_{ij}^{w}(d)$$

La función objetivo consta de varios términos distintos: los dos primeros representan el costo de viaje de los usuarios de transporte privado, el tercero los tiempos totales de viaje sobre el vehículo para los usuarios de transporte público, el cuarto el tiempo total de acceso y el quinto el tiempo total de espera, incurridos por los usuarios de transporte público y el término final (sexto) representa el costo total de operación de todas las líneas o servicios de transporte público.

Al igual que en el caso del PDI, el PDR así planteado es también un problema de programación en dos niveles. En el nivel superior, una autoridad central toma decisiones acerca de la especificación del vector d, de tal forma de minimizar el costo social total asociado con el sistema, pero tomando en cuenta las reacciones de los usuarios de transporte privado y transporte público que se encuentran en el nivel jerárquico inferior. A su vez, los usuarios toman decisiones de uso del sistema que determinan el valor del vector F, de tal forma de minimizar sus costos de operación individuales, pero restringidos por las especificaciones del vector d decididas en el nivel superior.

Es importante notar que el PDR no es un problema convexo, dado que la función F*(d) no es convexa. Sin embargo, si suponemos F dado y constante, podemos definir tantos problemas unidimensionales en función de d_r como líneas de transporte público existan. Para cada uno de estos problemas se supone que todas las frecuencias son fijas a excepción de una de ellas.

Los problemas unidimensionales así definidos son convexos ya que tanto las funciones C_a como $\bar{C}_{ar}$ son convexas en d_r para F dado y las funciones de tiempo de espera y costos de operación G_r son también convexas en d_r para F dado. Por último, la función de tiempo de acceso es constante para F dado, ya que los flujos V_{ij}^{w} son dados y el tiempo de acceso $\underline{t}_{ij}^{w}$ es independiente de las frecuencias d_r; por lo tanto, el término correspondiente de la función objetivo puede ser eliminado de consideración en el problema unidimensional en función de d_r.

Esta convexidad de la función objetivo en términos de cada uno de los d_r es una propiedad importante para la solución del PDR, dado que comunmente los algoritmos de solución incluyen dentro de su operación búsquedas unidimensionales, que tendrán solución única si es que la función unidimensional involucrada es convexa.

3. Principales Métodos de Solución

Los métodos o enfoques de solución más importantes que han sido propuestos para resolver los problemas de diseño de redes son los siguientes:

3.1. Algoritmo de Hooke and Jeeves (H-J)

Este método consiste en la utilización del algoritmo de Hooke and Jeeves (1961). Dicho algoritmo es suficientemente robusto como para no requerir convexidad, ni una expresión explícita de las derivadas de la función objetivo con respecto a las variables de decisión. Su aplicación sólo requiere que la función analizada sea continua y evaluable para cualquier valor factible de las variables y fue originalmente propuesta por Abdulaal y Leblanc (1979) para la solución del PDI.

El algoritmo de H-J trabaja sobre la base de la repetición de un procedimiento de dos fases: En la primera, partiendo de una solución factible x cualquiera que sirve como punto base, se realiza un proceso de búsqueda exploratoria a través de cada una de las coordenadas del espacio de soluciones, que permita definir una "buena" dirección local de descenso, (reducción del valor de la función obejtivo). Para ello, se toma una de las variables x_i 1/ se aumenta su valor en una cantidad δ pre-establecida y se evalúa la función objetivo en el nuevo punto $\bar{x}$ así definido. Si el valor obtenido resulta menor que el correspondiente al punto base, se acepta $\bar{x}$ y se pasa a examinar la siguiente variable x_i. Si por el contrario, la función objetivo resulta tener un mayor valor en $\bar{x}$, se disminuye el valor original de x_i en δ y se vuelve a evaluar la función objetivo. Si resulta menor que en el punto de partida, se acepta el nuevo punto y se continua con la siguiente coordenada. En caso de que el valor de la función objetivo resulte mayor, se vuelve al punto base y se pasa a examinar la siguiente coordenada. Este proceso se repite hasta completar el número total de las variables involucradas en el vector . Si ninguna de las búsquedas exploratorias, con respecto a todas las variables contenidas en el vector x, logra encontrar un nuevo punto que disminuya el valor de la función objetivo, se reduce el valor de δ en una cantidad pre-establecida y se repite todo el proceso de nuevo. El algoritmo se detiene cuando el valor de δ decrece más allá de un valor de "tolerancia" predefinido.

