



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

DISEÑO DE SERVICIOS EXPRESOS PARA BUSES

HOMERO LARRAIN IZQUIERDO

Tesis para optar al grado de
Doctor en Ciencias de la Ingeniería

Profesor Supervisor:
JUAN CARLOS MUÑOZ

Santiago de Chile, Septiembre, 2013.

© 2013, Homero Larrain



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

DISEÑO DE SERVICIOS EXPRESOS PARA BUSES

HOMERO LARRAIN IZQUIERDO

Tesis presentada a la Comisión integrada por los profesores:

JUAN CARLOS MUÑOZ

RICARDO GIESEN

JORGE VERA

ROBERTO COMINETTI

MIQUEL ESTRADA

CRISTIÁN VIAL

Para completar las exigencias del grado de
Doctor en Ciencias de la Ingeniería

Santiago de Chile, Agosto, 2013.



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

DISEÑO DE SERVICIOS EXPRESOS PARA BUSES

HOMERO LARRAIN IZQUIERDO

Tesis presentada a la Comisión integrada por los profesores:

JUAN CARLOS MUÑOZ

RICARDO GIESEN

JORGE VERA

ROBERTO COMINETTI

MIQUEL ESTRADA

CRISTIÁN VIAL

Para completar las exigencias del grado de
Doctor en Ciencias de la Ingeniería

Santiago de Chile, Agosto, 2013.

A mi abuelo, don Homero Larrain
Lorca.

AGRADECIMIENTOS

La realización de esta tesis cierra una etapa en mi vida muy significativa y enriquecedora. Muchas personas e instituciones participaron directa o indirectamente de este proceso, y merecen toda mi gratitud.

En primer lugar, quisiera agradecer a mis padres, Homero y Bernardita, y a mis hermanos y amigos por el constante apoyo y cariño entregado durante todos estos años. Sin ellos, nada de esto sería posible.

También quisiera reconocer el aporte brindado por la Escuela de Ingeniería Industrial de la Universidad Diego Portales, lugar donde me desempeñé como profesor durante los primeros años de este proceso. Esta institución, además de permitir y financiar la realización de una estadía en el extranjero como parte del proceso de esta investigación, proveyó un ambiente idóneo para la investigación y consolidó mi vocación por la docencia, principal impulso que alimentó mis esfuerzos para llevar a cabo esta investigación.

Quisiera agradecer también a la facultad de ingeniería de University College en Londres por acogerme por seis meses, y muy especialmente de los profesores Ben Heydecker y Andy Chow, tanto por la acogida brindada como por el interés mostrado en mi investigación. Agradezco también al CENIT en la Universidad Politécnica de Cataluña por recibirme durante cuatro meses, en especial a Francesc Robusté y Miquel Estrada, por orientar y apoyar mi investigación durante mi estadía en Barcelona.

También quisiera agradecer en forma muy especial a los profesores que integraron la comisión evaluadora de esta tesis. De las conversaciones compartidas surgieron numerosas ideas, que abren nuevos frentes de investigación que pretendo abordar en el futuro en mi vida académica.

No puedo dejar de agradecer a todo el departamento de Ingeniería en Transporte de la Universidad Católica, al cual me integro al terminar este doctorado y del que formaré parte con gran orgullo. Quisiera agradecer muy especialmente a Juan Carlos Muñoz, quien no

sólo hizo un excelente papel de guía durante la realización de esta investigación sino también ha sido siempre un gran amigo y orientador.

Finalmente, quisiera agradecer a Conicyt por el apoyo financiero prestado para la realización de esta investigación. Adicionalmente, esta tesis también contó con el apoyo del Fondecyt n° 1110720 Planning and Analysis of Express Services for an Integrated Public Transport System y del Centro BRT financiado por la Volvo Research and Educational Foundations.

ÍNDICE GENERAL

	Pág.
DEDICATORIA.....	ii
AGRADECIMIENTOS	iv
ÍNDICE DE TABLAS	ix
ÍNDICE DE FIGURAS	xii
RESUMEN.....	x
ABSTRACT.....	xvi
1. Introducción	18
1.1 Revisión bibliográfica	25
1.2 Metodología	38
2. Optimización de frecuencias de servicios expresos en un corredor	42
2.1 El modelo de Leiva	42
2.2 Mejoras y variantes del modelo de Leiva.....	75
3. Generación de servicios expresos para un corredor.....	103
3.1 Heurísticas para la generación de servicios.....	105
3.2 Evaluación de las heurísticas.....	149
3.3 Diseño de servicios expresos en un corredor	173
4. ¿Cuándo utilizar servicios expresos?.....	178
4.1 Experiencias previas.....	178
4.2 Construcción de los escenarios	186
4.3 Definición del indicador de desempeño	194
4.4 Calibración del modelo de regresión.....	197
4.5 Análisis de los resultados	207
5. Diseño de servicios expresos sobre una red.....	213
5.1 Algoritmo de diseño de servicios expresos para una red	213
5.2 Aplicación del algoritmo	224
6. Conclusiones.....	252

BIBLIOGRAFIA.....	258
-------------------	-----

ÍNDICE DE TABLAS

	Pág.
Tabla 1-1: Modelos de diseño de servicios expresos	37
Tabla 2-1: Parámetros de la instancia con múltiples óptimos locales.....	72
Tabla 2-2: Desglose de costos.....	74
Tabla 3-1: Heurísticas para la generación de servicios sin considerar capacidad....	150
Tabla 3-2: Heurísticas para ajuste de capacidad	151
Tabla 3-3: Heurísticas de generación con restricción de capacidad activa	152
Tabla 3-4: Parámetros de los escenarios	157
Tabla 3-5: Escenarios para el experimento	157
Tabla 3-6: Etapas de solución	158
Tabla 3-7: Costos sociales corregidos (MM \$/hr), sin capacidad activa	161
Tabla 3-8: Servicios óptimos para P1 y P2 sin capacidad	163
Tabla 3-9: Servicios óptimos para G1 y G2 sin capacidad.....	165
Tabla 3-10: Costos sociales corregidos (MM \$/hr), con capacidad activa	166
Tabla 3-11: Número de iteraciones del algoritmo para cada etapa y escenario.....	167
Tabla 3-12: Servicios óptimos para P1 con capacidad	169
Tabla 3-13: Servicios óptimos para P2 con capacidad	170
Tabla 3-14: Servicios óptimos para G1 con capacidad	171
Tabla 3-15: Servicios óptimos para G2 con capacidad	172

Tabla 4-1: Parámetros para la generación de escenarios.....	189
Tabla 4-2: Atributos de las matrices base	193
Tabla 4-3: Parámetros del modelo 1	200
Tabla 4-4: Parámetros del modelo 2	203
Tabla 4-5: Parámetros del modelo 3	206
Tabla 5-1: Matriz de viajes agregada por zona de la red a estudiar	225
Tabla 5-2: Parámetros del modelo	228
Tabla 5-3: Resultados iteración 1.....	229
Tabla 5-4: Indicadores iteración 1.....	230
Tabla 5-5: Resultados iteración 2.....	232
Tabla 5-6: Indicadores iteración 2.....	232
Tabla 5-7: Resultados iteración 3.....	234
Tabla 5-8: Indicadores iteración 3.....	234
Tabla 5-9: Resultados iteración 4.....	236
Tabla 5-10: Indicadores iteración 4.....	236
Tabla 5-11: Resultados iteración 5.....	237
Tabla 5-12: Indicadores iteración 5.....	238
Tabla 5-13: Resultados iteración 6.....	239
Tabla 5-14: Indicadores iteración 6.....	239
Tabla 5-15: Resultados iteración 13.....	241

Tabla 5-16: Indicadores iteración 13.....	241
Tabla 5-17: Resultados solución simplificada	244
Tabla 5-18: Indicadores solución simplificada	244
Tabla 5-19: Resultados solución con capacidad	245
Tabla 5-20: Indicadores solución con capacidad	246
Tabla 5-21: Resultados solución sin expresos con capacidad.....	246
Tabla 5-22: Indicadores solución sin expresos con capacidad.....	247

ÍNDICE DE FIGURAS

	Pág.
Figura 1-1: Servicios expresos alrededor del mundo	21
Figura 2-1: Cálculo del tiempo perdido por detención	49
Figura 2-2: Escenario para múltiples óptimos locales	67
Figura 2-3: Función objetivo con múltiples óptimos locales	73
Figura 2-4: Construcción del conjunto $\mathcal{V}_{(i,j)}$	83
Figura 3-1: Generación de servicios iniciales	106
Figura 3-2: Función de ahorros	109
Figura 3-3: Generación de servicios cortos.....	119
Figura 3-4: Generación de servicios zonales.....	123
Figura 3-5: Generación de servicios combinados para congestión.....	145
Figura 3-6: Perfil de carga matriz P1	153
Figura 3-7: Perfil de carga matriz P2	154
Figura 3-8: Perfil de carga matriz G1.....	155
Figura 3-9: Perfil de carga matriz G2.....	156
Figura 4-1: Definición de escenarios para el experimento.....	180
Figura 4-2: Ahorros en costos sociales v/s dispersión de los viajes.....	183
Figura 4-3: Reducción de costos sociales v/s nivel de demanda.....	184
Figura 4-4: Perfil de carga matriz tipo 1	190

Figura 4-5: Perfil de carga matriz tipo 2	191
Figura 4-6: Perfil de carga matriz tipo 3	192
Figura 4-7: Perfil de carga matriz tipo 4	192
Figura 4-8: Comparación entre indicadores de ahorro social	195
Figura 4-9: Regresión 1	201
Figura 4-10: Desglose regresión 1	202
Figura 4-11: Regresión 2	204
Figura 4-12: Desglose regresión 2	205
Figura 4-13: Regresión 3	207
Figura 5-1: Esquema de la red a estudiar	224
Figura 5-2: Principales movimientos de la red a estudiar	225
Figura 5-3: Perfil de carga zona A	226
Figura 5-4: Perfil de carga zona B	226
Figura 5-5: Perfil de carga zona C	227
Figura 5-6: Evolución de los indicadores.....	242
Figura 5-7: Diagrama solución.....	248

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

DISEÑO DE SERVICIOS EXPRESOS PARA BUSES

Tesis enviada a la Dirección de Investigación y Postgrado en cumplimiento parcial de los requisitos para el grado de Doctor en Ciencias de la Ingeniería.

HOMERO LARRAIN

RESUMEN

Los servicios expresos, que en general se pueden definir como servicios que atienden a un subconjunto de los paraderos de un determinado corredor, son una estrategia de operación que puede ser beneficiosa tanto para los operadores como para los usuarios. Una reducción en el número de detenciones beneficia a los usuarios al reflejarse en una disminución en los tiempos de viaje. Naturalmente, esto también favorece a los operadores, que al ver reducidos sus tiempos de ciclo pueden hacer un uso más eficiente de su flota. Estos beneficios deben compensar el perjuicio que experimentan los usuarios de los paraderos omitidos, cuyos tiempos de espera pueden verse elevados por este tipo de medida.

En este trabajo se propone una forma de diseñar servicios expresos en una red de transporte público urbano operada por buses. El algoritmo de diseño que se plantea genera un conjunto de servicios y entrega sus frecuencias de operación, tomando en cuenta la restricción que impone la capacidad máxima de los buses, y considerando que los pasajeros minimizan sus tiempos esperados de viaje considerando la existencia de líneas comunes y la posibilidad de realizar trasbordos.

El problema descrito es abordado enfocando inicialmente los esfuerzos en el caso de un corredor aislado, para posteriormente construir una heurística que diseña los servicios de una red descomponiéndola en corredores. El problema de diseño de servicios expresos para un corredor es a su vez descompuesto en dos subproblemas: el problema de optimización de frecuencias para un conjunto de servicios fijo, y el problema de generación de servicios para un corredor.

En Leiva et al. (2010) se formula un modelo para la optimización de frecuencias de servicios expresos en un corredor que fue utilizado como punto de partida para esta investigación. Un análisis en profundidad de este modelo permitió detectar algunas posibilidades de mejora, dentro de las que se cuentan nuevas formas de representar los servicios introduciendo restricciones de continuidad de frecuencia por nodo, y mejoras en la heurística de ajuste de capacidad que reducen drásticamente los tiempos de ejecución.

Para resolver el problema de generación de servicios se diseñaron diversas heurísticas. En términos generales, estas heurísticas evalúan el efecto sobre los costos sociales de permutaciones locales como eliminar o agregar paraderos sobre servicios existentes, o comparan los beneficios de diferentes configuraciones de servicios de formato predeterminado (cortos y zonales). Se proponen también heurísticas diseñadas especialmente para el caso con restricción de capacidad activa, donde centrando el análisis en el arco crítico del corredor se obtienen expresiones analíticas para frecuencias y costos óptimos. Todo esto permitió construir un algoritmo para el diseño de servicios expresos en un corredor que combina el modelo de optimización de frecuencias con las heurísticas de generación de servicios encontradas en esta etapa.

El modelo así obtenido fue implementado en cerca de mil escenarios, de forma de identificar los factores que hacen propicio un corredor para la implementación de servicios expresos. Calibrando un modelo de regresión lineal se concluye que estos beneficios son mayores cuando los viajes son en promedio más largos y están concentrados en menos pares, cuando el tiempo perdido por cada detención es mayor, cuando el volumen de viajes es más alto, cuando el tiempo de viaje es más valioso, cuando el valor del tiempo de espera es menor, y cuando la capacidad es menor.

El problema de diseño para una red es finalmente abordado mediante un algoritmo iterativo en dos fases: primero, se escoge un corredor de la red, se fija su demanda, y se aplica sobre él el algoritmo de diseño, y luego se optimizan las frecuencias de la red incorporando los nuevos servicios generados, reasignando a los usuarios sobre la red completa. Cuando no se obtiene mejora para ningún corredor de referencia, el resultado es ajustado para satisfacer las restricciones de capacidad. Este algoritmo fue implementado en una red de prueba con restricción de capacidad activa, donde mostró ser capaz de generar servicios expresos que entregaban ahorros considerables en los costos del sistema.

Palabras clave: Servicios expresos, transporte público, diseño de redes, BRT.

Miembros de la Comisión de Tesis Doctoral

Juan Carlos Muñoz
Ricardo Giesen
Jorge Vera
Roberto Cominetti
Miquel Estrada
Cristián Vial

Santiago, Septiembre, 2013.

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

EXPRESS SERVICE DESIGN FOR BUSES

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Doctor in Engineering Sciences by

HOMERO LARRAIN

ABSTRACT

Bus services that serve a subset of the bus stops along their route, usually known as express or stop-skipping services, are a powerful strategy that can benefit both users and operators. Decreasing the number of stops will favor users through a reduction in travel times. Naturally, this benefits operators as well, who can give their fleet a more efficient use thanks to lower cycle times. However, these upsides must compensate the delays suffered by users of the skipped stops, whose waiting times might rise due to this measure.

This work introduces a methodology for solving the bus-operated express service design problem on a public transport network. A solution for this problem consists on a set of generated services and their operational frequencies. The solution must take into account that buses have a maximum passenger capacity, and that users will minimize their expected travel times considering common lines and possible transfers.

The problem is tackled taking focus firstly on an isolated bi-directional corridor to later build on a heuristic for the network problem. The design problem for a corridor is in turn divided in two smaller problems: the frequency setting problem for a fixed set of express services, and the service generation problem.

In Leiva et al. (2010) a frequency optimization model for express services is formulated. This model served as a starting point to this research. An in-depth analysis of this model revealed some potential improvements, like the introduction of a new way to represent services using frequency continuity constraints, and some adjustments to the capacity heuristics, that help to reduce dramatically computation times.

Several heuristics were design for the service generation problem. In general terms, these heuristics evaluate the effect over social costs of local permutations such as adding or removing stops to existing services, or compare the benefits from different configurations of services of predefined format (short turning and zonal). In addition, some heuristics are designed for the specific case where buses operate at full capacity, where, focusing the analysis on the critical arc of the corridor, analytical expressions are obtained for the optimal frequencies and costs. An algorithm for the express service design problem for a corridor was finally built combining these heuristics with the aforementioned frequency optimization model.

This algorithm was tested on almost a thousand scenarios, in order to identify which factors make a corridor propitious for the implementation express services. Fitting the data to a linear regression model it is shown that the benefits are higher when trips are in average longer and more concentrated in fewer OD pairs, when dwell times are higher, when the total number of trips is higher, when travel time is more valuable, when waiting times are less valuable, and when bus capacity is lower.

The design problem for a network is finally solved by a two-step iterative algorithm, which first designs the services for an isolated corridor on the network fixing its demand, and then it optimizes the frequencies for the whole network, reassigning the trips it carries. When it is no longer possible to improve the solution for any of the corridors in the network, the solution is adjusted to meet capacity constraints. This algorithm was applied on a test network with active capacity constraints, where it proved able to generate express services which yielded great social costs savings for the system.

Keywords: Express services, network design, public transport, BRT.

Members of the Doctoral Thesis Committee:

Juan Carlos Muñoz
Ricardo Giesen
Jorge Vera
Roberto Cominetti
Miquel Estrada
Cristián Vial

Santiago, September, 2013.

1. INTRODUCCIÓN

Los esfuerzos por proveer de un mejor estándar de calidad en los sistemas de transporte público basados en buses se han intensificado alrededor de todo el mundo durante la última década, y siguen ganando cada día más popularidad. Esto se debe a que, en primer lugar, la experiencia ha demostrado que el transporte público de calidad es vital para el desarrollo sustentable de una ciudad, y a que esta opción ha probado que puede ser altamente eficiente en términos de costos y beneficios tanto para operadores como para usuarios. Desde comienzos del presente siglo se han propuesto y acuñado distintos términos para referirse a este esfuerzo por dotar a los corredores de buses de atributos que permitan acercar su nivel de servicio al ofrecido por Metro. Mientras que en Europa se habla de Buses with High Level of Service (BHLS), en Latinoamérica ha surgido con fuerza el concepto de Bus Rapid Transit (BRT), que ha permeado en gran parte del mundo. El BRT es, de acuerdo con la Metropolitan Transportation Authority (MTA) de Nueva York, “un sistema de transporte público flexible, integrado y de alto desempeño, con imagen de calidad y una fuerte identidad” y combina la velocidad, confiabilidad y otras virtudes de los sistemas de transporte público rodado con la flexibilidad de los buses (www.mta.info). BRT también ha sido definido como un “transporte público de calidad, orientado al usuario, que entrega movilidad urbana en forma rápida, cómoda y eficiente” (Wright, 2003). De acuerdo a Global BRT Data, se estima que en la actualidad existen aproximadamente 160 ciudades a nivel mundial que han implementado alguna forma de BRT o BHLS en sus calles. De estos sistemas cerca de un 75% fueron inaugurados

posteriormente al año 2000, y casi la mitad de los sistemas en el mundo llevan 6 años o menos en operación (www.brtdata.org).