Cuando la primera fase ha dado como resultado un nuevo punto o solución factible, en el cual la función objetivo tiene un valor menor que en el punto base anterior, éstos dos puntos definen una dirección de avance en el espacio X. La longitud de avance se determina por la distancia entre los dos puntos, multiplicada por un parámetro α .

Las características de convergencia del método están gobernadas fundamentalmente por los valores de los dos parámetros α y δ los que deben determinarse en forma númerica para cada tipo de problema a tratar. En el caso del PDI, se ha encontrado que una buena combinación de valores es:

1/ Supongamos que x representa el vector de las variables de decisión. En el caso del PDI tendremos $x \equiv u$ y en el caso del PDR $x \equiv d$

$\alpha = 1$ y $\delta = 0,5$ (ver Elton, 1983) cuando g_a es del tipo $g_a = \beta \, |u_a|^n$, en que el exponente n determina las características de rendiemientos a escala en la cosntrucción de infraestructura (en el presente trabajo estamos suponiendo $n = 1$).

Está claro que no puede asegurarse que el método obtenga un óptimo global a menos que el problema a resolver fuera convexo, lo que no ocurre en el caso del PDI ni del PDR. Sin embargo, dada la forma en que funciona, el método permite explorar cualquier función continua no convexa por extraña que sea su forma. A fin de mejorar las posibilidades de obtener un óptimo, la aplicación del método puede repetirse usando distintos puntos de partida escogidos aleatoriamente.

Como acabamos de ver, cada iteración del método de H-J requiere de una búsqueda exploratoria y un avance en la dirección escogida. Si suponemos que en la primera fase, para la mitad de las variables x_i basta explorar en un sólo sentido del eje correspondiente y para la otra mitad es necesario moverse en ambos sentidos ($x_i + \delta$ y $x_i - \delta$), lo que podría considerarse un caso promedio, serán necesarias $1,5n$ evaluaciones de la función objetivo en esta fase, en que n es el número de variables del problema. Por lo tanto, si se realizan N iteraciones del método, se requerirán en promedio $N(1,5n + 1)$ evaluaciones, dado que una evaluación adicional es necesaria en la segunda fase de avance (en el peor de los casos serían $N(2n + 1)$. Para evaluar la función objetivo del PDI es necesario realizar una asignación de equilibrio a la red, proceso que es computacionalmente caro, por lo que no resulta muy atractivo que el número de evaluaciones dependa de n. Sin embargo, Abdulaal y Leblanc (1979) propusieron una modificación del algoritmo que ha dado muy buena resultados prácticos. Ella está basada en la observación de que los flujos de equilibrio $f^*(u)$ ó $F^*(d)$ no cambian significativamente cuando sólo una de las componentes de u ó d es levemente modificada. Por lo tanto, no es necesario calcular nuevos flujos de equilibrio, para cada modificación de las componentes de x durante la fase exploratoria, sino que basta con utilizar, para las evaluaciones de la función objetivo, los mismos flujos obtenidos para el punto base de la búsqueda. Así, cada iteración del método requiere sólo una asignación de flujos de equilibrio, necesitándose en total sólo N asignaciones, independientemente del número de variables.

Es importante recalcar que la gran ventaja del método de H-J es que no exige ningún atributo especial a la función objetivo del problema a resolver, salvo continuidad. Además este algoritmo pertenece a la familia de métodos de coordenadas que son muy sencillos de implementar ya que no requieren el cálculo de derivadas o Hessianas de la función objetivo (lo que en nuestro caso además no es posible dado que no conocemos la expresión de las funciones $f^*(u)$ o $F^*(d)$).

Sin embargo, estas mismas características pueden constituirse en un problema que afecte a la eficiencia computacional del algoritmo, ya que este no hace uso de niguna característica de regularidad que la función objetivo pueda presentar. En especial hemos visto que la función objetivo del PDR es convexa en cada una de las variables d_r para F dado y algo similar ocurre en el caso del PDI, cuya función objetivo es convexa en cada uno de los u_a para f dado.

Esta propiedad es utilizada por el método de optimización- equilibrio que describimos a continuación.

3.2. Método de descomposición: optimización-equilibrio

Este algoritmo fue originalmente propuesto por Steenbrink, (1974) para la solución del PDI y consiste en un procedimiento heurístico de dos fases: En la primera, se supone que el vector u es fijo y se resuelve el PEU para encontrar f*(u). La segunda fase supone a su vez que el vector f es fijo, e igual al conjunto de flujos de equilibrio determinados en la fase anterior, y procede a calcular los correspondientes valores óptimos de las componentes del vector u. Dado que f es fijo, esto último puede realizarse independientemente para cada arco de la red. Por lo tanto, el algoritmo consiste en una secuencia de asignaciones de equilibrio a la red seguidas por minimizaciones unidimensionales, para todos los arcos considerados en el vector u. Las dos fases del proceso se repiten hasta que se satisfaga algún criterio de convergencia pre-definido.

Dado un valor fijo del vector u, la primera fase, que calcula los flujos de equilibrio f*(u), consiste en resolver el PEU. Esto puede realizarse mediante la aplicación del algoritmo de Frank-Wolfe, técnica que es hoy día estandard para la solución de este problema (ver, Fernández y Friesz, 1983).

Por otra parte, la segunda fase, que supone que el vector f es constante, consiste en la solución de una serie de problemas de minimización unidimensionales e independientes entre sí.

Cada uno de estos problemas tiene la forma especificada en (5) cuya solución, como vimos en la sección 2.2. conduce a la expresión analítica explicita de $u_a(f_a)$ presentada en (6).

Esto permite que la segunda fase sea computacionalmente muy eficiente ya que se reduce a evaluar una serie de expresiones del tipo (6), en las que todos los términos son constantes y f_a corresponde al valor obtenido como resultado de la primera fase.

Si la función $g_a(u_a)$ es no lineal, como en (9), $u_a(f_a)$ tiene una expresión como la dada en (10).

Harker y Friesz (1984) caracterizan este algoritmo como un juego no cooperativo del tipo Cournot-Nash, entre los usuarios y los planificadores del sistema de transporte. En dicho juego, cada jugador actúa con una visión miope del comportamiento de su oponente, tratando de maximizar su utilidad individual bajo el supuesto de que sus acciones no producirán ninguna reacción en su contricante. Este juego alcanza una solución de equilibrio cuando ninguno de los participantes puede aumentar su utilidad mediante un cambio unilateral de estrategia. En nuestro caso, en cada iteración del algoritmo, los planificadores no están tomando en cuenta que los usuarios reaccionarán con nuevos flujos de equilibrio ante las modificaciones introducidas en el vector u. Harker y Friesz, también plantean que la solución exacta al PDI corresponderá más bien a la de un juego del tipo Stackelberg y en consecuencia argumentan que el método de descomposición, optimización-equilibrio, no conducirá a la solución deseada del PDI.