Existen varios elementos que definen a un sistema como BRT. Dentro de los más importantes se pueden incluir la segregación de las vías para el uso de los buses, colección de tarifas en forma centralizada, paraderos modernos con pago anticipado de tarifa y abordaje expedito a los buses, uniformización de la imagen de los buses, y priorización de señales, entre otras (Levinson, 2002). Un elemento que también se suele asociar a BRT corresponde a la utilización de servicios especiales en corredores de alta demanda. Furth y Day (1985) mencionan dentro de estas estrategias de operación a los servicios cortos (*short turning*), servicios expresos, y la estrategia de *deadheading*, que se describe más adelante en este documento.

Los servicios expresos, que en general se pueden definir como servicios que atienden a un subconjunto de los paraderos de un determinado corredor, son una estrategia de operación que puede ser beneficiosa para todos los actores involucrados. Una reducción en el número de detenciones implica una disminución en los tiempos de viaje, ya sea por el ahorro de tiempo que proviene directamente de no detenerse en los paraderos, o a la posibilidad de cambiar partes del trazado para aprovechar rutas más rápidas. Esta disminución en los tiempos de viaje no sólo beneficia a los usuarios: una reducción en los tiempos de ciclo permite al operador hacer un uso más eficiente de sus vehículos, disminuyendo los costos de operación. De todas formas, para que los servicios expresos sean una alternativa atractiva, los beneficios mencionados deben compensar los aumentos en los tiempos de espera y eventuales trasbordos que enfrentan los usuarios de los paraderos omitidos.

En el contexto de este trabajo, se entenderá como servicio expreso cualquier servicio que omita algún paradero de un corredor de referencia. De esta forma, se está de alguna manera haciendo una generalización de la definición, de forma de englobar lo que habitualmente en la literatura se conoce como servicio expreso (también a veces referido como servicio zonal, *limited-stop*, o *stop-skipping*, aunque este último nombre también se puede referir a un problema de omitir paraderos para dar regularidad el servicio y que se resuelve en tiempo real, y que escapa al alcance de esta investigación), con otros formatos tales como los servicios cortos (que pueden o no omitir paraderos al interior de su trazado), o servicios tipo *deadheading*, definidos más adelante.

Existen ejemplos de implementación de servicios expresos en numerosas ciudades del mundo: según Global BRT Data existen 158 ciudades distintas, en países de todo el mundo, en las que opera algún servicio de este tipo (Figura 1-1). Algunas de las ciudades donde estos servicios operan exitosamente son Bogotá, Los Angeles y Nueva York (entre muchas otras ciudades de los Estados Unidos), Curitiba (célebre por ser una de las primeras ciudades en implementar BRT), Pekín, Amsterdam, Auckland, Lagos, y un largo etcétera. En el caso de Transmilenio, el ampliamente reconocido sistema BRT de la ciudad de Bogotá, los servicios expresos forman la espina dorsal de la operación del sistema. Durante días hábiles, un 88% de la flota de este sistema se destina a este tipo de servicio, y un 90% de los servicios que se ofrecen son expresos¹. Sin embargo, y a pesar de la propagación del uso de este tipo de servicios alrededor del mundo y del gran potencial que representan como herramienta para mejorar el desempeño de las redes de transporte

¹ Fuente: Conversación vía e-mail con Carlos Acosta, profesional de Subgerencia de Desarrollo de Negocios de Transmilenio S.A.

público, los modelos comerciales o herramientas científicas actualmente existentes para el diseño de transporte público no son capaces de modelar efectivamente los servicios expresos y menos aún diseñarlos.

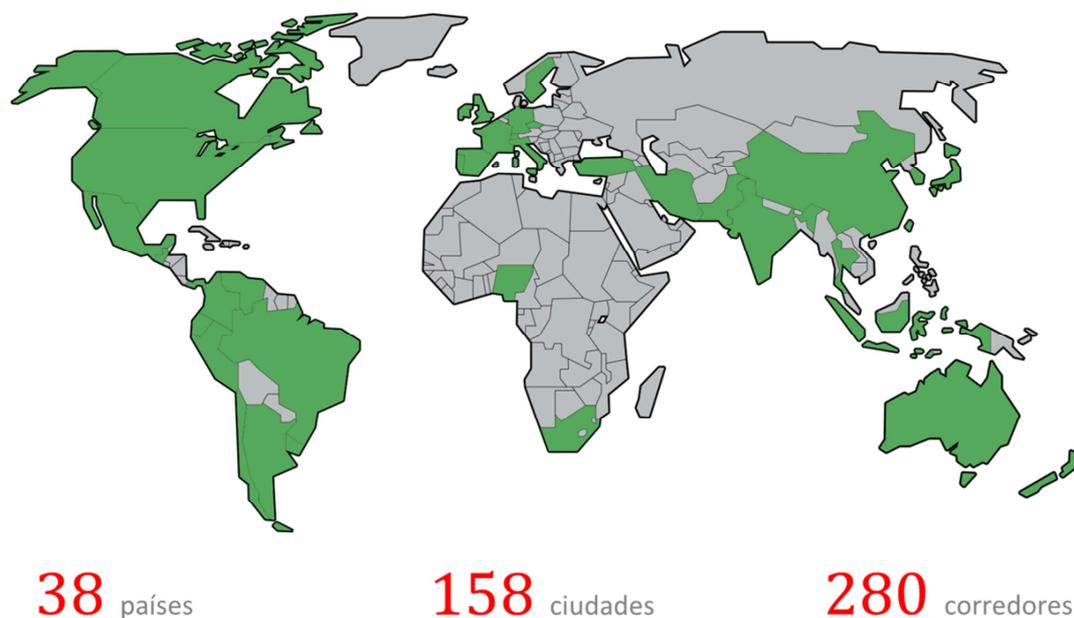


Figura 1-1: Servicios expresos alrededor del mundo

El diseño de servicios expresos es un problema altamente complejo. Para entender mejor el problema, podemos representarlo como un problema en tres niveles. En un nivel superior se encuentra el problema de diseño de trazado de los servicios. Este problema consiste básicamente en determinar qué paraderos atender con cada servicio. Si suponemos por simplicidad que estamos diseñando servicios para un sentido de un corredor de n paraderos, existen del orden de 2^n servicios distintos que podrían operar sobre él. La naturaleza combinatorial de este subproblema lo hace ser de por sí considerablemente complejo.

En un segundo nivel se encuentra el problema de optimizar las frecuencias para un conjunto de servicios ya conocido. Este problema se traduce en encontrar un equilibrio entre los costos del operador, que crecen con la frecuencia, y los tiempos de espera de los usuarios, que decrecen con ésta, y que representa un problema no lineal. Este problema se encuentra íntimamente ligado con un tercer nivel del problema, que corresponde a predecir el comportamiento de los usuarios en un sistema de transporte dado. Dada la existencia de líneas paralelas, trasbordos, y capacidad limitada en los vehículos, que se traducen en una asimetría en las interacciones de los flujos de la red, la modelación del comportamiento de los usuarios se complica considerablemente (Sheffi, 1985).

Para ilustrar la complejidad que implica el comportamiento de los usuarios, consideremos la siguiente situación: un cierto número de usuarios desea realizar un viaje entre dos paraderos, entre los cuales operan dos servicios: un servicio lento y un servicio rápido. Los usuarios, que supondremos uniformes, deben escoger entre estas dos estrategias: esperar al servicio rápido, o tomar el primer servicio que pase, sea del tipo que sea. Además, si asumimos que los usuarios no cuentan con información en línea, qué estrategia es mejor dependerá de las frecuencias de los servicios. Esto implica que en el dominio de combinaciones de frecuencias aparecen fronteras que separan distintos tipos de conducta de los usuarios. El problema que esto conlleva es que las distintas estrategias óptimas implican distintos usos de los servicios: a un lado de esta frontera, todos los usuarios utilizan sólo el servicio rápido; al otro lado, la carga se reparte entre ambos servicios en forma proporcional a sus frecuencias. Esta discontinuidad en la carga de los servicios complica el cumplimiento de las restricciones de capacidad del problema, ya que variaciones infinitesimales en las frecuencias pueden generar saltos en el nivel de carga de

los buses. Cabe señalar que este análisis supone que la capacidad de los buses no es restrictiva, ya que si además los buses comienzan a llenarse y los usuarios se ven impedidos de escoger su primera opción el análisis se vuelve mucho más complejo aún.

Una forma de modelar el comportamiento de los usuarios consiste en buscar el diseño de servicios que minimice los costos totales del operador y los usuarios. Esto implícitamente asume que si todos los usuarios minimizan sus propios costos, los costos del sistema también se ven minimizados, y por ende ambos enfoques son equivalentes. Sin embargo, esto es cierto sólo si los costos individuales de los usuarios no se ven afectados por lo que otros usuarios realizan (como sucede al considerar la comodidad como factor de decisión), y si la capacidad limitada de los vehículos no es tomada en consideración. Cuando los buses tienen capacidad limitada el óptimo social se suele alcanzar asignando usuarios a rutas que no utilizarían en forma espontánea, por ejemplo para guardar capacidad en un bus para otros usuarios del sistema. Esto obliga a hacer algún tipo de corrección en la modelación que permita diseñar un sistema que presente una asignación de equilibrio de usuarios (es decir, que en la solución ningún usuario pueda mejorar su situación cambiando unilateralmente su decisión), y que satisfaga la restricción de capacidad de los vehículos.

El problema de diseño crece aún más en complejidad si trabajamos con una red de corredores en lugar de un corredor aislado. En primer lugar, normalmente una red consta de más paraderos que un corredor, lo que dispara el número de servicios factibles y la complejidad de los problemas de optimización involucrados. Incluso asumiendo que el número de paraderos permanece constante, el problema sobre una red es más complejo que sobre un corredor. En el caso de un corredor, por ejemplo, existe una solución trivial que

consiste en proveer un único servicio de ida y vuelta que visite todos los paraderos del corredor (en lo que sigue de esta tesis, llamaremos a este tipo de servicios regulares), cuya frecuencia óptima puede ser fácilmente obtenida en forma analítica. Cuando se trabaja con una red, dependiendo de la complejidad de ésta, puede existir un alto número de posibles corredores cuya optimización simultánea resulta de por sí desafiante. Otra dificultad que surge al trabajar sobre una red, es que un servicio ya no queda completamente definido a partir de sus paraderos, ya que se pueden generar ambigüedades respecto al orden de atención de éstos.

Esta tesis tiene como objetivo principal construir un método para el diseño de servicios expresos que cumpla con los siguientes requisitos:

- Que sea capaz de generar nuevos servicios que podrían ser atractivos en la solución óptima.
- Que incorpore restricciones de capacidad en vehículos.
- Que represente adecuadamente el comportamiento de los usuarios.
- Que sea aplicable en una red.

Los resultados que este modelo entrega son principalmente dos: los diseños de los servicios (es decir, la secuencia de nodos a visitar por cada uno de ellos), y la frecuencia con la que cada servicio debe operar. Este modelo fue aplicado en un experimento orientado a obtener una visión del potencial que representan los servicios expresos, identificando cuáles son los principales elementos de una red que indican que representa una instancia favorable para la implementación de dichos servicios. En palabras sencillas, los objetivos de esta tesis se pueden condensar en dos preguntas centrales: ¿cómo diseñar servicios expresos? y ¿cuándo utilizar servicios expresos?

A continuación se presenta una revisión de la literatura existente relacionada con el problema de diseño de servicios expresos. Esta sección está organizada de la siguiente forma: en primer lugar se realiza una revisión sobre los modelos existentes para el problema de diseño de trazados, que es un problema íntimamente relacionado con el de diseño de servicios expresos. Luego se discute sobre el problema de optimización de frecuencias y de qué manera ha sido abordado en la literatura. Posteriormente se entrega un análisis sobre las publicaciones existentes que abordan en forma directa el problema del diseño de servicios expresos. A continuación se cierra la sección exponiendo distintas formas reportadas para modelar los costos de los operadores y los usuarios de una red de transporte público, y distintos enfoques existentes para predecir el comportamiento de los usuarios.

1.1 Revisión bibliográfica

De acuerdo con Ceder y Wilson (1986), el problema general de planificación de un sistema de transporte público puede ser entendido como un problema en cinco etapas. La primera etapa correspondería al diseño de trazados, seguido por determinación de frecuencias, construcción de horarios de operación, y asignación de flota y de personal. En Desaulniers y Hichkman (2007) se provee una completa revisión del estado del arte para las diferentes etapas del problema de planificación de transporte público. El diseño de servicios expresos combina elementos de los problemas de diseño de trazados con optimización de frecuencias, por lo que el énfasis de esta revisión bibliográfica se pondrá en estos dos problemas en particular, y en cómo han sido estudiados en relación a los servicios expresos.

El problema de diseño de trazados consiste en determinar los recorridos que minimizan los costos sociales para una red. Existen varios ejemplos en la literatura que abordan este tema, de los que se pueden citar, además del ya mencionado de Ceder y Wilson (1986), a LeBlanc (1988), Baaj y Mahmassani (1995), Fan y Machemehl (2006), y Mauttone y Urquhart (2009), y las revisiones del estado del arte de Ceder (2003), Desaulniers y Hickman (2007), y Guihaire y Hao (2008). En general, estos trabajos no consideran para el análisis la restricción de capacidad de los vehículos. En este caso, basta con minimizar los costos del sistema para obtener una solución que corresponde a una asignación espontánea de los usuarios, sin embargo, cuando la restricción de capacidad es incorporada y ésta se hace activa para algunos servicios, la asignación obtenida puede no responder a un comportamiento racional de los usuarios. En Fernández et al. (2008) se presenta una formulación binivel que permite diseñar servicios de buses sobre una red considerando la capacidad de los vehículos.

Una línea de trabajo que merece mención aparte por su posible relevancia para la solución del problema de diseño de servicios expresos, corresponde al enfoque continuo de modelación, en el que se asume una distribución continua de la demanda (generalmente uniforme), y se representa el trazado de los servicios a través de una función a optimizar a lo largo del trazado, y que puede representar, por ejemplo, densidad de paraderos o densidad de los corredores en una grilla. Dos trabajos recientes que aplican este tipo de enfoque son el de Daganzo (2010) y el de Estrada et al. (2011), donde se aplica este tipo de modelo para obtener una red de servicios para la ciudad de Barcelona. El enfoque continuo también ha sido aplicado en otros

problemas relacionados con el diseño de transporte público, como la localización de paraderos (Medina-Tapia et al., 2013) o el diseño de servicios para trenes que, compartiendo una misma vía, alternan los paraderos a servir para mejorar sus tiempos de viaje (Freyss et al., 2013). Este último esquema de operación opera exitosamente en la actualidad en el metro de Santiago, Chile.

La principal diferencia entre el problema de diseño de recorridos con el de los servicios expresos es que el primero asume que los recorridos atienden a todos los paraderos de los corredores por los que pasan (o, simplemente, no abordan el tema de los paraderos). Ambos problemas podrían ser entendidos como distintos niveles de detalle de un problema más amplio de diseño: en un nivel superior existe una optimización de trazados que determinan las vías por las cuáles circularían los buses, y en un nivel inferior se optimizan los servicios sobre los corredores formados.

A pesar de ser consideradas como dos etapas diferentes del problema de planificación de un sistema de transporte público según la clasificación de Ceder y Wilson (1986), muchos autores no hacen distinción entre los problemas de optimización de frecuencias y el de definición de horarios de operación. De hecho, el mismo Ceder colapsa estas dos etapas del modelo de planificación en una sola en Ceder (2001). Esto tiene sentido ya que por lo general los sistemas de transporte público operan por diseño bajo una de estas dos configuraciones. Normalmente se dice que un sistema basado en frecuencias es un sistema donde los vehículos operan a una frecuencia lo suficientemente alta para que no tenga mucho sentido especificar los horarios a los que pasan los buses, debido a la variabilidad de los tiempos de operación, y a lo bajo de los tiempos de espera involucrados. En este tipo de sistemas se asume que los

pasajeros llegan a los paraderos en forma aleatoria e independiente de los horarios de los buses. Por otra parte, un sistema basado en horarios corresponde al caso contrario, donde los buses sí tienen y publican un horario de funcionamiento, y donde se puede esperar que la conducta de los usuarios se ajuste a esta información lo que complica considerablemente el análisis (Turnquist, 1978). Naturalmente, este segundo enfoque implica trabajar con un mayor nivel de detalle. En el contexto de esta tesis se optó por el primero de estos enfoques, que permite trabajar con una aproximación robusta de los costos sociales, sin complicar excesivamente la formulación del problema.

La literatura relativa al problema de optimización de frecuencias y diseño de horarios es abundante. Dentro de los trabajos más relevantes con respecto a optimización de frecuencias, podemos mencionar los de Vickrey (1955), Mohring (1972) y Jansson (1980) que obtienen a partir de un análisis microeconómico del problema las primeras versiones de la fórmula para el diseño de frecuencias conocida como “fórmula de la raíz”, que muestra que en un corredor la frecuencia óptima es proporcional a la raíz cuadrada del nivel de flujo total de éste. Otros trabajos como el de Kocur y Hendrickson (1982) y el de Chang y Schonfeld (1991) incorporan al análisis el concepto espacial, probando que un aumento en la demanda debe ser compensado no sólo con aumentos en la frecuencia, sino también en la densidad de los servicios. Jara-Díaz y Gschwender (2003) realizan un análisis dentro de esta misma línea donde extienden el problema desde un corredor a redes pequeñas tratables analíticamente. Un estudio de caso para la ciudad de Santiago, Chile, se puede ver en Medina-Tapia et al. (2013), donde se resuelve en forma simultánea el

problema de localización de paraderos con el de optimización de frecuencias para el corredor de Av. Grecia.

El problema de optimización de horarios de operación (o *timetables*, también a veces referido como optimización de líneas *-line setting-* en contraposición a optimización de frecuencias) también ha sido investigado profusamente. Además de la sección dedicada a horarios de operación en Desaulniers y Hickman (2007), un detallado estado del arte sobre este problema en particular se puede encontrar en Nuzzolo y Crisalli (2009). De particular interés para este trabajo resulta la tesis doctoral de Bradley (2010), que presenta un modelo para la optimización de horarios diseñado específicamente para sistemas BRT. De todas formas, y como se menciona anteriormente, en esta tesis se opta por no entrar en tal nivel de detalle para la modelación, dado que el problema abordado ya combina elementos estratégicos (diseño de servicios) y tácticos (frecuencias).