Por su parte Fisk (1984) caracteriza el método de descomposición como una solución iterativa al problema del "lider y el seguidor". En nuestro caso, los planificadores representarían al "lider", que optimiza los valores de u dados ciertos valores de f y los usuarios serían los "seguidores", que reacomodan sus flujos como consecuencia de las decisiones tomadas por los planificadores. Fisk hace notar también que el método de descomposición tiene la forma del algoritmo de aproximaciones sucesivas usado para resolver el problema de punto fijo y = G(y) (ver Ortega y Rheinboldt, 1970), pero que la solución óptima al PDI, u*, no corresponde al punto fijo de G. Indica también que, (de acuerdo a Marcotte) el método de descomposición corresponde a resolver el problema de Cournot-Nash usando el enfoque de Gauss-Seidel. Sin embargo, Fisk muestra que este método puede producir resultados intermedios que poseen un valor de la función objetivo menor que el correspondiente a la solución de equilibrio del problema Cournot-Nash.

El método puede también ser aplicado a la solución del PDR. En dicho caso, la primera etapa consiste en una asignación de flujos de equilibrio multimodal y la segunda en una minimización unidimensional en las variables d_r.

La asignación multimodal se realiza utilizando el método de relajación o diagonalización. En nuestro caso y dado que en la primera etapa del método de descomposición las frecuencias son fijas, el método de diagonalización converge en una sola iteración del siguiente tipo (ver, Florian y Spiess, 1983):

Paso 0 : (inicialización). Encuentre una solución factible $(\tilde{f}, \tilde{d})$

Paso 1 : Diagonalice las funciones de costo

$$c_a(f_a) = c_a(f_a, \tilde{d})$$
$$\bar{c}_a(d) = \bar{c}_a(\tilde{f}_a, d)$$

Paso 2 : Resuelva el siguiente problema diagonalizado a fin de encontrar $(\hat{f}, \hat{d})$

$$\underset{(f,d)}{\text{Min}}\ Z = \sum_a \int_o^{f_a} c_a(x, \tilde{d})\, dx + \sum_a \int_o^{d} \bar{c}_a(\tilde{f}_a, y)\, dy$$

$$\text{s.a.: } f \in \Phi \quad f_a \geq 0, \quad \forall a \in A$$

Paso 3 : Test de convergencia, si $(\hat{f}, \hat{d})$ es suficientemente cercano a $(\tilde{f}, \tilde{d})$ pare. Sino prosiga.

Paso 4 : Haga $(\tilde{f}, \tilde{d}) = (\hat{f}, \hat{d})$ y regrese al Paso 1.

Note que dado que en esta etapa las frecuencias d son constantes, el segundo término de la función Z es constante y por lo tanto puede eliminarse de consideración en el correspondiente problema de minimización. Por lo tanto, el problema a resolver en el Paso 2 se reduce a:

$$\text{Min } Z : \Sigma \int_{o}^{f_a} c_a(x, \tilde{d})\, dx$$

$$\text{s.a. : } f \in \Phi,\ f_a \geq 0,\ \forall\, a \in A$$

Como resultado de este proceso se obtendrían los flujos de equilibrio de transporte privado f* , los costos de operación de equilibrio de transporte privado (c_a, $\forall$ a ε A) y los tiempos de viaje en transporte público (c_a, $\forall$ a ε L).

A continuación y para completar la primera etapa del método de descomposición, se procede a asignar la matriz O-D de viajes en transporte público a la red R de rutas de dicho nodo, utilizando los tiempos de viaje c_a obtenidos arriba. Por lo tanto, como resultado de la primera etapa obtendremos una especificación del vector de flujos multimodal F.

En la especificación del PDR hemos supuesto que la partición modal es fija entre transporte privado y transporte público y por lo tanto existe una matriz O-D fija para cada modo. Es importante anotar que la especificación puede ser extendida para incluir partición modal variable. En tal caso, es posible también resolver el problema de asignación multimodal con demanda variable utilizando una variación del método de diagonalización presentado arriba, que básicamente consiste en una modificación del Paso 2, que implica uncluir en la función Z un término que considere las demandas variables por transporte privado.

La segunda etapa del método de descomposición consiste en resolver un problema unidimensional en d_r para cada ruta de transporte público. Los problemas a resolver son de la forma:

$$\begin{aligned} \underset{d_k}{\text{Min}} : & \sum_{a \varepsilon L_k} C_a(\tilde{f}_a, d', d_k) + \sum_{r \varepsilon R} \sum_{a \varepsilon L_k} \overline{C}_{ar} (\tilde{f}_a, \tilde{\tilde{f}}_{ar}, d', d_k) \\ & + \sum_{r \varepsilon R} \sum_{(i,j)_s^k} \left[\tilde{\tilde{f}}_{ij}^{r} / 2(d_r + \sum_{\ell_{ij}} {}_{r'} d_{r'}) \right] \end{aligned} \tag{17}$$

$$\text{s.a. : } d_k \geq 0$$

en que d' es el vector de frecuencias sin incluir la correspondiente a la línea k, f_a, f_{ar}, $\bar{f}^{r}_{ij}$ son los flujos obtenidos como resultado de la primera etapa y $\bar{\ell}^{r'}_{ij}$ es el conjunto de líneas comunes a la línea r en el segmento (i, j) pero con excepción de la línea k.

Es interesante hacer notar que en este caso los problemas de minimización unidimensional planteado en (17) no son completamente separables en las variables d_r, para F dado, ya que al existir líneas comunes, la variación de la frecuencia de una línea específica afecta a los costos de operación, y tiempos de viaje y espera de todos los usuarios, de las líneas que poseen algún segmento común con ella.

3.3. Aproximación a una solución óptima del sistema

Como vimos en la sección 2.2 una forma de simplificar el PDI es suponer que la distribución de flujos sobre la red f*(u) será de acuerdo a un óptimo de operación del sistema y no a un equilibrio de usuarios. Con ello la formulación del PDI se simplifica notablemente y su solución se puede obtener en forma muy eficiente mediante la utilización de algoritmos conocidos y probadamente convergentes.

Varios autores (Los, 1979, Marcotte,1983b Abdulaal y Leblanc, 1984) han propuesto utilizar el resultado obtenido del PDI' como una aproximación razonable al resultado del PDI.

3.4. Acotamiento del PDI

Harker y Friesz (1984) proponen utilizar los métodos descritos en los puntos 3.2 y 3.3 para acotar la solución exacta al PDI.

Los citados autores justifican este enfoque basados en los siguientes argumentos:

i) El método de H-J es un procedimiento heurístico que no asegura la obtención de la solución exacta al PDI y además es demasiado caro en términos computacionales como para ser aplicado en el caso de redes de tamaño real.

ii) El enfoque simplificado, que considera una distribución de flujos de acuerdo a un óptimo del sistema, entrega una solución con un valor que constituye una cota inferior para la solución exacta al PDI. Esto es obvio si se considera que ningún otro principio de distribución de flujos puede presentar un costo total de operación menor. Así, para cualquier valor dado del vector u, el costo total de operación es mínimo cuando los flujos se distribuyen de acuerdo a un óptimo del sistema.

iii) El método de descomposición optimización-equilibrio no es capaz de entregar el valor de la solución exacta al PDI, pero sin embargo, provee una cota superior para esta.

Este enfoque presenta la dificultad de que el rango definido por las dos cotas puede ser demasiado amplio en casos con elevada congestión y la elección del vector u dentro de dicho rango plantea graves problemas de definición.

4. Agradecimientos

El presente trabajo fue realizado como parte de una investigación que cuenta con el financiamiento del Departamento de Investigaciones de la Universidad Católica de Chile (DIUC), el Fondo Nacional de Desarrollo Tecnológico y el International Development Research Center del Gobierno Canadiense.

5. Referencias

ABDULAAL, M.S. y LEBLANC, L.J. (1979) Continuous equilibrium network design models. Transportation Research, Vol 13B, 19-32.