No es posible abordar ninguna de las etapas de interés del problema de planificación de transporte público relevantes para el diseño de servicios expresos sin modelar de alguna forma la asignación de los pasajeros sobre la red que tienen disponible. En general los problemas de diseño de trazados y de optimización de frecuencias incorporan el problema de asignación en forma simultánea, lo que es relativamente sencillo de hacer cuando no existe congestión, aunque modelos que incorporan otros elementos tales como capacidad de los vehículos, llegada no aleatoria de los pasajeros, o estrategias mínimas (que serán explicadas a continuación) pueden forzar a separar los problemas en niveles.

Los primeros modelos de asignación para transporte público corresponden esencialmente a adaptaciones de problemas de rutas mínimas, como es el caso del algoritmo de Dial (1967). Chriqui y Robillard (1975) introducen un nuevo nivel de realismo al problema incorporando el concepto de líneas comunes. Estas líneas comunes surgen de suponer que los usuarios cuando viajan escogen la ruta que les conviene, entendiendo una ruta como una secuencia de viajes entre distintos puntos de trasbordo. En cada una de estas etapas de viaje, conocidas como secciones de ruta, el usuario puede utilizar diferentes líneas, de las que escoge el subconjunto que minimiza su tiempo esperado de viaje. Los trabajos de Spiess (1983) y Spiess y Florian (1983) incorporan una nueva dimensión al problema al reconocer que los pasajeros no necesariamente definen sus trasbordos desde el comienzo del viaje, sino que pueden definirlos de acuerdo a los servicios que pasen primero en cada punto donde esperan. Este tipo de comportamiento es conocido como estrategia de viaje.

El problema de asignación con capacidad es también estudiado por diversos autores, de los que destacan Nguyen y Pallottino (1988) que introduce el concepto de hiperrutas, y de Cea y Fernández (1993) que utiliza las denominadas frecuencias efectivas, que se ven reducidas en presencia de congestión (definida como la tasa de ocupación de los vehículos) aumentando así los tiempos de espera de los usuarios. Esta línea de trabajo es continuada por Wu y Florian (1993), Wu et al. (1994), Cominetti y Correa (2001), y Cepeda et al. (2006), donde se entrega una caracterización de las condiciones de equilibrio de este problema y se propone una formulación como un problema de optimización equivalente. A pesar de la gran cantidad de investigación existente en torno a los distintos problemas de diseño de

transporte público, y de la atención que han acaparado los sistemas BRT en la última década, es escasa la literatura existente dedicada al estudio de los servicios expresos. Uno de los primeros trabajos donde se plantean los esquemas de operación expresos como una alternativa atractiva para mejorar el desempeño de sistemas de transporte público es el artículo de Furth y Day (1985), donde se proponen las siguientes estrategias de operación:

- i. Servicios cortos (*short turn*): Una proporción de los vehículos operan en un servicio que consiste en un ciclo más corto dentro del corredor, de forma de aumentar la capacidad en dicho sector.
- ii. *Deadheading*: Una proporción de los vehículos opera en el sentido menos cargado sin llevar pasajeros (o, al menos, sin detenerse en las paradas intermedias) con el fin de comenzar un nuevo viaje en el sentido más cargado lo antes posible, aumentando de esta forma la frecuencia en este último sentido. Normalmente este esquema de operación se justifica en corredores que presentan un desbalance significativo en la demanda entre ambos sentidos de operación.
- iii. Servicios expresos (*express services, limited-stop services*): Una proporción de los servicios visitan sólo un subconjunto de los paraderos disponibles de un corredor. Un caso particular de este tipo de servicios corresponde a los servicios zonales, donde los servicios atienden u omiten grupos de paraderos consecutivos en un corredor. Otro caso particular es el servicio expreso que atiende únicamente dos paraderos, entregando un viaje rápido entre dos nodos

de alta demanda, normalmente apartados entre sí. Este tipo de servicios son actualmente conocidos como superexpresos en la ciudad de Santiago, Chile.

A pesar de que existen varios trabajos que abordan distintos aspectos de las dos primeras estrategias, es poca la literatura que se ha podido encontrar publicada con respecto a los servicios expresos.

De los trabajos más relevantes sobre estas dos primeras estrategias se pueden destacar a Furth (1987) para los servicios cortos y Ceder y Stern (1981) para *deadheading*. El trabajo de Cortés et al. (2011), que construye sobre el modelo para servicios cortos propuesto en Tirachini et al. (2011), adopta una metodología que permite el análisis de una estrategia que integra *deadheading* con servicios cortos, diseñando servicios que no necesariamente cubren la extensión completa del corredor, y que en sus puntos de retorno se desplazan a lo largo de un tramo sin entregar servicio. Una idea similar es abordada en la sección del Capítulo 2 de esta tesis que se refiere a permitir que los servicios no sean cíclicos en el modelo base de optimización de frecuencias.

De acuerdo a Desaulniers y Hickman (2007), el primer trabajo en los que se aborda el problema de diseño para los servicios expresos es el de Jordan y Turnquist (1979), donde se optimiza la operación de un corredor cuyos viajes son atraídos (o generados) en el último paradero del corredor (o en el primero). El modelo de programación dinámica que el autor propone encuentra el diseño óptimo para la red, consistente en un conjunto de servicios que atienden porciones del corredor y luego se dirigen al destino sin hacer más detenciones, y la asignación de flota para cada uno de estos servicios. Dada la estructura de demanda considerada, este modelo no

requiere considerar trasbordos ni un modelo complejo de comportamiento de usuarios, ya que éstos nunca se ven enfrentados a más de una opción para realizar su viaje. El modelo de Jordan y Turnquist es generalizado en Furth (1986), donde se estudia el problema para el caso de corredores bidireccionales y ramificados, manteniendo la estructura de demanda.

Un problema relacionado con el planteado por Jordan y Turnquist es discutido previamente en el trabajo de Clarence y Hurdle (1975), aunque este último no trabaja con servicios expresos directamente. En dicho trabajo se aplica el enfoque continuo para estudiar el problema donde los viajes comparten un origen y los destinos se encuentran dispersos sobre una ciudad. El objetivo del problema consiste en determinar el tamaño óptimo de las zonas. El trabajo concluye que este tamaño óptimo es proporcional a la raíz cuadrada de la densidad de los viajes cuando se opera a capacidad, y a la raíz cúbica de dicha densidad cuando los buses no operan llenos.

La mayor parte de los trabajos que se conocen sobre servicios expresos están orientados a estudios de casos y evaluación del desempeño de estos sistemas, más que al estudio del problema de diseño de servicios en general. Dentro de esta línea se encuentran los trabajos de Ercolano (1984) que estudia el desempeño de los servicios expresos en comparación con los servicios regulares en el barrio de Manhattan, Nueva York, y el de Silverman (1998) que realiza un análisis a partir de datos de 25 servicios expresos en la ciudad de Nueva York. Este último trabajo concluye que estos servicios generan ahorros en los tiempos de viaje, y al mismo tiempo disminuyen los costos de los operadores. La actitud de los usuarios hacia el sistema

es favorable: más de la mitad de los usuarios opinan que el servicio mejoraba con estos servicios, y más de un 80% de ellos estaban de acuerdo con que se debía mantener este esquema. Incluso se reporta que existe una tendencia de los usuarios a sobredimensionar los ahorros percibidos en sus tiempos de viaje. Este trabajo también identifica a partir de este caso particular algunos de los factores clave para el éxito de la implementación de los servicios expresos, identificando entre otros el nivel de demanda del corredor y el largo promedio de los viajes.

En esta misma línea, Tétreault y El-Geneidy (2010) y El-Geneidy y Surprenant-Legault (2010) proponen y evalúan un servicio expreso para un corredor en Montreal. Para diseñar las paradas del servicio, los autores generan cuatro escenarios que gradualmente incorporan paraderos en el trazado, priorizando los paraderos de mayor demanda. En la evaluación de la implementación del servicio diseñado, los autores observaron ahorros en los tiempos de viaje del orden de los 5 minutos en promedio, pero que eran percibidos por los usuarios como un ahorro todavía mayor. Los tiempos de espera de los usuarios perjudicados por la medida (los que no tienen acceso directo al servicio expreso) vieron aumentados sus tiempos de espera en a lo más 2 minutos.

Un modelo para la evaluación del impacto de los servicios expresos es presentado en la tesis de doctorado de Scortia (2010), que trabaja con un modelo primero introducido por Schwarcz (2004). El modelo de Schwarcz predice, para un corredor donde opera un servicio regular, la asignación de los pasajeros y los beneficios de incorporar un determinado servicio expreso, en términos de una serie de indicadores. Scortia extiende esta formulación para el caso donde un BRT se superpone a un

servicio regular, e identifica algunos de los factores decisivos para el éxito de la implementación de servicios expresos entre los que destacan el nivel de frecuencia del servicio regular existente (homologable al nivel de demanda del corredor, suponiendo que éste se encuentra razonablemente diseñado), al largo promedio de los viajes, y la concentración de los viajes en algunos pares OD.

En el trabajo de Leiva et al. (2010) se propone un modelo para el diseño de servicios expresos en un corredor de transporte público con capacidad restringida. Este modelo minimiza los costos sociales (compuestos por los costos del operador más los costos de los usuarios) para un conjunto predefinido de servicios expresos candidatos. Los resultados que este modelo entrega son compatibles con un comportamiento racional de los usuarios (que minimizan sus itinerarios considerando la existencia líneas comunes), incluso en presencia de capacidad, por medio de la imposición de cotas mínimas crecientes a las frecuencias de los servicios hasta alcanzar una oferta coherente con el comportamiento deseado. En el Capítulo 2 de esta tesis se muestra en detalle el modelo de Leiva, que constituye un punto de partida para el modelo desarrollado en esta tesis.

En Leiva et al. (2010) y en Larrain et al. (2010a) se aplica el modelo de Leiva para realizar diversos análisis sobre la potencialidad de los servicios expresos frente a corredores en diferentes escenarios. Se observa que en la implementación del modelo sobre un corredor de la ciudad de Santiago, la utilización de servicios expresos reportaría ahorros en los costos sociales totales del orden del 10%. Se concluye además que tanto en este caso como otros estudiados es posible obtener una gran parte de los ahorros utilizando un número reducido de servicios distintos. Se observa,

finalmente, que el indicador que más correlación guarda con los ahorros obtenidos por los servicios expresos es el largo promedio de los viajes de la matriz del corredor, aunque también se observa (aunque no se indica en forma categórica) que los servicios expresos generan mayores ahorros porcentuales cuando los niveles de demanda son a su vez mayores.

Existen algunos trabajos recientes que proponen metodologías para el problema de diseño de los servicios expresos para corredores que incorporan la etapa de generación de servicios candidatos. Sun et al. (2008) presenta un modelo de programación entera que optimiza la configuración y frecuencia de tres servicios (un servicio regular, un expreso y un zonal) para minimizar los costos sociales. En este trabajo a los usuarios no se les permite realizar trasbordos, ni se admite la existencia de líneas comunes, modelando implícitamente el comportamiento de los usuarios como una minimización de itinerarios. El trabajo no hace mención a la restricción de capacidad en los vehículos.

En Chen et al. (2012) también se presenta un modelo que minimiza los costos sociales para un corredor, en este caso atendido por un servicio regular y un expreso. Este modelo, al igual que el anterior, asume que los usuarios no realizan trasbordos. En presencia de líneas comunes se supone sencillamente que los usuarios siempre prefieren el servicio expreso. Este modelo incorpora una restricción de capacidad de los vehículos en forma directa, lo que, como se explica en mayor detalle en el Capítulo 2, puede generar soluciones que no respetan la racionalidad de la asignación de los usuarios.

En Chiraphadhanakul y Barnhart (2013) se plantea un modelo de características similares al anterior, que tampoco permite trasbordos, pero que modela la demanda por el servicio expreso como una parte fija, proporcional a su participación en la frecuencia, más una parte variable, proporcional al ahorro de tiempo ofrecido. Este trabajo explora algunas propiedades interesantes del problema, y propone un método incremental para su resolución. La restricción de capacidad también es incorporada al modelo en forma directa como restricción. Este trabajo concluye que tanto la frecuencia original del servicio regular como el largo de los viajes se correlacionan con los beneficios entregados por los servicios expresos, pero no así los niveles totales de demanda del sistema.

La Tabla 1-1 a continuación resume los principales atributos de los modelos para el diseño de servicios expresos hallados en la literatura.

Tabla 1-1: Modelos de diseño de servicios expresos

	Many to many	Opt. de frecuencias	Generación de servicios	Líneas comunes	Trasbordos	Capacidad	Equilibrio de usuarios	Redes
Jordan & Turnquist (1979)		✓	✓				✓	
Sun et al. (2008)	✓	✓	✓				✓	
Leiva et al. (2010)	✓	✓		✓	✓	✓	✓	
Chen et al. (2012)	✓	✓	✓	✓		✓		
Chiraphadhanakul & Barnhart (2013)	✓	✓	✓	✓		✓		

En la tabla se indican las restricciones y supuestos del enfoque adoptado por cada autor. En este contexto, una matriz de demanda se dice que es tipo *many to many* si no impone que los orígenes o destinos de los viajes se encuentren concentrados en un único nodo. Todos los modelos reportados optimizan las frecuencias de los servicios considerados, pero el de Leiva et al. (2010) no genera sus propios servicios, sino que recibe un conjunto propuesto por el modelador, como se indicó anteriormente. Con respecto a la modelación del equilibrio de los usuarios, algunos trabajos consideran la existencia de líneas comunes, aunque sólo Leiva et al. (2010) admite la posibilidad de realizar trasbordos. En general, la presencia de restricciones de capacidad hace que los resultados de los modelos revisados invalide el equilibrio de los usuarios en los resultados, salvo en el caso de Leiva et al. (2010) donde este problema es abordado en forma explícita a través de una heurística. Finalmente, se agregó la columna referente a la aplicabilidad de los modelos sobre una red en la tabla para hacer notar que ninguno de los modelos revisados son aplicables en forma directa sobre una red.

1.2 Metodología

Como se menciona en la introducción, el problema de diseño de servicios expresos se puede entender como un problema en dos grandes niveles: el nivel superior, correspondiente al problema de diseño de trazados (es decir, determinar qué paraderos son atendidos por cada servicio), y un nivel inferior, que corresponde a la optimización de frecuencias de los servicios diseñados en el nivel superior. El problema de diseño de frecuencias también a su vez puede ser entendido como un

problema en dos niveles, separando la optimización de frecuencias de la asignación de los usuarios a la red, aunque lo más común es encontrarlo formulado simultáneamente. En esta tesis el problema será enfrentado mediante este enfoque binivel, aunque también se discute y experimenta con la opción de separar en algunas partes la asignación del problema, que como se verá en el capítulo 5 puede resultar de gran utilidad.

Las etapas seguidas para enfrentar el problema de diseño y responder las preguntas centrales de esta tesis se describen a continuación:

1. Primero, los esfuerzos fueron concentrados en resolver el problema de las frecuencias para un corredor aislado. De esta manera se resuelve un problema de complejidad moderada, pero que sirve como bloque básico para la construcción de un modelo más complejo. Dado al papel crítico que cumple esta pieza en todo el resto de los algoritmos desarrollados, era crucial dedicar mucha atención a este paso de forma de obtener un modelo eficiente y que a la vez entregara buenos resultados. En esta etapa se toma como punto de partida el modelo de Leiva, que provee una forma razonable de resolver el problema de optimización de frecuencias con capacidad, pero que no se hace cargo del diseño de los servicios ni es capaz de optimizar una red, además de presentar varias oportunidades de mejora en términos de eficiencia.
2. Una vez diseñado un modelo para la optimización de frecuencias que trabajaba con restricciones de capacidad y respetaba el comportamiento racional de los usuarios, se procedió a estudiar diferentes estrategias para el diseño de servicios. Este tipo de estrategias se basaron en distintos

indicadores generados a nivel de paraderos y de pares OD, tomando en consideración el efecto sobre los costos del operador y de los usuarios de la inclusión u omisión de paraderos en los trazados de posibles servicios. Con esto se construyeron heurísticas para la generación de servicios, que eran evaluados aplicando el modelo de optimización de frecuencias. También fueron diseñadas heurísticas especiales para la generación de servicios de configuraciones particulares, como servicios cortos y un tipo de servicio que en el contexto de esta tesis fue denominado como zonales. En esta etapa también se estudiaron indicadores y heurísticas diseñados específicamente para situaciones donde el corredor opera con restricción de capacidad activa, sacando provecho de ciertas propiedades particulares a este caso. Al final de esta etapa se propone una heurística consolidada para la generación de servicios expresos para un corredor aislado, que combina las técnicas desarrolladas que mostraron ser más eficientes en los pasos anteriores.

3. La heurística consolidada fue luego aplicada en un experimento donde se generaron cerca de mil escenarios con diferentes atributos de demanda, parámetros de operación y de costos, de forma de identificar económicamente las principales características de un corredor que lo hacen ser un buen candidato a ser atendido con servicios expresos, y así responder la segunda pregunta propuesta como objetivo de este trabajo.
4. Los modelos de optimización de frecuencias y de diseño de servicios fueron finalmente generalizados en una heurística para la optimización de servicios expresos en una red de transporte público. Esta heurística resuelve en forma

iterativa el problema, separando la red en corredores, aplicando las heurísticas para diseñar servicios y optimizar frecuencias en cada uno de estos en forma secuencial. Como resultado de esta etapa se obtiene un algoritmo que permite resolver el problema de diseño satisfaciendo las cuatro condiciones impuestas en el planteamiento del problema: cumplimiento de la capacidad, comportamiento racional de los usuarios, generación de servicios, y posibilidad de trabajar con redes, proporcionando de esta forma respuesta a la pregunta central de esta tesis.

Lo que sigue de este documento está organizado de acuerdo a la metodología recién expuesta. El capítulo 2 presenta e identifica las posibilidades de mejora que existen para el modelo y el algoritmo de Leiva, explora algunas propiedades interesantes del problema, y luego propone un nuevo algoritmo para el problema de optimización de frecuencias de servicios expresos en un corredor y una serie de variantes del problema que resultan de interés en algunos casos especiales. El capítulo 3 plantea un conjunto de heurísticas de generación de servicios, y una heurística consolidada que resuelve para un corredor el problema de diseño de servicios expresos, combinando generación con optimización de frecuencias. En el capítulo 4 se realiza el experimento descrito en la etapa 4 de la metodología, con el fin de determinar cuándo resulta conveniente aplicar servicios expresos. El capítulo 5 presenta una generalización de la heurística del capítulo 3, que permite resolver el problema de diseño en redes, y entrega un ejemplo detallado de una implementación de este modelo. Finalmente, el capítulo 6 sintetiza los principales resultados obtenidos en este trabajo.