BARD, J.F. y FALK, J.E. (1982) An explicit solution to the multilevel programming problem. Computers and Operations Research, Vol 9, 77-100.

DE CEA, J. y BUNSTER, J.P. (1985) Asignación todo o nada a redes de transporte público: tratamiento de líneas comunes y transbordos. Primer Congreso Iberoamericano de Métodos Computacionales en Ingeniería, Madrid, 1 - 5 de Julio 1985, España.

ELTON, P. (1983) Análisis de Métodos Continuos de Solución al Problema de Diseño de Redes. Tesis de Ingeniero Civil, Departamento de Ingeniería de Transporte, Pontificia Universidad Católica de Chile, Santiago.

FERNANDEZ, J.E. (1982) Peculiaridades de la aplicación del método de Branch and Bound al diseño óptimo de redes de transporte. Apuntes de Ingeniería 7, 107-147.

FERNANDEZ, J.E. y FRIESZ, T.L. (1983) Equilibrium predictions in transportation markets: the state of the art. Transportation Research, Vol. 17B, N°2, 155-172.

FERNANDEZ, J.E. (1984) Métodos de diseño de redes: aplicaciones al caso de transporte. Segundo Congreso Latinoamericano de Investigación Operativa e Ingeniería de Sistemas, Buenos Aires, 20-24 Agosto 1984, Argentina.

FERNANDEZ, J.E. (1985) Características computacionales de heurísticas continuas para el diseño de redes de transporte. Primer Congreso Iberoamericano de Métodos Computacionales en Ingeniería, Madrid, 1 - 5 Julio 1985, España.

FISK, C.S. (1984) Game theory and transportation systems modelling. Transportation Research, Vol, 18B, N°4/5, 301-313.

FLORIAN, M. y SPIESS, H. (1983) On binary mode choice/assignment models. Transportation Science, Vol. 17, N°1, 32-46.

FLORIAN, M., GUELAT, J. y SPIESS, H. (1985) An efficiente implementation of the "Partan" variant of the linear approximation method for the network equilibrium problem. Publication N° 395, Centre de Recherche sur les Transports, Universitê de Montreal, Canadá.

GUTIERREZ, F.R. (1982) Análisis de Mêtodos de Solución al Problema de Diseño de Redes. Tesis de Ingeniero Civil, Departamento de Ingeniería de Transporte, Pontificia Universidad Católica de Chile, Santiago.

HARKER, P.T. y FRIESZ, T.L. (1984) Bounding the solution of the continuous network design problem. En J. Volmuller y R. Hamerslag (eds.), Proceedings of the Ninth International Symposium on Transportation and Traffic Theory. VNU Science Press, Uthrech.

HOOKE, R. y JEEVES, T. (1961) Direct search solution of numerical statistical problems. Journal of the Association of Computer Machinery, Vol 8, 219-229.

LEBLANC, L.J., MORLOK, E. y PIERSKALLA, W. (1975) An efficient approach to solving the road network equilibrium traffic assignment problem. Transportation Science, Vol 9, 309-318.

LEBLANC, L.J. y ABDULAAL, M. (1984) A comparison of user-optimum versus system-optimum traffic assignment in transportation network design. Transportation Research, Vol. 18B, N° 2, 115-121.

LOS, M. (1979) A discrete convex programming approach to the simultaneous optimization of land-use and transportation. Transportation Research, Vol. 13B, 33-48.

MARCOTTE, P.(1983a) An analysis of heuristics for the network design problem. En V.F. Hurdle, E. Hauer y G. N. Steuart (eds.), Proceedings of tl Eight International Symposium on Transportation and Traffic Theory. University of Toronto Press, Canadá.

MARCOTTE, P. (1983b) Network optimization with continuous control parameters. Transportation Science, Vol. 17, 181-197.

ORTEGA, J. y RHEINBOLDT, W. (1970) Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, Nueva York.

STEENBRINK, P.A. (1974) Optimization of Transport Networks, Wiley, Nueva York.