2. OPTIMIZACIÓN DE FRECUENCIAS DE SERVICIOS EXPRESOS EN UN CORREDOR

Como primer paso en busca de resolver el problema de diseño de servicios para una red, se estudió el problema restringido a un corredor aislado. Este problema de diseño, como se mencionaba en la introducción, puede ser abordado como un problema en dos niveles: diseño de los servicios, en un nivel superior, y optimización de las frecuencias, en un nivel inferior. Este capítulo se concentra en analizar el segundo de estos problemas y proponer un método de solución.

En el trabajo de Leiva et al. (2010), propusimos un modelo para el problema de optimización de frecuencias de servicios expresos en un corredor. En lo que sigue, se presenta el modelo de Leiva, se discuten algunas de sus fortalezas, debilidades y propiedades, para posteriormente proponer una serie de mejoras y variantes al modelo.

2.1 El modelo de Leiva

En esta sección se presenta el modelo de Leiva, que fue utilizado como punto de partida para el trabajo presentado en esta tesis, que culmina en la formulación de un algoritmo para el diseño de servicios expresos para redes de transporte público.

Algunas de las características más importantes del modelo de Leiva son:

- El modelo recibe un conjunto de servicios definidos previamente por el modelador y optimiza las frecuencias de cada uno de ellos.
- La función objetivo del modelo corresponde a los costos sociales del sistema, compuestos por los costos de operación y los costos de los usuarios. Los costos

de los usuarios se descomponen a su vez en costos por tiempo de viaje en vehículo, costos por esperas, y costos por trasbordos.

- El modelo trabaja sobre un corredor aislado y no considera interacciones con su entorno.
- La matriz de demanda se considera fija, en parte como consecuencia de lo anterior. Esto implica que la demanda no se ve modificada frente a variaciones en los servicios provistos, ni en magnitud ni en estructura.
- Como consecuencia de lo anterior, las caminatas no son consideradas para la optimización al representar un costo fijo para los usuarios, que no consideran la elección de paraderos como una opción.
- La operación de los buses se asume basada en frecuencias, lo que implica que las llegadas de los pasajeros son independientes de las llegadas de los buses. Esto implica a su vez que se asume que los usuarios no utilizan ningún tipo de información en línea, como por ejemplo aplicaciones para teléfonos celulares, que les permita sincronizar su llegada con la de los buses.
- Los usuarios minimizan el costo promedio de sus viajes. Estos costos están conformados por tiempos de espera y de viaje en vehículo, que pueden ser minimizados escogiendo un conjunto de líneas atractivas cuando existen líneas paralelas, y costos de trasbordo.
- La capacidad de los buses (y de otros factores restrictivos como la dimensión de los paraderos) se considera ilimitada en el modelo de optimización. Sin embargo, el artículo presenta una heurística que permite ajustar las frecuencias de los servicios a los flujos demandados por los usuarios de modo de alcanzar

una solución compatible con una asignación espontánea de los usuarios a los servicios del corredor.

- Los tiempos de viaje de cada servicio entre cualquier par de estaciones se asumen conocidos y determinísticos. Esto a su vez implica que el tiempo de detención en una parada es conocido y fijo también (es decir, independiente de los flujos de subida y bajada).
- Las valoraciones de los tiempos de viaje y espera y de los trasbordos idénticas para todos los usuarios, e iguales a las del modelador.
- La penalización por trasbordo (que se suma a la espera adicional durante el trasbordo) es la misma para todos los paraderos e independiente de los servicios que se conectan.
- La red a considerar posee una tarifa fija asociada a la primera etapa del viaje, que es independiente del servicio o del par en que se realiza el viaje, más una tarifa fija por cada etapa adicional. Dado que la parte fija la pagan una vez todos los usuarios, ésta no influirá en la asignación de los usuarios sobre la red. La tarifa asociada a los trasbordos se puede sumar a la penalización por trasbordo.
- Los servicios a optimizar son ciclos completos, es decir, incluyen una definición de sus paraderos de ida y vuelta, aunque no necesariamente deben ser simétricos (lo que implica además que distintos servicios pueden tener la misma configuración en un sentido del corredor) y tampoco deben necesariamente abarcar el corredor completo, dejando abierta la posibilidad de modelar servicios cortos.

Para la modelación del problema de optimización de frecuencias en un corredor se definen los siguientes conjuntos:

- \mathcal{N} : Conjunto de paraderos. Estos paraderos van de 1 a N . Se denominará como r al nodo final del sentido de ida, y $r + 1$ al nodo inicial del sentido de vuelta.
- \mathcal{L} : Conjunto de servicios o líneas. Un servicio se define como un conjunto ordenado de paraderos. Al trabajar sobre un corredor se asume que el orden de los paraderos es el mismo que el orden en el que aparecen sobre el corredor, por lo que en este caso la configuración de un servicio queda determinada por cuáles paraderos del corredor son los que atiende.
- \mathcal{W} : Conjunto de pares origen destino. Consiste en el conjunto de pares de paraderos entre los que puede existir demanda por viajes. En el caso de un corredor de ida y vuelta un elemento $w \equiv (i, j)$ con $i, j \in \mathcal{N}$ pertenece al conjunto \mathcal{W} si corresponde a un viaje en el sentido de ida (es decir $i < j \leq r$) o si corresponde a un viaje en el sentido de vuelta ($r + 1 \leq i < j$).
- \mathcal{S} : Conjunto de secciones de ruta. Este conjunto está conformado por las etapas de viaje que pueden ser utilizadas por los pasajeros como parte de un viaje que podría contener trasbordos. En principio, sus elementos coinciden con los del conjunto \mathcal{W} . En algunos contextos, sobre todo cuando los trasbordos no son tomados en consideración, los conjuntos \mathcal{S} y \mathcal{W} se pueden usar indistintamente. Por ejemplo, los tiempos de viaje, que en general se denotan como $t_l^{\mathcal{S}}$, para algunas demostraciones y cálculos serán denotados por simplicidad como $t_l^{\mathcal{W}}$.

- \mathcal{A} : Conjunto de arcos del corredor. Contiene todos los pares de paraderos consecutivos de un corredor. En el caso de un corredor aislado, este conjunto está conformado por los pares $(i, i + 1)$ donde $i \in \mathcal{N} - \{r, N\}$. En el caso de una red este conjunto depende de la forma en que los corredores sean definidos.

Las variables de decisión del modelo son:

- f_l : Frecuencia del servicio $l \in L$.
- f_l^s : Frecuencia efectiva del servicio $l \in L$ que es atractiva para un viaje en la sección de ruta s .
- V_s^w : Cantidad de viajes que utilizan la sección de ruta s como parte de un viaje en el par w .

Como se puede apreciar, además de las variables f_l , que son las que responden a la pregunta central del modelo, es necesario definir dos tipos de variables auxiliares. Las variables f_l^s sirven para modelar un comportamiento racional de los usuarios. Esto se debe a que al momento de viajar por una sección de ruta, muchas veces resulta conveniente para el usuario no abordar un bus de un servicio para esperar por uno más rápido. Esto implica que f_l^s sólo puede tomar valor 0 o f_l . Por su parte, las variables V_s^w son necesarias para permitir la existencia de trasbordos. En este capítulo se utilizan algunas otras variables auxiliares que resultaría poco práctico incluir en esta lista, y que son definidas donde son utilizadas por primera vez.

Los principales datos y parámetros del modelo se listan a continuación:

- c_l : Costo unitario por ciclo de operación del servicio $l \in L$.

- λ : Proporción del intervalo promedio del servicio que corresponde a la espera promedio de un usuario. En caso de un servicio perfectamente regular, donde los intervalos entre buses son constantes, se tendrá que $\lambda = 0,5$. Cuando la llegada de buses sigue un proceso de Poisson, se tendrá que $\lambda = 1$.
- θ_{wt} : Valor del tiempo de espera.
- θ_{tt} : Valor del tiempo de viaje.
- θ_{tr} : Penalización percibida por el usuario por trasbordar.
- t_l^s : Tiempo de viaje de la línea $l \in \mathcal{L}$ en la sección $s \in \mathcal{S}$.

Tal como en la presentación de las variables del problema, acá se listan los parámetros básicos del modelo (los que aparecen directamente en el problema de optimización del modelo de Leiva). Los tiempos de viaje t_l^s normalmente deben ser preprocesados de acuerdo a la naturaleza del problema a estudiar. En algunos casos puede ser conveniente utilizar mediciones tomadas en terreno, sobre todo cuando los servicios expresos aprovechan de utilizar alternativas no disponibles para el resto de los servicios. En el caso donde todos los servicios comparten físicamente el mismo corredor, estos tiempos pueden ser preprocesados a partir de las distancias entre paraderos, la velocidad de operación (que puede variar a lo largo del corredor), del número de detenciones intermedias del servicio (dependiente del trazado de éste), y del tiempo fijo que se pierde en cada detención τ , como se muestra en la fórmula (2.1):

$$t_l^s = \begin{cases} \sum_{a \in \mathcal{A}} \zeta_a^s \frac{d_a^l}{v_a^l} + \sum_{i=O_s}^{D_s-1} \pi_l^i \cdot \tau, & \forall (s, l) | \pi_l^{O_s} \cdot \pi_l^{D_s} = 1 \\ M, & e. o. c. \end{cases} \quad (2.1)$$

En esta expresión se incorporan algunas funciones y parámetros nuevos. Los operadores O_s y D_s indican el nodo de inicio y de término de la sección s (o de un par w , o un arco a). El parámetro binario ζ_a^s indica si el arco a está contenido en la sección s , que en el caso de un corredor se traduce en que $O_s \leq O_a \wedge D_a \leq D_s$. El parámetro también binario π_l^i adopta un valor unitario cuando el paradero i forma parte del trazado del servicio l . Los valores d_a^l y v_a^l representan la distancia y la velocidad de operación para un arco a y un servicio l . El valor M corresponde a un valor alto, que supere cualquier tiempo de viaje razonable para la red, y que asegure que en la sección s no se utilice el servicio l a menos que $\pi_l^{O_s}$ y $\pi_l^{D_s}$ valgan 1. Notar que un viaje entre dos paraderos seguidos (o entre dos paraderos sin detención intermedia) considerará de todas formas el tiempo de una detención, que representa al tiempo en el paradero de origen, el tiempo de aceleración desde el paradero de origen, y el tiempo de frenado en el paradero de destino. Es importante recalcar que esta expresión no es necesariamente parte del modelo de Leiva, sino que una forma particular de modelar los tiempos de viaje. En el resto de esta tesis se asumirá que los tiempos de viaje responden a esta formulación en particular, aunque en estricto rigor este supuesto sólo se hace necesario al diseñar las heurísticas de generación presentadas en el Capítulo 3.

El tiempo perdido por detención τ se puede calcular como la suma entre un tiempo fijo de detención en el paradero, más un tiempo de frenado, más un tiempo de aceleración, como se muestra en la Figura 2-1 a continuación.

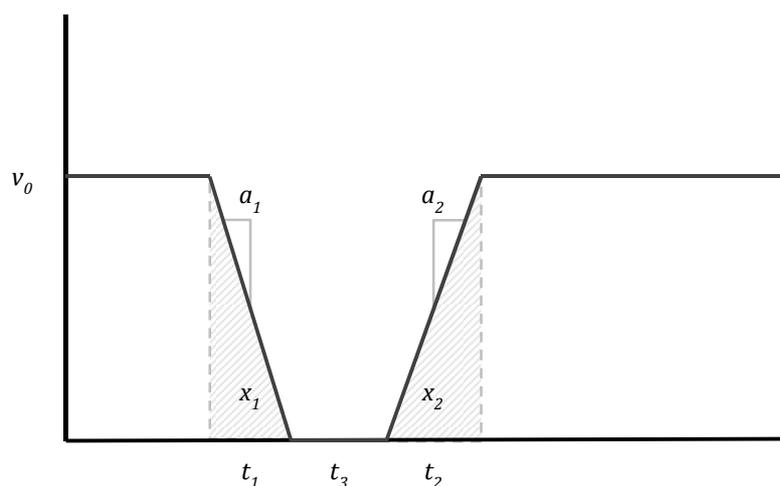


Figura 2-1: Cálculo del tiempo perdido por detención

De acuerdo al gráfico, que asume aceleración y desaceleración constante, el tiempo perdido por detención se puede calcular como:

$$\tau = \frac{v_0}{a_1} + \frac{v_0}{a_2} + t_3 \quad (2.2)$$

En la fórmula anterior aparecen los parámetros v_0 , la velocidad de operación de los buses, a_1 y a_2 , la deceleración y aceleración promedio de los buses, y t_3 , el tiempo de detención en paradero del bus.

El problema de optimización de frecuencias puede considerar distintos objetivos, dependiendo de las necesidades del caso. El modelo de Leiva considera como función objetivo los costos sociales del sistema, compuestos por los costos del operador, más los costos de los usuarios. Estos últimos a su vez se descomponen en

tres partes: costos de espera, costos de viaje en vehículo, y costos de trasbordo. A continuación se presenta la formulación escogida para modelar cada uno de estos elementos:

- **Costos del operador**

El operador en la práctica enfrenta tanto costos fijos como costos variables. Dentro de la primera categoría caben todos los costos que no están relacionados con la frecuencia de los buses o con el tamaño de la flota, como por ejemplo los gastos de administración y gestión de la empresa. Sin embargo, los costos que interesan realmente para esta formulación son sólo los costos variables, es decir, que dependen del tamaño de flota y la frecuencia de los servicios. En Fernández y De Cea (2003) se mencionan, dentro de esta categoría, los costos asociados directamente a la operación de los buses (que incluyen principalmente combustible, lubricantes, repuestos y mantenciones periódicas), los costos por mano de obra, costos de depreciación de la flota, y el pago de patentes e impuestos. Algunos de estos costos del operador dependen fuertemente del kilometraje recorrido por los vehículos, como los costos por combustible, mientras que otros dependerán más directamente del tiempo de utilización de los vehículos, como los salarios de los conductores (notar que el tiempo de ciclo de los servicios también tiene directa relación con el tamaño requerido para la flota de buses). De esta forma, para un servicio l , de frecuencia f_l , es razonable asumir que los costos del operador tienen la siguiente forma:

$$OC_l = (c_L c l_l + c_T c t_l) \cdot f_l = c_l f_l \quad (2.3)$$

En la fórmula anterior los valores c_L y c_T corresponden a los costos variables de operación por unidad de distancia y de tiempo, respectivamente. Estos valores pueden depender del tipo de bus a utilizar, lo que se debe tomar especialmente en cuenta si se consideran distintos tipos de buses a optimizar (en Leiva et al.,2010, se entrega un análisis del efecto de la utilización de buses de distinto tamaño en un corredor con servicios expresos). Los parámetros cl_l y ct_l , que corresponden a la distancia y a la duración del ciclo del servicio l , se obtienen a partir de los tiempos de viaje de acuerdo a las siguientes fórmulas:

$$cl_l = \sum_{a \in \mathcal{A}} \zeta_a^{s_1^l} d_a^l + \sum_{a \in \mathcal{A}} \zeta_a^{s_2^l} d_a^l + kl_l \quad (2.4)$$

$$ct_l = t_l^{s_1^l} + t_l^{s_2^l} + kt_l \quad (2.5)$$

En las ecuaciones (2.4) y (2.5) los valores s_1^l y s_2^l representan las secciones que van de comienzo a fin del servicio l en sus sentidos de ida y de vuelta respectivamente. Los parámetros kl_l y kt_l representan la distancia y tiempo asociadas a las maniobras de retorno de los buses.

A partir de (2.3) se obtiene el costo por unidad de frecuencia de operar un bus, c_l , que permite obtener el costo total del operador asociado a un servicio l con frecuencia f_l . Por lo tanto, para un sistema donde opera un conjunto \mathcal{L} de servicios, los costos totales del operador se pueden expresar como la siguiente sumatoria:

$$OC = \sum_{l \in \mathcal{L}} c_l f_l \quad (2.6)$$

Es importante notar que esta función de costos del operador incorpora los costos por tamaño de flota y no es necesario contabilizarlos aparte. En otras palabras, cuando la

frecuencia de un servicio es proporcional al número de buses que requiere para operar, como sucede en el período del día donde la flota es completamente utilizada (normalmente punta mañana), los costos por aumento de flota pueden ser incorporados en el parámetro c_l de la fórmula anterior.

Por simplicidad, en lo que resta de este documento se utilizará la expresión “costos de operación” indistintamente de costos del operador, es decir, incluyendo todas las componentes de este costo, no sólo las directamente relacionadas con la operación.

- **Costos de los usuarios**

Los costos de usuarios se pueden descomponer en cuatro partes: un tiempo de acceso a la red, un tiempo de espera sumado para cada etapa de viaje, un tiempo total de viaje en vehículo, y un costo asociado a realizar trasbordos. Dado que el tiempo de acceso depende del paradero escogido para acceder y egresar del sistema, y que se asumió que la matriz de viajes es fija y conocida, esta componente del costo es constante y se puede excluir de la formulación. Los costos asociados a los otros tres componentes se detallan a continuación:

a) Costos de espera

El tiempo de espera promedio para un servicio de frecuencia f_l se suele asumir proporcional a su intervalo promedio (Mohring, 1972, Jansson, 1980):

$$wt_l = \frac{\lambda}{f_l} \quad (2.7)$$

En esta expresión el parámetro λ refleja el efecto de la aleatoriedad de los intervalos: entre menos regulares son los intervalos (por ejemplo, por efecto del

apelotonamiento de los buses) mayor se hace la espera promedio, ya que es mucho más probable llegar al paradero durante un intervalo largo que uno corto. En el caso más favorable, donde los buses llegan idénticamente espaciados al paradero, el tiempo esperado de espera es la mitad de un intervalo, es decir, $\lambda = 0,5$. En un escenario donde los buses llegan según un proceso de Poisson es posible demostrar que en promedio los usuarios esperan un intervalo completo (o bien, que $\lambda = 1$). En este último caso, la fórmula se puede extender a múltiples servicios, considerando la suma de las frecuencias de los servicios atractivos.

Si consideramos que el número de viajes a lo largo de cada sección de ruta es de V_s^w , y que el costo asociado al tiempo de espera es de θ_{wt} , es posible concluir que los costos totales de espera son de:

$$WTC = \theta_{wt} \sum_{w \in \mathcal{W}} \sum_{s \in \mathcal{S}} V_s^w \frac{\lambda}{\sum_{l \in \mathcal{L}} f_l^s} \quad (2.8)$$

En esta expresión la variable f_l^s se hace cargo de que es posible que el modelo nos indique que la línea l es atractiva para un viaje en la sección s , caso en el que f_l^s toma valor f_l (existe una restricción en el modelo que impone esta cota para f_l^s). En caso contrario, f_l^s adopta valor nulo.

b) Costos de viaje en vehículo

Como asumimos que se conoce el tiempo de viaje para un servicio l en una sección s , el tiempo esperado de viaje corresponderá al promedio ponderado de los tiempos de viajes de las líneas que el usuario está dispuesto a tomar, por la probabilidad de que dicha línea sea la primera en llegar al paradero entre las atractivas. Considerando

un costo de θ_{tt} para el tiempo de viaje, y definiendo como t_l^s al tiempo de viaje (conocido) para cada línea l en cada sección s , el costo por tiempo de viaje del sistema será:

$$TTC = \theta_{tt} \sum_{w \in \mathcal{W}} \sum_{s \in \mathcal{S}} V_s^w \frac{\sum_{l \in \mathcal{L}} t_l^s f_l^s}{\sum_{l \in \mathcal{L}} f_l^s} \quad (2.9)$$

a) Costos de trasbordo

Asumiendo que todos los trasbordos representan el mismo costo para los usuarios, independiente del paradero donde se efectúan y del par de líneas que hacen conexión, podemos calcular el costo por trasbordos como un costo unitario, θ_{tr} , multiplicado por el número de trasbordos. El número de trasbordos del sistema se puede obtener a partir de la diferencia entre el número total de secciones de viajes realizadas, y el total de viajes en la matriz origen destino:

$$TRC = \theta_{tr} \left\{ \sum_{s \in \mathcal{S}} \sum_{w \in \mathcal{W}} V_s^w - \sum_{w \in \mathcal{W}} T_w \right\} \quad (2.10)$$

Cabe señalar que el costo asociado a un trasbordo, θ_{tr} , no incluye los costos por tiempo de espera, debido a que la expresión de los tiempos de espera está formulada a nivel de sección de ruta. Los costos a considerar por trasbordo provienen de otro tipo de efectos involucrados, como el costo por caminata, por incertidumbre, o simplemente por la molestia de interrumpir un viaje.

- **Modelo de Leiva**

Con estas definiciones se puede construir la función de costos sociales del sistema, que es la función objetivo a minimizar del modelo de diseño:

$$SC = OC + WTC + TTC + TRC \quad (2.11)$$

Con esto, podemos plantear el modelo de diseño de servicios expresos como el siguiente problema de optimización:

$$\begin{aligned} \min_{f_l, f_l^s, V_s^w} & \left[\sum_{l \in \mathcal{L}} c_l f_l + \sum_{w \in \mathcal{W}} \sum_{s \in \mathcal{S}} V_s^w \left\{ \theta_{wt} \frac{\lambda}{\sum_{l \in \mathcal{L}} f_l^s} + \theta_{tt} \frac{\sum_{l \in \mathcal{L}} t_l^s f_l^s}{\sum_{l \in \mathcal{L}} f_l^s} \right\} \right. \\ & \left. + \theta_{tr} \left\{ \sum_{w \in \mathcal{W}} \sum_{s \in \mathcal{S}} V_s^w - \sum_{w \in \mathcal{W}} T_w \right\} \right] \end{aligned} \quad (2.12)$$

Sujeto a:

$$f_l^s \leq f_l, \quad \forall l \in \mathcal{L}, \forall s \in \mathcal{S} \quad (2.13)$$

$$f_l^s \geq 0, \quad \forall l \in \mathcal{L}, \forall s \in \mathcal{S} \quad (2.14)$$

$$\sum_{s \in \mathcal{S}_i^+} V_s^w - \sum_{s \in \mathcal{S}_i^-} V_s^w = \begin{cases} T_w, & \forall i \in \mathcal{P} | i = O_w \\ -T_w, & \forall i \in \mathcal{P} | i = D_w \\ 0, & e. o. c. \end{cases}, \quad \forall w \in \mathcal{W}, \forall s \in \mathcal{S}, \forall i \in \mathcal{N} \quad (2.15)$$

El modelo formado por la función objetivo (2.12) más las restricciones (2.13), (2.14) y (2.15) corresponde al modelo de optimización de frecuencias para servicios expresos de Leiva. Este modelo considera tres conjuntos de restricciones:

- Restricción de elección de líneas atractivas: La expresión (2.13) impone que un usuario que desea viajar en la sección s y que considera utilizar la línea l , su frecuencia no puede exceder la frecuencia provista para este servicio.
- Restricción de signo: Las frecuencias deben ser no negativas. Esta restricción en conjunto con la restricción (2.14) asegura que a su vez $f_l \geq 0, \forall l \in \mathcal{L}$.
- Restricción de continuidad de viajes de cada par OD: La expresión (2.15) impone al modelo la continuidad en los flujos en cada parada de la red. En esta

restricción se introducen los conjuntos \mathcal{S}_i^+ y \mathcal{S}_i^- para referirse a las secciones que salen o que entran a un nodo i . También se utiliza la notación O_w y D_w para referirse al nodo de origen y de destino de un determinado par w .

El modelo acá planteado, a pesar de minimizar costos sociales totales, es consistente con un comportamiento espontáneo de usuarios, si asumimos que éstos escogen sus rutas con el fin de minimizar sus costos esperados de viaje y que valoran los distintos atributos de la misma forma como se han dispuesto en la función objetivo. Dado que en esta versión del modelo no existe una restricción de capacidad, no hay interacción alguna entre las opciones de los distintos individuos. Así, la asignación en la que cada individuo minimiza su tiempo total esperado de viaje (incluyendo trasbordos, esperas, y viaje en vehículo) coincide con la que minimiza los costos totales del sistema. Acá radica la necesidad de asumir que los parámetros θ_{tt} , θ_{wt} y θ_{tr} son los mismos para los usuarios y para la evaluación social por parte del modelador.

La minimización de los tiempos esperados de viaje de cada usuario es consistente con el modelo de comportamiento propuesto por Chriqui y Robillard (1975). Este comportamiento consiste en que, para cada sección de ruta s , el usuario escogerá el conjunto de líneas (llamadas líneas atractivas) que le conviene esperar de forma de minimizar el tiempo esperado de su viaje. El modelo de Leiva captura este mismo efecto por medio de la incorporación de las variables f_l^s , dado que asignar frecuencia $f_l^s = f_l$ al servicio l en el arco s equivale a decir que este servicio pertenece al conjunto de líneas atractivas, y asignar $f_l^s = 0$ equivale a decir que no pertenece. Para completar este argumento, se demuestra a continuación que efectivamente las

variables f_l^s siempre toman valores extremos en el óptimo, salvo en casos de indiferencia.

Una demostración de que el comportamiento de los usuarios subyacente a este modelo equivale al comportamiento de Chriqui y Robillard (1975) donde los usuarios escogen un conjunto de líneas atractivas se presenta a continuación:

Lema 1: Existe una solución óptima para el problema de Leiva donde las frecuencias f_l^s toman valor f_l o nulo.

Demostración:

Si asumimos conocidas las variables f_l y V_s^w , la función objetivo del modelo de Leiva se puede separar a nivel de secciones de ruta s , es decir, el modelo equivale a minimizar el tiempo esperado de viaje de cada sección s que se entrega en (2.16), sujeto a las restricciones (2.13) y (2.14).

$$\min ETT_s = \theta_{wt} \frac{\lambda}{\sum_{l \in \mathcal{L}} f_l^s} + \theta_{tt} \frac{\sum_{l \in \mathcal{L}} t_l^s f_l^s}{\sum_{l \in \mathcal{L}} f_l^s} \quad (2.16)$$

La derivada de ETT_s con respecto a una frecuencia efectiva en particular es la siguiente:

$$\frac{\partial ETT_s}{\partial f_l^s} = \frac{\theta_{tt} \sum_{k \in \mathcal{L}} (t_l^s - t_k^s) f_k^s - \lambda \theta_{wt}}{(\sum_{l \in \mathcal{L}} f_l^s)^2} \quad (2.17)$$

El numerador de la expresión (2.17) no depende de f_l^s , ya que cuando $k = l$ el término de la sumatoria se anula. Esto implica que el signo de la derivada es el mismo en todo el dominio, es decir, que la función objetivo (2.16) es

monótonamente creciente cuando el numerador de (2.17) es positivo y que encuentra su mínimo cuando f_l^s es máximo ($f_l^s = f_l$), y que es monótonamente decreciente cuando este signo es negativo y que la función se minimiza con $f_l^s = 0$.

Cuando el numerador de (2.17) se anula, estamos en presencia de un servicio que resulta indiferente para el usuario, es decir, que el tiempo esperado de viaje resulta el mismo si éste se incluye o excluye del conjunto de líneas atractivas. En este caso se puede asumir arbitrariamente que $f_l^s = 0$, valor que lleva al óptimo de la función objetivo.

De esta forma, hemos probado que para cualquier valor de f_l y V_s^w el modelo, al minimizar los costos de los usuarios en la función objetivo, construye un conjunto de líneas atractivas para cada sección de ruta s , representando un comportamiento de usuarios acorde al planteado por Chriqui.

- **Restricción de capacidad**

El modelo de Leiva planteado en la sección anterior no considera la capacidad de los vehículos de ninguna manera. Un posible enfoque para asegurar que la solución posea capacidad suficiente para transportar a los pasajeros consiste directamente en incorporar al modelo una restricción de capacidad del tipo:

$$cap_l f_l \geq \sum_{w \in \mathcal{W}} \sum_{s \in \mathcal{S}} V_s^w \zeta_a^s \frac{f_l^s}{\sum_{i \in \mathcal{L}} f_i^s}, \quad \forall l \in \mathcal{L}, \forall a \in \mathcal{A} \quad (2.18)$$

Desafortunadamente, esta forma de incorporar la restricción de capacidad puede llevar a soluciones que no respetan la racionalidad del comportamiento de los usuarios. En ausencia de límites para la capacidad, el óptimo del sistema (u óptimo social) coincide con el equilibrio de usuarios ya que cada uno de ellos puede optimizar individualmente sus costos sin estorbarse unos a otros. Desafortunadamente, y como ya se explicó antes, al restringir la capacidad la asignación de los usuarios que resulta de minimizar los costos sociales no se condice con la que se obtiene cuando los usuarios velan por sus propios intereses. En el artículo de Leiva et al. (2010) se propone un enfoque alternativo para abordar el problema de la capacidad sin atentar contra el modelo de comportamiento. El método consiste en agregar una cota inferior para la frecuencia de cada servicio en las restricciones como se muestra en (2.19), e ir sucesivamente elevando estas cotas en los puntos donde se ve violada la capacidad máxima de los vehículos, hasta que la restricción (2.18) sea satisfecha en todo el corredor, para todos los servicios. Esta heurística la llamaremos HC_1 en el resto de este trabajo.

$$\check{f}_l \leq f_l, \quad \forall l \in \mathcal{L} \quad (2.19)$$

Heurística para capacidad HC_1 :

1. Inicializar las cotas en $\check{f}_l = 0$.
2. Resolver el problema (2.12) - (2.15) con las restricciones adicionales (2.19). Si la solución encontrada no presenta arcos con insuficiencia de capacidad, detenerse.

3. Para la solución actual, escoger una línea \hat{l} donde exista un déficit de capacidad, es decir donde la inecuación (2.18) no es satisfecha. El criterio de selección será escoger la línea que presente el arco donde este déficit sea mayor, es decir:

$$\hat{l} = \operatorname{argmax}_{l \in \mathcal{L}} \left[\max_{a \in \mathcal{A}} \left(\sum_{w \in \mathcal{W}} \sum_{s \in \mathcal{S}} V_s^w \zeta_a^s \frac{f_l^s}{\sum_{i \in \mathcal{L}} f_i^s} - \operatorname{cap}_l f_l \right) \right] \quad (2.20)$$

4. Aumentar en un incremento previamente definido la cota mínima para la frecuencia de la línea \hat{l} . Dado que la frecuencia actual de dicha línea es $f_{\hat{l}}^*$, el aumento de cota a realizar es $\check{f}_{\hat{l}} \rightarrow f_{\hat{l}}^* + \Delta$.
5. Volver al paso 2.

Dado que este enfoque no restringe las capacidades, la solución a la que converge este procedimiento es coherente con un equilibrio de usuarios. Sin embargo, este método no garantiza optimalidad social, sino que simplemente hace una búsqueda de una solución factible en torno a la solución óptima del problema relajado. A pesar de esto, y como se puede apreciar en los análisis realizados en Leiva et al. (2010), este método permite encontrar en muchos casos una solución que reduce el costo social al compararlo con el correspondiente a un único servicio regular optimizado.

Otra debilidad de este enfoque es que, para obtener soluciones con una precisión razonable, se debe resolver el modelo base un gran número de veces. Esto puede alargar considerablemente los tiempos de solución del problema, haciéndolo inviable para instancias de mayor tamaño o complejidad como lo sería una red de transporte público real.

En la ecuación (2.18), el parámetro de capacidad de los buses se considera dependiente del servicio, por eso lleva el subíndice l . Esto es así porque Leiva, como se mencionó antes, plantea una variante del modelo que considera la posibilidad de utilizar buses de diferentes capacidades y costos asociados. Sin embargo, el modelo planteado no es el idóneo para resolver el problema con buses distintos, ya que en la obtención de la solución inicial con restricción de capacidad relajada el modelo no percibe los beneficios de trabajar con buses de mayor capacidad (que normalmente serán más caros) y tenderá a utilizar buses pequeños. Por su parte, el algoritmo de ajuste de capacidad que se plantea no es capaz de corregir esta situación, ya que sólo aumenta las frecuencias de servicios que tienen algún nivel de utilización, lo que no sucederá cuando se entregue la opción de utilizar distintos tamaños de buses en servicios de un mismo trazado.

El problema anterior podría ser abordado mediante modificaciones sobre el algoritmo de ajuste de capacidad. Una forma sencilla (pero quizás ineficiente en términos de tiempos de solución) consistiría en resolver primero el problema sólo con buses pequeños, y sobre la solución que se encuentre ir secuencialmente tomando el servicio de mayor frecuencia (siempre y cuando opere a capacidad), cambiar el tipo de bus que utiliza actualizando cap_l y c_l , y volviendo a optimizar desde cero. De todas formas, por un tema de simplicidad, en lo que resta de esta tesis se trabajará sólo con buses idénticos, de forma que la capacidad será denotada desde ahora simplemente por cap .

Hasta el momento, se ha dado por entendido que al hablar de capacidad en este contexto nos estamos refiriendo a la capacidad de los vehículos de la flota. Sin

embargo, existen otros recursos que pueden resultar limitantes para el diseño. Por una parte existen las limitaciones asociadas a la operación en paraderos, tales como la capacidad de los andenes para atender buses, la capacidad de carga y descarga de los buses, o la capacidad de los paraderos para albergar pasajeros.

Otro recurso que puede imponer una restricción al modelo consiste en la capacidad de las vías. En un corredor exclusivo, donde los buses no compiten por capacidad con vehículos particulares, el cuello de botella en términos de capacidad vial a menudo se encontrará en las intersecciones semaforizadas.

En la segunda sección del presente capítulo se discuten algunas modificaciones que se pueden realizar al modelo de Leiva y al algoritmo HC_1 para incorporar restricciones de capacidad al modelo.

- **Otras propiedades del modelo de Leiva**

El problema de optimización de frecuencias tiene algunas propiedades analíticas que podrían resultar de interés para el desarrollo de nuevos algoritmos u otro tipo de avances en el problema. Tres de estas propiedades se demuestran a continuación.

a) Igualdad de costos del operador y de tiempos de espera en el óptimo

En diversas implementaciones del modelo se pudo constatar una propiedad interesante de sus soluciones: cuando la capacidad no es limitante, en la solución los costos totales del operador OC se igualan con los de espera WTC . Para demostrar este hecho, primero estudiaremos el problema donde no existen trasbordos, cuya función objetivo se muestra en la fórmula (2.21) a continuación:

$$SC = \sum_{l \in \mathcal{L}} f_l c_l + \theta_{tt} \sum_{w \in \mathcal{W}} T_w \left(\frac{\sum_{l \in \mathcal{L}} \delta_l^w f_l t_l^w}{\sum_{l \in \mathcal{L}} \delta_l^w f_l} \right) + \theta_{wt} \sum_{w \in \mathcal{W}} T_w \left(\frac{\lambda}{\sum_{l \in \mathcal{L}} \delta_l^w f_l} \right) \quad (2.21)$$

Al eliminar los trasbordos, se vuelve innecesario trabajar con secciones de ruta, y la función objetivo se puede expresar a nivel de pares OD. Es por esto que en lugar de V_s^w se trabaja directamente con T_w en la ecuación anterior, y en general los subíndices de sección de ruta s son reemplazados directamente por pares w . Para facilitar la demostración, se han reemplazado las frecuencias efectivas f_l^w por $\delta_l^w f_l$, donde la variable binaria δ_l^w representa si el servicio l es atractivo para el par w . Este último reemplazo es válido como consecuencia del Lema 1, demostrado anteriormente. El problema de optimización asociado a (2.21) para estar completo debe considerar además restricciones de no negatividad en sus frecuencias, y de binariedad para las variables δ_l^w .

Lema 2: En el óptimo del problema definido por (2.21) considerando restricciones de signo y binariedad, los costos de operación se igualan con los costos de espera.

Demostración:

Asumiremos por un momento que todas las líneas poseen frecuencia positiva en el óptimo. Esto implica que, en el óptimo:

$$\frac{\partial SC}{\partial f_i} = 0, \quad \forall i \in \mathcal{L} \quad (2.22)$$

Multiplicando ambos lados de la ecuación por f_i y sumando para todas las líneas se obtiene que:

$$\sum_{i \in \mathcal{L}} \frac{\partial SC}{\partial f_i} f_i = 0 \quad (2.23)$$

Ahora bien,

$$\begin{aligned} \frac{\partial SC}{\partial f_i} = c_i + \theta_{tt} \sum_{w \in \mathcal{W}} T_w \delta_i^w \left(\frac{t_i^w}{\sum_{l \in \mathcal{L}} \delta_l^w f_l} - \frac{\sum_{l \in \mathcal{L}} \delta_l^w f_l t_l^w}{(\sum_{l \in \mathcal{L}} \delta_l^w f_l)^2} \right) \\ - \theta_{wt} \sum_{w \in \mathcal{W}} T_w \frac{\lambda \delta_i^w}{(\sum_{l \in \mathcal{L}} \delta_l^w f_l)^2} \end{aligned} \quad (2.24)$$

Reemplazando (2.24) en (2.23):

$$\begin{aligned} \sum_{i \in \mathcal{L}} c_i f_i + \underbrace{\sum_{i \in \mathcal{L}} f_i \theta_{tt} \sum_{w \in \mathcal{W}} T_w \delta_i^w \left(\frac{t_i^w}{\sum_{l \in \mathcal{L}} \delta_l^w f_l} - \frac{\sum_{l \in \mathcal{L}} \delta_l^w f_l t_l^w}{(\sum_{l \in \mathcal{L}} \delta_l^w f_l)^2} \right)}_* \\ - \sum_{i \in \mathcal{L}} f_i \theta_{wt} \sum_{w \in \mathcal{W}} T_w \frac{\lambda \delta_i^w}{(\sum_{l \in \mathcal{L}} \delta_l^w f_l)^2} = 0 \end{aligned} \quad (2.25)$$

Desarrollando el término marcado con asterisco:

$$* = \theta_{tt} \sum_{w \in \mathcal{W}} T_w \left(\frac{\sum_{i \in \mathcal{L}} \delta_i^w f_i t_i^w}{\sum_{l \in \mathcal{L}} \delta_l^w f_l} - \frac{\sum_{i \in \mathcal{L}} \delta_i^w f_i \cdot \sum_{l \in \mathcal{L}} \delta_l^w f_l t_l^w}{(\sum_{l \in \mathcal{L}} \delta_l^w f_l)^2} \right) \quad (2.26)$$

$$* = \theta_{tt} \sum_{w \in \mathcal{W}} T_w \left(\frac{\sum_{i \in \mathcal{L}} \delta_i^w f_i t_i^w}{\sum_{l \in \mathcal{L}} \delta_l^w f_l} - \frac{\sum_{l \in \mathcal{L}} \delta_l^w f_l t_l^w}{\sum_{l \in \mathcal{L}} \delta_l^w f_l} \right) = 0 \quad (2.27)$$

Por lo tanto, la ecuación (2.25) se puede reescribir como:

$$\sum_{i \in \mathcal{L}} c_i f_i - \theta_{wt} \sum_{w \in \mathcal{W}} T_w \frac{\lambda \sum_{i \in \mathcal{L}} f_i \delta_i^w}{(\sum_{l \in \mathcal{L}} \delta_l^w f_l)^2} = 0 \quad (2.28)$$

$$\sum_{i \in \mathcal{L}} c_i f_i = \theta_{wt} \sum_{w \in \mathcal{W}} T_w \frac{\lambda}{\sum_{l \in \mathcal{L}} \delta_l^w f_l} \quad (2.29)$$

El lado izquierdo de la última ecuación corresponde a los costos del operador (2.6). El lado derecho corresponde a su vez a los costos de espera del usuario.

Para probar el Lema 2 falta generalizar esta propiedad para cuando los servicios pueden presentar frecuencias nulas en el óptimo. Sin embargo, para cualquier conjunto de líneas iniciales podemos construir un problema equivalente que considere \mathcal{L} como el conjunto de las líneas con frecuencia óptima positiva del problema original. La solución de este problema equivalente cumple con la propiedad demostrada, y posee costos equivalentes al problema original, lo que prueba que ambos deben cumplir con la igualdad demostrada. Con esto queda demostrado el Lema 2.

El Lema 2 acá demostrado es válido para problemas sin trasbordo. Este resultado se puede extender para el caso con trasbordos haciendo una conversión: supongamos que conocemos, para el problema con trasbordos, una asignación válida de usuarios sobre el corredor, es decir, un conjunto de valores para las variables V_s^w que satisface la restricción de continuidad (2.15). El problema entonces puede ser convertido al del Lema 2 simplemente tomando como nueva matriz de viajes la matriz resultante de separar todos los tramos de los viajes asignados como si fueran viajes de usuarios distintos (en este cambio de problema se deja fuera el costo por trasbordos, pero no es relevante ya que es fijo para una asignación dada). En esta nueva asignación sin trasbordos las frecuencias óptimas cumplirán la condición de igualdad (2.29) para cualquier asignación que se escoja, y en particular también para la asignación óptima. Con esto queda demostrada la propiedad para el caso de un corredor con trasbordos.

b) Umbrales de comportamiento de usuarios

En algunas ocasiones se desea saber qué condiciones deben cumplirse para que un servicio expreso sea atractivo por sí solo, es decir, que los usuarios estarán dispuestos a dejar pasar los servicios más lentos para esperar al servicio expreso. Esta condición resulta útil, por ejemplo, cuando se optimiza analíticamente la frecuencia de un servicio (como cuando se construyen servicios en el Capítulo 3), ya que permite realizar simplificaciones sobre el comportamiento de usuarios. El Lema 3 a continuación propone un umbral para la frecuencia del servicio expreso que determina cuándo puede ser atractivo por sí solo.

Lema 3: En un corredor donde existe un servicio regular y un expreso este último será atractivo por sí solo en una sección de ruta s cuando su frecuencia cumpla que:

$$f_e \geq \frac{\theta_{wt}\lambda}{\theta_{tt}(t_a^s - t_e^s)} \quad (2.30)$$

Demostración:

Consideremos una sección de ruta atendida por dos servicios, un servicio regular y un servicio expreso (representado por los subíndices a y e). Los costos de los usuarios, al no considerar trasbordos, corresponden a la suma de los costos de espera y de tiempos de viaje. Existen dos estrategias razonables para los usuarios: tomar cualquiera de los servicios que pase primero (opción 1), o esperar a que pase el servicio expreso (opción 2). Esperar al servicio

regular no tiene sentido ya que por definición el servicio expreso es más rápido.

Los costos esperados de viaje para la opción 1 son:

$$ETT_s = \frac{\theta_{tt}f_a t_a^s + \theta_{tt}f_e t_e^s + \theta_{wt}\lambda}{f_a + f_e} \quad (2.31)$$

Los costos de la opción 2 son:

$$ETT_s = \theta_{tt}t_e^s + \frac{\theta_{wt}\lambda}{f_e} \quad (2.32)$$

El servicio expreso será atractivo por si solo cuando los costos de la primera opción (2.31) sean mayores o iguales que los de la segunda opción (2.32), es decir:

$$\frac{\theta_{tt}f_a t_a^s + \theta_{tt}f_e t_e^s + \theta_{wt}\lambda}{f_a + f_e} \geq \theta_{tt}t_e^s + \frac{\theta_{wt}\lambda}{f_e} \quad (2.33)$$

Desarrollando la desigualdad se obtiene la condición a demostrar.

Este resultado puede seguir siendo desarrollado cuando los tiempos de viaje en vehículo responden a la estructura mostrada en la expresión (2.1), ya que en este caso el término $(t_a^s - t_e^s)$ que aparece en (2.30) puede ser reemplazado por τX_s , donde X_s es el número de detenciones omitidas por el servicio expreso en s . Para estos casos, entonces, se puede asegurar que el servicio expreso es atractivo por sí solo cuando el número de paraderos omitidos cumple que:

$$X_s \geq \frac{\theta_{wt}\lambda}{\theta_{tt}f_e\tau} \quad (2.34)$$

c) Multiplicidad de óptimos locales

Una fuente de complejidad del problema de optimización de frecuencias proviene de la existencia de múltiples óptimos locales. Esto sucede, a grandes rasgos, por dos motivos. Por una parte, cuando el problema es alimentado de un número alto de servicios, existen soluciones que utilizan diferentes conjuntos de servicios que pueden ser óptimas localmente, lo que en cierta forma le da una dimensión combinatorial a la solución del problema (ya que distintas combinaciones de servicios llevan a distintos óptimos locales). Por otra parte, incluso cuando se trabaja con pocos servicios, los distintos tipos de comportamiento también pueden generar multiplicidad de óptimos locales. Por ejemplo, un servicio expreso que puede no ser beneficioso para frecuencias bajas, ya que sus ahorros son percibidos por una fracción pequeña de los usuarios al no ser atractivo por sí solo, puede resultar beneficioso para frecuencias más altas.

En esta sección demostraremos la existencia de múltiples óptimos locales construyendo un ejemplo sencillo del segundo tipo mencionado. Para esto, supondremos que un corredor es atendido por dos servicios: un servicio regular de frecuencia f_a y costo de operación por ciclo c_a , y un servicio expreso que atiende solo los paraderos extremos del corredor de frecuencia f_e y costo de operación c_e . La demanda total del par atendido por el servicio expreso es de T_e , y la del resto del corredor, que sólo tiene acceso al servicio regular, es de T_a . En el sentido de regreso del corredor no existe demanda. Los tiempos de viaje de extremo a extremo para ambos servicios son t_a y t_e . Asumiremos que

no existen trasbordos en el corredor (o, equivalentemente, que θ_{tr} es muy alto).

La situación descrita se ilustra en la Figura 2-2.

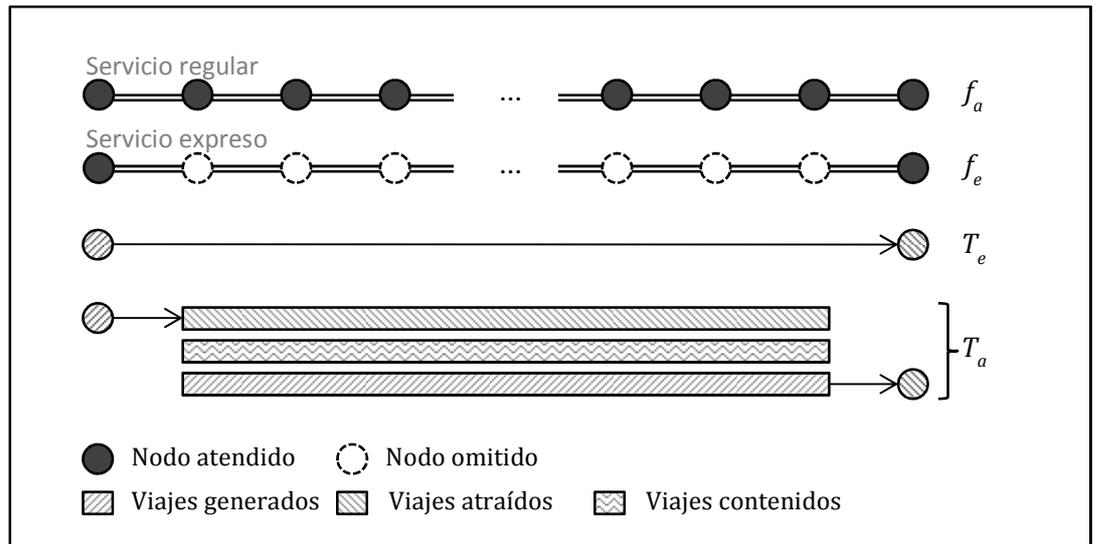


Figura 2-2: Escenario para múltiples óptimos locales

La figura anterior muestra la configuración por paraderos y la demanda para el corredor. Como se puede apreciar, T_e corresponde al número de viajes que se generan en el nodo inicial y que poseen como destino el nodo final, y T_a contiene al resto de los viajes.

Los costos de operación de este escenario, de acuerdo a la fórmula (2.6), serán los siguientes:

$$SC = f_a c_a + f_e c_e \quad (2.35)$$

Del Lema 3 podemos desprender que existe un umbral \hat{f}_e para la frecuencia del servicio expreso, bajo el cual los usuarios de los pares extremos utilizarán el primer servicio en pasar, y que al ser superado los usuarios optarán por esperar

únicamente al servicio expreso. Ese umbral es el que se entrega en la fórmula (2.30) de la sección anterior.

Con esto, podemos plantear la función de costos sociales como una función por casos. Los tiempos de viaje para flujos fuera de los pares extremos son constantes, por lo que se eliminan de la función de costos. Para simplificar la formulación de los tiempos de viaje en vehículo se trabajará con los ahorros en costo de tiempo de viaje en vez de los costos netos, que llevan a un problema equivalente (la diferencia entre ambos es una constante). Con esto, los costos de los usuarios, que se desprenden de (2.8) y (2.9), son:

$$UC = \begin{cases} T_a \frac{\theta_{wt}\lambda}{f_a} + T_e \frac{\theta_{wt}\lambda}{f_a + f_e} - T_e(t_a - t_e) \frac{\theta_{tt}f_e}{f_a + f_e}, & f_e \leq \hat{f}_e \\ T_a \frac{\theta_{wt}\lambda}{f_a} + T_e \frac{\theta_{wt}\lambda}{f_e} - T_e(t_a - t_e)\theta_{tt}, & f_e > \hat{f}_e \end{cases} \quad (2.36)$$

El problema para determinar las frecuencias óptimas para este caso corresponderá a minimizar los costos sociales, definidos como la suma de SC y UC , sujeto a que ambas frecuencias deben ser no negativas.

Una forma de que este problema presente más de un óptimo local sería que exista un óptimo local diferente para cada tipo de comportamiento. Para que estos óptimos sean efectivamente distintos se debe verificar además que no ocurran ambos en la frontera de las dos regiones definidas. En el caso donde $f_e > \hat{f}_e$ la función objetivo es separable para f_a y f_e , lo que es consecuencia de que ambos servicios atienden su propia porción de la demanda sin interactuar. Derivando la función de costos sociales, igualando a cero y despejando se obtienen los valores de f_a^* y f_e^* para este caso:

$$f_a^* = \sqrt{\frac{\lambda\theta_{wt}T_a}{c_a}} \quad (2.37)$$

$$f_e^* = \sqrt{\frac{\lambda\theta_{wt}T_e}{c_e}} \quad (2.38)$$

Se puede notar que las frecuencias óptimas para este caso corresponden, naturalmente, a la que se obtendría a partir de la fórmula de la raíz de Mohring (Mohring, 1972) considerando ambos servicios como si operaran sobre corredores separados. Como se desea poder asegurar que la frecuencia óptima aparezca dentro del sector donde ocurre este tipo de comportamiento, podemos asegurar que existe un óptimo local donde los usuarios esperan exclusivamente al expreso cuando $f_e^* > \hat{f}_e$.

Para el otro tipo de comportamiento, es decir, para el caso en que $f_e > \hat{f}_e$, evaluamos las derivadas parciales de la función objetivo:

$$\frac{\delta SC}{\delta f_a} = c_a - T_a \frac{\theta_{wt}\lambda}{f_a^2} - T_e \frac{\theta_{wt}\lambda}{(f_a + f_e)^2} - T_e(t_a - t_e) \frac{\theta_{tt}f_e}{(f_a + f_e)^2} \quad (2.39)$$

$$\frac{\delta SC}{\delta f_e} = c_e - T_e \frac{\theta_{wt}\lambda}{(f_a + f_e)^2} - T_e(t_a - t_e) \frac{\theta_{tt}f_a}{(f_a + f_e)^2} \quad (2.40)$$

El mínimo de la función objetivo para este lado del dominio se podría obtener igualando las dos últimas derivadas a cero y resolviendo el sistema de ecuaciones. Desafortunadamente, este procedimiento no lleva a una solución explícita.

Para probar la existencia de múltiples óptimos locales se entrega a continuación una instancia del problema que cumple con esta condición. Consideremos para el problema planteado los siguientes datos:

Tabla 2-1: Parámetros de la instancia con múltiples óptimos locales

Parámetro	Valor
T_a	1.100 pax/hr
T_e	400 pax/hr
tt_a	60 min
tt_e	50 min
c_a	2.500 \$/bus-hr
c_e	2.400 \$/bus-hr
θ_{tt}	15 \$/min
θ_{wt}	25 \$/min
λ	1

A partir de las fórmulas (2.30) y (2.43) podemos concluir que $\hat{f}_e = 10 \text{ bus/hr}$ y que sobre este umbral las frecuencias óptimas son $f_a^* = 25,7 \text{ bus/hr}$ y $f_e^* = 15,8 \text{ bus/hr}$, lo que corresponde a un óptimo local del problema con un costo asociado de 146.977,8 \$/hr. Para $f_e \leq \hat{f}_e$ se puede resolver numéricamente el sistema de ecuaciones que se obtiene de igualar a cero las expresiones (2.45) y (2.46), del que se obtiene el segundo óptimo local del problema correspondiente a $f_a^* = 28,5 \text{ bus/hr}$ y $f_e^* = 2,5 \text{ bus/hr}$ con un costo de 149.660,8 \$/hr. En la figura a continuación se muestran los costos sociales del problema en función de las frecuencias de ambos servicios.

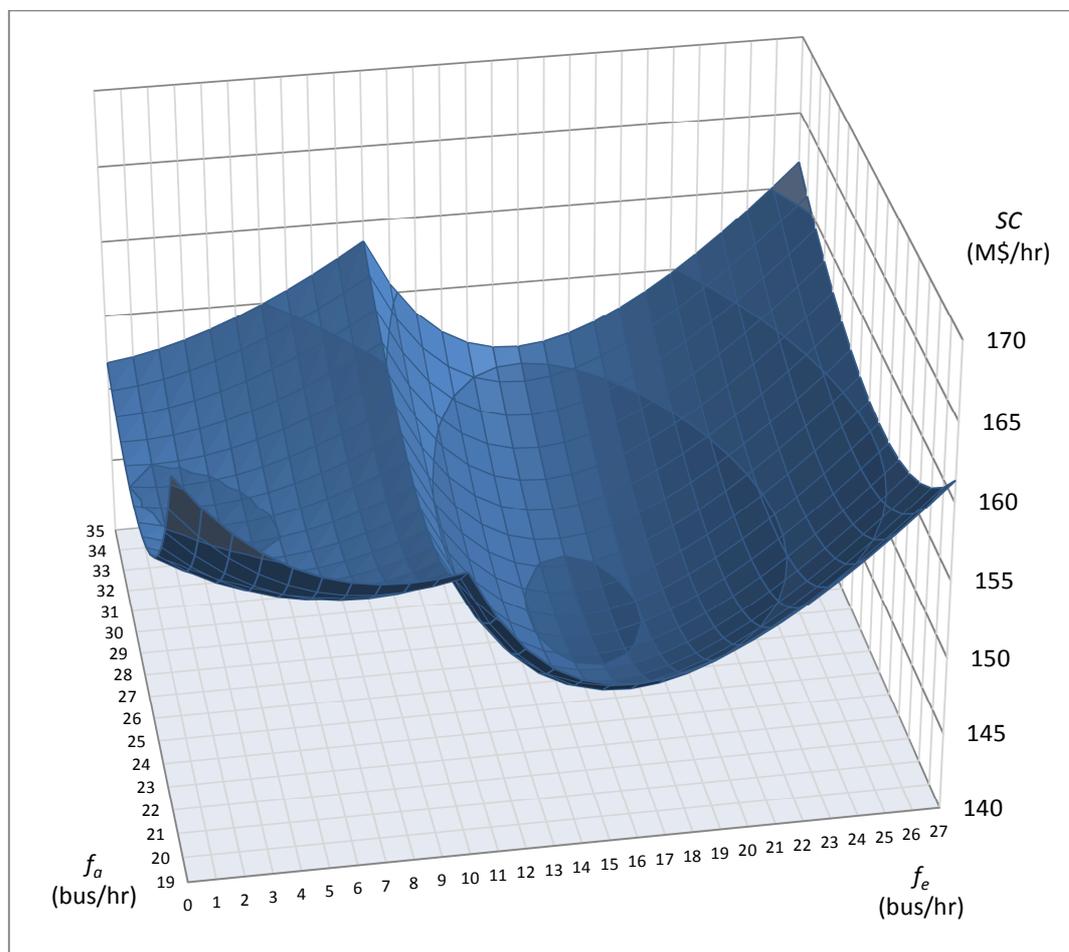


Figura 2-3: Función objetivo con múltiples óptimos locales

En el gráfico se puede apreciar con claridad la existencia de los dos óptimos locales mencionados antes. El gráfico permite ver además claramente el umbral donde ocurre el cambio de comportamiento, en el que se puede apreciar cómo la función objetivo del problema sufre un cambio discreto de pendiente, pero que no deja de ser continua. Esto, como se menciona antes, se debe a que esta frontera entre tipo de comportamientos se define precisamente como un punto de indiferencia donde los costos de los usuarios para dos opciones se igualan.

Esto puede ser entendido también de la siguiente forma: en cada par, los usuarios eligen el conjunto de servicios que minimizan su comportamiento. Si imaginamos entonces que cada comportamiento tiene asociada una función propia de costos para el dominio completo, la función resultante donde los usuarios se comportan racionalmente correspondería al mínimo de todas estas funciones en cada punto, lo que se puede también apreciar para este caso en el gráfico.

La tabla a continuación muestra un desglose de los componentes del costo social para ambos óptimos locales, y para el óptimo sin servicio expreso.

Tabla 2-2: Desglose de costos

Instancia	f_a (bus/hr)	f_e (bus/hr)	OC (\$/hr)	WTC (\$/hr)	TTC (\$/hr)	SC (\$/hr)
Óptimo local 1	25,7	15,8	102.173,5	102.173,5	-60.000,0	144.347,0
Óptimo local 2	28,5	2,5	77.284,6	77.284,7	-4.908,4	149.660,8
Óptimo s/ expreso	30,0	0,0	75.000,0	75.000,0	0,0	150.000,0

Se puede verificar, en primer lugar, que los costos de espera se igualan con los del operador en ambos óptimos locales del problema tal como se desprende del Lema 1. Esta propiedad también se cumple en el caso sin expresos, ya que el óptimo de ese caso es el que se obtiene definiendo el problema alimentando al modelo sólo del servicio regular, que es una instancia particular del modelo y por lo tanto también cumple sus propiedades. Un hecho que llama la atención es que en el óptimo global los costos del operador suben con respecto a las otras soluciones. Esto sucede porque los ahorros que se obtiene en los tiempos

de viaje cuando el servicio expreso es atractivo compensan un aumento en las otras componentes del costo.

2.2 Mejoras y variantes del modelo de Leiva

El modelo de Leiva es una buena herramienta para estudiar el problema de diseño en corredores de tamaño moderado y permite hacer varios análisis interesantes. Sin embargo, el modelo presenta tres principales limitaciones que dificultan su implementación en casos más complejos y realistas.

La primera de estas limitaciones es la ineficiencia del modelo para la solución de problemas con congestión (es decir, donde la capacidad de los vehículos se ve alcanzada en el óptimo). La heurística que se propone para la búsqueda de soluciones factibles para congestión es altamente ineficiente en términos del número de iteraciones, lo que si puede ser un inconveniente cuando se desea obtener las frecuencias óptimas de un escenario en particular, puede llegar a ser inviable si se desea optimizar frecuencias como parte de un modelo más complejo en más niveles. Una alternativa al algoritmo de ajuste de capacidades es presentada más adelante en esta sección.

La segunda limitación, también relacionada con el procedimiento de ajuste de capacidad, es que las soluciones que obtiene no aseguran optimalidad. La solución para el problema relajado que entrega el modelo es óptima, pero al agregar cotas al problema se restringe el dominio en forma arbitraria, y se excluyen de éste soluciones que potencialmente podrían entregar más ahorros que la solución a la que converge el algoritmo. De todas formas, a pesar de no asegurar optimalidad, las

pruebas realizadas con el modelo alcanzan buenas soluciones que representan ahorros con respecto al caso base donde no operan los servicios expresos.

La tercera limitación del modelo, desde el punto de vista del problema de diseño de servicios expresos, es que no genera sus propios servicios sino que depende del criterio del planificador para obtener buenas soluciones. En el Capítulo 3 de esta tesis se presenta una heurística que incorpora esta etapa al modelo de diseño.

En lo que sigue de esta sección se presentan algunas propuestas de mejora para el modelo de optimización de frecuencias orientadas a hacer más eficiente su solución, o a incorporar dimensiones del problema que no son capturadas en su formulación original.

- **Relajación de exigencia de servicios cíclicos**

El modelo de Leiva trabaja con servicios que son forzosamente cíclicos, es decir, que terminan en el lugar donde comienzan. Sin embargo, esta condición puede ser relajada, siempre y cuando se tenga cuidado de que las frecuencias de los servicios obtenidos sean factibles en términos de continuidad de frecuencias en cada punto del corredor. Definir servicios por tramos por separado puede ser conveniente porque permite considerar un mayor número de combinaciones de ida y vuelta entregando menos servicios al modelo, y porque facilita la generación de servicios, como se muestra en el Capítulo 3.

Cuando el problema se restringe a un corredor, es decir, cuando la matriz de viajes se puede separar entre un sentido de ida y uno de vuelta, es posible entregar al modelo los servicios codificados para cada sentido por separado. Supondremos inicialmente

que los paraderos del corredor son $\mathcal{N} = \{1, \dots, r, r + 1, \dots, N\}$, donde los paraderos de 1 a r corresponden al sentido de ida (subconjunto que denotaremos como \mathcal{N}_1), y de $r + 1$ hasta N representan el sentido de vuelta, \mathcal{N}_2 . Si asumimos que todos los servicios a considerar van de 1 a r o de $r + 1$ a N (es decir, que no existen servicios cortos), y que los costos asociados a la maniobra de retorno son despreciables, la única modificación necesaria para que el modelo esté completo sería agregar una restricción que asegure que la frecuencia total en el sentido de ida coincida con la frecuencia total ofrecida en el sentido de vuelta, es decir:

$$\sum_{l \in \mathcal{L}_1} f_l = \sum_{l \in \mathcal{L}_2} f_l \quad (2.41)$$

En esta última fórmula se utilizan los conjuntos \mathcal{L}_1 y \mathcal{L}_2 para representar a los servicios que operan en cada sentido en forma respectiva. Un supuesto implícito en esta formulación del modelo es que cualquier bus puede operar en cualquier servicio, aunque evidentemente es posible extender esta formulación imponiendo una restricción para cada tipo de bus.

Este tipo de formulación implica adaptar la forma en que se calculan los costos de operación de los buses. En este caso la fórmula (2.1) todavía aplica, pero es necesario reemplazar el largo y el tiempo de ciclo del servicio, cl_l y ct_l , por el largo y el tiempo del servicio desde su nodo inicial hasta su nodo final.

Esta formulación se puede ampliar para considerar servicios cortos (es decir, que comienzan o terminan en paraderos intermedios del corredor). Para esto asumiremos por ahora que el número de paraderos en el sentido de ida coincide con el número de paraderos en el sentido de vuelta ($2r = N$), y que cada paradero de ida corresponde

geográficamente con un paradero de vuelta (el paradero i correspondería al paradero $N + 1 - i$, $\forall i \in \mathcal{N}$). Utilizando los parámetros ϑ_{ii}^+ y ϑ_{ii}^- , que toman valor 1 cuando la línea l comienza o termina en el paradero i , respectivamente, las restricciones de continuidad de frecuencia se pueden expresar como:

$$\sum_{l \in \mathcal{L}} \vartheta_{ii}^+ f_l = \sum_{l \in \mathcal{L}} \vartheta_{i(N+1-i)}^- f_l, \quad \forall i \in \mathcal{N}_1 \quad (2.42)$$

Una versión más general de este modelo consiste en no limitar las paradas a coincidir geográficamente, y permitir que un servicio que termina en un paradero viaje hasta un paradero en el sentido contrario para iniciar su servicio. Esta modificación, además de ser más realista con respecto a la configuración de los paraderos, permite soluciones que la versión anterior descarta incluso cuando los paraderos coinciden. Más aún, esta formulación permite incorporar el tiempo que tarda el bus en sus maniobras de retorno para retomar el servicio en sentido opuesto. Para esto es necesario incorporar al análisis el costo de mover el bus entre dos paraderos i y j de sentidos opuestos, que llamaremos c_{ij} , y que en forma análoga a la fórmula (2.1) se puede estimar como:

$$c_{ij} = (c_L l_{ij} + c_T t_{ij}) \quad (2.43)$$

En esta expresión l_{ij} y t_{ij} corresponden a la distancia que debe recorrer y al tiempo que se demora el bus al realizar sus maniobras de retorno entre los paraderos i y j . Además, se hace necesario incluir nuevas variables que representan la frecuencia que se mueve entre cada par de nodos. Estas variables las llamaremos, para los movimientos de ida hacia vuelta y al revés, f_{ij} y f_{ji} , con $i \in \mathcal{N}_1$ y $j \in \mathcal{N}_2$. Las

restricciones de continuidad de frecuencias para este caso se muestran a continuación:

$$\sum_{l \in \mathcal{L}} \vartheta_{li}^+ f_l + \sum_{j \in \mathcal{N}_2} f_{ij} = \sum_{l \in \mathcal{L}} \vartheta_{li}^- f_l + \sum_{j \in \mathcal{N}_2} f_{ji}, \quad \forall i \in \mathcal{N}_1 \quad (2.44)$$

$$\sum_{l \in \mathcal{L}} \vartheta_{li}^+ f_l + \sum_{j \in \mathcal{N}_1} f_{ij} = \sum_{l \in \mathcal{L}} \vartheta_{li}^- f_l + \sum_{j \in \mathcal{N}_1} f_{ji}, \quad \forall i \in \mathcal{N}_2 \quad (2.45)$$

Las nuevas variables del tipo f_{ij} deben ser positivas, lo que hace necesario agregar restricciones de no negatividad:

$$f_{ij} \geq 0, \quad \forall i \in \mathcal{N}_1, \forall j \in \mathcal{N}_2 \quad (2.46)$$

$$f_{ji} \geq 0, \quad \forall i \in \mathcal{N}_1, \forall j \in \mathcal{N}_2 \quad (2.47)$$

Dado que los servicios ahora incurren en un trayecto adicional, se deben modificar los costos de operador en la función objetivo:

$$OC = \sum_{l \in \mathcal{L}} c_l f_l + \sum_{i \in \mathcal{N}_1, j \in \mathcal{N}_2} (c_{ij} f_{ij} + c_{ji} f_{ji}) \quad (2.48)$$

El tipo de servicio que este modelo podría diseñar es comparable con el formato de operación propuesto y optimizado en Cortéz et al. (2011), que definen como una combinación entre servicio corto, debido a que los servicios pueden cubrir el corredor parcialmente si resulta beneficioso, y *deadheading*, ya que pueden poseer un tramo donde operan sin entregar servicio para comenzar su operación en sentido contrario (o para volver al comienzo del servicio en el mismo sentido, en caso extremo).

Como se explicaba anteriormente, estas variaciones permiten modelar un número mayor de posibles servicios a partir de un número reducido de tramos que se

combinan. Esto se traduce en una formulación más eficiente, capaz de abarcar un dominio mayor para el mismo número de servicios a optimizar, que lleva a mejores soluciones a su vez. Las variantes acá presentadas para el modelo son progresivamente más flexibles, en el sentido de que consideran más posibilidades de combinaciones de servicios, lo que implica menores costos sociales, pero a costa de incorporar también más restricciones. En los capítulos siguientes se muestran aplicaciones de la variante de nivel intermedio de complejidad (la que trabaja sobre un corredor en el que paraderos de ida coinciden con los de vuelta, utilizando las restricciones de continuidad de frecuencias representadas por la ecuación (2.42)). En el Capítulo 3 se saca provecho de la separación por sentidos de los servicios para concentrar los esfuerzos de diseño de servicios en un sentido en particular, lo que tiene especial relevancia en presencia de capacidad limitada.

Sería interesante extender este enfoque de codificación de servicios por sentido a un modelo que combine porciones de servicios para construir servicios para el corredor. La idea general consistiría en determinar en cada nodo qué parte de la frecuencia de cada porción de servicio entrante se le transfiere a cada porción de servicio saliente, y unir estas partes en un servicio más largo. La dificultad para resolver este problema radicaría en encontrar la forma de informar al modelo que dos servicios están posiblemente conectados y que los usuarios que viajan entre ellos no requieren realizar un trasbordo ni esperar nuevamente por el servicio.

- **Incorporación de trasbordos en sentido contrario**

La existencia de servicios expresos puede generar situaciones donde un usuario puede encontrar conveniente realizar algún tramo de su viaje en la dirección contraria del sentido de su desplazamiento. Esto sucede, por ejemplo, cuando devolviéndose algunos paraderos puede acceder a un servicio expreso que lo lleva directo a su destino y que no se puede acceder desde otros paraderos aguas abajo, o cuando un servicio expreso que atiende el paradero de origen de un viaje posee como destino algún nodo más allá del destino deseado.

Cuando los servicios son codificados como ciclos, sin embargo, este tipo de trasbordos son impedidos por el modelo a menos que se tomen precauciones adicionales, ya que al utilizar nodos para ambos sentidos por separado en la codificación, el flujo no tiene forma de moverse entre los paraderos de sentidos contrarios del corredor.

Un posible enfoque para abordar este problema consistiría en agregar un servicio ficticio que entregue conexión entre los nodos correspondientes a un mismo paradero. Este servicio debe poseer costo nulo para el operador, y tiempos nulos de viaje entre paraderos donde es factible trasbordar. Esto último se debe a que, a pesar de que estos tiempos podrían ser aprovechados para incorporar de mejor forma los diferentes tiempos de caminata de los distintos lugares de trasbordo, cuando la demanda se encuentra codificada a nivel de paradero se podrían generar costos indeseados al final del viaje, donde el pasajero se ve obligado a cambiar de sentido de paradero para llegar a su destino. La frecuencia de este servicio ficticio debe ser fijada en un valor alto al momento de optimizar, dado que la caminata entre

paraderos no conlleva una espera. Es importante notar que los flujos sobre este servicio ficticio no deben ser considerados en la expresión (2.10) para los costos de trasbordo, de forma de no generar dobles conteos.

Otra forma de enfrentar este problema, válida sólo para el caso donde los paraderos de ida coinciden con los de vuelta, corresponde a modelar los paraderos de ambos sentidos como un único nodo, de forma de que queden efectivamente conectados ambos sentidos del corredor (o, equivalentemente, colapsar las restricciones de continuidad expresadas en (2.15) para los nodos que corresponden a dos sentidos de un mismo paradero en una sola). Esta forma de modelar los corredores es particularmente conveniente para el caso de redes (abordado en el Capítulo 5), donde no tiene sentido asignar los orígenes o destinos de los viajes a un sentido en particular de los corredores ya que los viajes podrían ser realizados por diferentes partes de la red.

- **Limitación de puntos de trasbordo**

Otra forma de reducir la complejidad del problema de optimización de frecuencias consiste en restringir los puntos de trasbordo a un subconjunto de los nodos de la red. Llamaremos a este conjunto \mathcal{N}_{tr} . Estos nodos pueden ser escogidos de acuerdo al nivel de actividad de los paraderos, ya que si el diseño del sistema va a incentivar los trasbordos o incluso los va a hacer necesarios para algunos viajes, lo natural es que estos puntos coincidan con las paradas de mayor movimiento (entendiendo movimiento en este caso como viajes originados más viajes atraídos). De todas formas, pueden existir en la práctica otros criterios técnicos para seleccionar estos

paraderos, como por ejemplo la capacidad de los paraderos en sí: se desearía desincentivar los trasbordos en paraderos que no tienen la capacidad de atender a un número muy alto de buses o de pasajeros.

Ya definido el conjunto \mathcal{N}_{tr} , el próximo paso es identificar cuál es el nuevo conjunto de secciones de ruta s que se pueden utilizar para realizar un viaje en el par w . Llamaremos conjunto de tramos válidos, \mathcal{V}_w , al conjunto de secciones s que cumplan la siguiente condición:

$$s \in \mathcal{V}_w \Leftrightarrow [O_s \in \mathcal{N}_{tr} \cup \{O_w\}] \wedge [D_s \in \mathcal{N}_{tr} \cup \{D_w\}] \quad (2.49)$$

El conjunto \mathcal{V}_w contiene las secciones s cuyo inicio coincide con el del par w o con un nodo de trasbordo, y cuyo final coincide con el nodo de destino de w o con un nodo de trasbordo. En la Figura 2-4 se entrega un ejemplo de la construcción del conjunto $\mathcal{V}_{(i,j)}$ para un corredor con dos nodos de trasbordo.

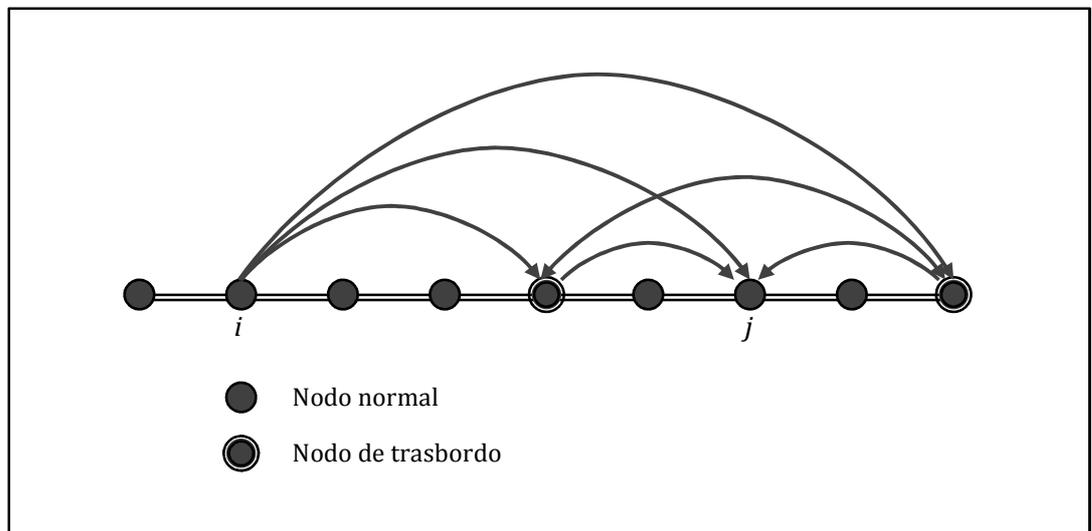


Figura 2-4: Construcción del conjunto $\mathcal{V}_{(i,j)}$

El nuevo modelo con trasbordos restringidos consistirá entonces en un modelo similar al modelo base, donde las variables V_s^w estén definidas sólo para los pares $s \in \mathcal{V}_w$. Alternativamente, es posible mantener las variables definidas originalmente para todas las combinaciones de (s, w) y simplemente imponer que todos los flujos fuera del conjunto de tramos válidos sean nulos, es decir:

$$V_s^w = 0, \quad \forall s \notin \mathcal{V}_w \quad (2.50)$$

De implementar esta simplificación, es necesario poner atención a qué sucede con los posibles trasbordos en sentido contrario al del par w , de acuerdo a lo discutido en el punto anterior.

- **Enfoque binivel**

Una de las dificultades que surgen al trabajar con el modelo de frecuencias es que debido a la gran cantidad de variables que involucra y la multiplicidad de óptimos locales que presenta, no siempre se puede confiar en la calidad de las soluciones que un solver comercial es capaz de encontrar para el problema. En particular, se ha podido apreciar en la práctica que cuando crece la complejidad del problema, la herramienta probada (en este caso el solver Minos) entrega soluciones donde la asignación de los usuarios es evidentemente subóptima, ignorando trasbordos que claramente podrían disminuir los tiempos de viaje, o escogiendo mal el conjunto de líneas atractivas para alguna sección de ruta.

Cuando esto sucede, se hace conveniente separar las etapas de optimización de frecuencias de la asignación de los usuarios y resolver el problema como un problema en dos niveles. Una de las ventajas de este enfoque es que el problema de

asignación de los usuarios a veces puede ser resuelto (o al menos simplificado) en forma directa mediante algoritmos, sin necesidad de resolver un problema de optimización.

Para resolver el problema mediante este enfoque, en un nivel superior se encontraría el problema de optimización de frecuencias para una asignación dada. Esta asignación será representada por la matriz \hat{T}_w , que se obtiene dividiendo en etapas cada viaje realizado sobre la red. Debido a que al fijar la asignación se eliminan del problema los trasbordos, resulta conveniente en este caso trabajar con la función objetivo presentada en (2.21). Esta función objetivo utiliza los parámetros δ_l^w (fijos en esta etapa) en lugar de las frecuencias efectivas. De esta forma se previenen los conflictos que podrían surgir al fijar las frecuencias efectivas, ya que éstas impondrían una cota en las frecuencias a optimizar del problema. Con esto, el problema del nivel superior a resolver es el siguiente:

$$\min_{f_l} \sum_{l \in \mathcal{L}} f_l c_l + \theta_{tt} \sum_{w \in \mathcal{W}} \hat{T}_w \left(\frac{\sum_{l \in \mathcal{L}} \delta_l^w f_l t_l^w}{\sum_{l \in \mathcal{L}} \delta_l^w f_l} \right) + \theta_{wt} \sum_{w \in \mathcal{W}} \hat{T}_w \left(\frac{\lambda}{\sum_{l \in \mathcal{L}} \delta_l^w f_l} \right) \quad (2.51)$$

El problema queda sujeto exclusivamente a restricciones de signo de las variables:

$$f_l \geq 0, \quad \forall l \in \mathcal{L}, \quad (2.52)$$

El nivel inferior del problema corresponde a la asignación de los pasajeros a la red. Este problema consiste en determinar, dadas las frecuencias de los servicios, la asignación de los pasajeros en la red. Esta asignación equivale a determinar dos cosas: en primer lugar, los conjuntos de líneas atractivas para cada sección de ruta s (es decir, determinar δ_l^w), y, conocidos estos valores, la asignación de los pasajeros a la red considerando trasbordos (es decir, determinar V_s^w).

La primera parte del problema antes descrito se puede resolver por separado para cada par s en forma casi trivial aplicando el algoritmo propuesto por en Chriqui y Robillard (1975). A grandes rasgos, el algoritmo consiste en ordenar los servicios disponibles para la sección s por orden creciente en tiempo de viaje, y agregarlos secuencialmente al conjunto de líneas atractivas, calculando el tiempo esperado de viaje resultante para cada caso. Este procedimiento alcanza en una iteración un valor mínimo para el tiempo esperado de viaje, que comienza a aumentar si se siguen agregando servicios. El conjunto de líneas con el que se alcanza dicho mínimo corresponde al conjunto de líneas atractivas. Con esta solución se actualizan las variables δ_l^w .

La segunda parte del problema de asignación se puede resolver mediante el siguiente problema de optimización:

$$\begin{aligned} \min_{V_s^w} & \left[\sum_{w \in \mathcal{W}} \sum_{s \in \mathcal{S}} V_s^w \left\{ \theta_{wt} \frac{\lambda}{\sum_{l \in \mathcal{L}} \delta_l^s f_l} + \theta_{tt} \frac{\sum_{l \in \mathcal{L}} t_l^s \delta_l^s f_l}{\sum_{l \in \mathcal{L}} \delta_l^s f_l} \right\} \right. \\ & \left. + \theta_{tr} \left\{ \sum_{w \in \mathcal{W}} \sum_{s \in \mathcal{S}} V_s^w - \sum_{w \in \mathcal{W}} T_w \right\} \right] \end{aligned} \quad (2.53)$$

Sujeto a las restricciones de continuidad de flujo, que se muestran en (2.15).

El problema recién planteado, si lo separamos por par w , es similar a un problema de rutas mínimas donde se incurre en un costo fijo de θ_{tr} por cada vez que se pasa por un nodo (excepto el inicial). Esto significa que si se le agrega al costo de cada arco el costo de trasbordo, el problema de asignación en cada par equivale un problema de rutas mínimas que se puede resolver usando algoritmos como el de Dijkstra. Las rutas resultantes de esta conversión están todas sobredimensionadas en un trasbordo

(el del nodo inicial), pero dado que esta magnitud es la misma en todas las alternativas, este enfoque lleva a la solución óptima del problema (2.53).

Para volver al nivel superior del problema es necesario recalcular la matriz \hat{T}_w de etapas de viajes consecuencia de la nueva asignación, como se muestra en la fórmula a continuación:

$$\hat{T}_w = \sum_{k \in \mathcal{W}} V_w^k \quad (2.54)$$

El enfoque binivel acá descrito no asegura optimalidad, aunque sí convergencia. La secuencia de asignaciones y optimizaciones de frecuencias pueden llevar a un óptimo local del problema, que en teoría la formulación simultánea del problema podría ser capaz de evitar si el método de solución del problema simultáneo fuera lo suficientemente potente. Con respecto a la convergencia, basta con observar que cada iteración de cada subproblema siempre reduce o mantiene los costos sociales del corredor. Por un lado, cuando la asignación es fija, redefinir las frecuencias sólo puede mejorar la solución actual, no empeorarla. Por otra parte, reasignar a los usuarios hace que cada uno de ellos minimice individualmente sus costos, sin afectar los costos del operador, lo que necesariamente lleva a una solución mejor (o igual) que la anterior.

- **Otros comportamientos de usuario**

El comportamiento de los usuarios puede ser modelado mediante distintos enfoques. Anteriormente se discutió la existencia de modelos más complejos para modelar este comportamiento, basados en la minimización de estrategias de viaje en lugar de tan sólo rutas mínimas o itinerarios. En palabras sencillas, minimizar estrategias equivale

a permitir que los usuarios decidan dónde hacer sus trasbordos en función del servicio que pase primero. Esto permite, en el caso de una red, que los mismos usuarios carguen distintas partes de la red sin concentrarse todos ellos en sólo un sector.

Cualquier cambio en el modelo de comportamiento de usuarios acepta dos posibles enfoques de modelación. Por un lado, se puede intentar formular un modelo simultáneo de optimización de frecuencias y de asignación de usuarios de forma similar a lo que hace el modelo de Leiva. Otro posible enfoque consistiría en plantear el problema en dos niveles como se explica en el apartado anterior, lo que permite incorporar un mayor nivel de complejidad al comportamiento de usuarios, pero que requiere de algún método iterativo de solución como el propuesto anteriormente. A continuación se aplica el primer enfoque para construir un modelo para la optimización de frecuencias que considera estrategias mínimas en forma simultánea. Para incorporar un comportamiento de usuarios a estrategias mínimas, se plantea un modelo que se puede entender como una combinación del modelo de Spiess y Florian (1983) con el de Leiva. El modelo de Spiess y Florian para obtener las estrategias mínimas de los viajes que comparten un destino d se puede plantear como sigue:

$$\min_{v_s, V_i, x_s} \theta_{tt} \sum_{s \in \mathcal{S}l} v_s t_s + \theta_{wt} \sum_{i \in \mathcal{N}} \frac{\lambda V_i}{\sum_{s \in \mathcal{S}l_i^+} f_s x_s} \quad (2.55)$$

En la fórmula anterior se introduce $\mathcal{S}l$ para representar al conjunto de secciones de línea del corredor. Cada elemento del conjunto $\mathcal{S}l$ está asociado a una única línea. El conjunto $\mathcal{S}l_i^+$ contiene a las secciones de línea que salen desde un nodo i . Las

variables v_s corresponden al flujo en la sección s , mientras que t_s y f_s representan el tiempo de viaje y la frecuencia (fijos y conocidos) de dicha sección. Las variables V_i corresponden al flujo total que espera en el nodo i . La función objetivo planteada en (2.55) corresponde entonces a la suma de los costos por tiempos de viaje más tiempo de espera para los usuarios. El modelo está sujeto a las siguientes restricciones:

$$\sum_{s \in \mathcal{S}l_i^+} v_s = \sum_{s \in \mathcal{S}l_i^-} v_s + T_{(i,d)}, \quad \forall i \in \mathcal{N} - \{d\} \quad (2.56)$$

$$v_s = \frac{x_s f_s V_i}{\sum_{r \in \mathcal{S}l_i^+} f_r x_r}, \quad \forall s \in \mathcal{S}l_i^+, \forall i \in \mathcal{N} \quad (2.57)$$

$$v_s \geq 0, \quad \forall s \in \mathcal{S}l \quad (2.58)$$

$$x_s = \{0,1\}, \quad \forall s \in \mathcal{S}l \quad (2.59)$$

En las restricciones anteriores se introducen los conjuntos $\mathcal{S}l_i^+$ y $\mathcal{S}l_i^-$, que contienen a las secciones de línea salientes y entrantes desde un nodo i . La restricción (2.56) representa una ecuación de continuidad de flujos. La variable binaria x_s en la restricción (2.57) representa si la sección de línea s es atractiva o no, de forma que esta restricción reparte el flujo que espera un bus en i , V_i , entre las variables de flujo v_s , con $s \in \mathcal{S}l_i^+$. Las restricciones (2.58) y (2.59) imponen las condiciones de signo y binariedad de las variables del modelo.

El modelo anterior puede ser simplificado aplicando el siguiente cambio de variables:

$$w_i = \frac{V_i}{\sum_{s \in \mathcal{S}l_i^+} f_s x_s} \quad (2.60)$$

Con esto, la función objetivo se puede reescribir como:

$$\min_{v_s, w_i, x_s} \theta_{tt} \sum_{s \in \mathcal{S}l} v_s t_s + \theta_{wt} \lambda \sum_{i \in \mathcal{N}} w_i \quad (2.61)$$

Y la restricción (2.57) debe ser reemplazada por:

$$v_s = x_s f_s w_i, \quad \forall s \in \mathcal{S}l_i^+, \forall i \in \mathcal{N} \quad (2.62)$$

Spiess y Florian demuestran que la variable x_s se puede hacer continua obteniendo un problema equivalente de programación continua. Con esto, además de descartar la restricción de binariedad (2.59), se puede reemplazar (2.62) por:

$$v_s \leq f_s w_i, \quad \forall s \in \mathcal{S}l_i^+, \forall i \in \mathcal{N} \quad (2.63)$$

Si se asume que no hay congestión, el problema puede ser agregado para todos los destinos d . Dado que los flujos y esperas dependen del destino, se les debe agregar un superíndice. El problema agregado se puede plantear como:

$$\min_{v_s^d, w_i^d} \theta_{tt} \sum_{d \in \mathcal{N}} \sum_{s \in \mathcal{S}l} v_s^d t_s + \theta_{wt} \lambda \sum_{d \in \mathcal{N}} \sum_{i \in \mathcal{N}} w_i^d \quad (2.64)$$

Sujeto a:

$$\sum_{s \in \mathcal{S}l_i^+} v_s^d = \sum_{s \in \mathcal{S}l_i^-} v_s^d + T_{(i,d)}, \quad \forall i \in \mathcal{N}, \forall d \in \mathcal{N} \quad (2.65)$$

$$v_s^d \leq f_s w_i^d, \quad \forall s \in \mathcal{S}l_i^+, \forall i \in \mathcal{N}, \forall d \in \mathcal{N} \quad (2.66)$$

$$v_s^d \geq 0, \quad \forall s \in \mathcal{S}l, \forall d \in \mathcal{N} \quad (2.67)$$

Al modelo anterior se le pueden agregar los costos del operador, completando los costos sociales, y así obteniendo un problema de optimización de frecuencias:

$$\min_{v_s^d, w_i^d, f_l} \sum_{l \in \mathcal{L}} f_l c_l + \theta_{tt} \sum_{d \in \mathcal{N}} \sum_{s \in \mathcal{S}l} v_s^d t_s + \theta_{wt} \lambda \sum_{d \in \mathcal{N}} \sum_{i \in \mathcal{N}} w_i^d \quad (2.68)$$

La restricción (2.66) debe ser adaptada para utilizar f_l en lugar de f_s . Esto se logra incorporando un parámetro binario δ_s^l , que indica si la línea l corresponde a la sección de línea s :

$$v_s^d \leq \delta_s^l f_l w_i^d, \quad \forall s \in \mathcal{S}l_i^+, \forall i \in \mathcal{N}, \forall d \in \mathcal{N} \quad (2.69)$$

El modelo acá propuesto optimiza las frecuencias en un corredor para un conjunto de servicios expresos dados, donde los usuarios se comportan de acuerdo a estrategias mínimas. Este modelo puede ser utilizado en forma alternativa al modelo de Leiva si se desea incorporar este tipo de comportamiento. Con respecto a la adición de una restricción en la capacidad de los vehículos, las consideraciones a tomar en cuenta son las mismas que en otro caso, y se pueden aplicar la misma heurística HC_1 descrita anteriormente para realizar el ajuste de capacidad o alguna de las que se proponen en el siguiente apartado.

Como se mencionaba antes, si se resuelve el problema de optimización de frecuencias en dos niveles, separando la asignación de los usuarios de la optimización de frecuencias propiamente tal, es posible resolver el problema utilizando modelos de comportamiento más complejos que el de minimización de rutas. Sin embargo, para que esto funcione es necesario que el comportamiento de los usuarios corresponda una minimización de costos que pueda ser incorporada en los costos sociales, con tal de poder asegurar que cada iteración (optimización de frecuencias para una asignación fija y asignación de los usuarios) contribuya a disminuir los costos sociales y así asegurar que el algoritmo converja.

Otra posible variante de comportamiento de usuarios que el enfoque binivel permitiría incorporar consistiría en el caso donde existen usuarios con diferentes

percepciones para el valor del tiempo. En este caso, los costos de los usuarios se sumarían para clases de usuarios con diferentes valores del tiempo, lo que no es complicado de incorporar en la función objetivo. Una ventaja que puede tener la incorporación de clases de usuarios en la modelación consistiría en que las discontinuidades en la asignación de flujo en la red ocasionadas por umbrales de flujo que provocan cambios en el comportamiento de usuarios se verían reducidas en magnitud, dándole mayor robustez al modelo de comportamiento. De todas formas, esta ventaja viene acompañada de un aumento en el número de puntos de discontinuidad de asignación en el dominio, y de la obligación de utilizar como parámetros de optimización los mismos valores del tiempo de los usuarios y sus distintas clases. Vale la pena remarcar, en todo caso, que las discontinuidades mencionadas corresponden exclusivamente a la asignación de usuarios a buses, y no a los costos de la función objetivo. De hecho, los puntos donde existen cambios de comportamiento son por definición puntos de indiferencia donde los costos experimentados por los usuarios son los mismos en una vecindad del dominio.

Un último punto que vale la pena hacer notar es que todos los modelos de comportamiento acá mencionados funcionan correctamente siempre y cuando la capacidad no sea limitante, por los motivos que ya han sido explicados anteriormente. Podría ser interesante intentar modelar la asignación de los usuarios frente a capacidad limitada, incorporando efectos tales como la posibilidad de que un pasajero pueda dejar pasar un bus lleno para abordar el próximo que pase con capacidad disponible. Este tipo de modelación, en todo caso, sería altamente compleja de implementar y de resolver, incluso adoptando un enfoque binivel.

- **Mejoras a la heurística de ajuste de capacidad**

Las simplificaciones y variantes del modelo acá sugeridas sirven para facilitar la solución del problema, sobre todo en ausencia de congestión. Sin embargo, ninguna de estas variantes del modelo aborda cómo lidiar con la presencia de congestión. En Leiva et al. (2010) se propone una heurística para la solución del problema con capacidad, que fue explicada en la primera parte de este capítulo. Ahí también se mencionan las que a juicio del autor son las dos mayores debilidades de dicha heurística: su gran número de iteraciones, y la suboptimalidad de las soluciones que entrega. El primero de estos problemas puede ser resuelto con una pequeña modificación al algoritmo. Debido a que por cada paso que el algoritmo da (es decir, cada vez que se aumenta en un delta una cota mínima de la frecuencia) se debe resolver una vez el algoritmo de optimización, se hace necesario evitar iteraciones innecesarias. Una forma simple de hacer esto podría ser aumentar el tamaño del incremento adoptado para las cotas inferiores, pero una medida de este tipo evidentemente repercute en la precisión de la solución obtenida. Una mejor forma de disminuir el número de iteraciones se propone a continuación:

Heurística para el ajuste de capacidad HC₂:

1. Inicializar las cotas en $\check{f}_i = 0$.
2. Resolver el problema (2.12) - (2.15) más la restricción (2.19). Si la solución encontrada no presenta arcos con insuficiencia de capacidad, detenerse.

3. Para la solución actual, escoger una línea \hat{l} donde exista un déficit de capacidad, es decir donde la inecuación (2.18) no es satisfecha. El criterio de selección será escoger la línea \hat{l} que presente el arco donde este déficit sea mayor, de acuerdo a la fórmula (2.20).
4. Aumentar la cota mínima para la frecuencia de la línea \hat{l} hasta el valor donde la frecuencia sea la suficiente para transportar la carga crítica del servicio, es decir, asignar a la cota inferior el valor $\check{f}_{\hat{l}} = P_{\hat{l}}^{max} / cap_{\hat{l}}$, donde $P_{\hat{l}}^{max}$ corresponde a la carga máxima del servicio \hat{l} en la solución actual.
5. Volver al paso 2.

Como se puede apreciar, la diferencia entre este algoritmo y HC_1 es que se cambia el criterio con el que se define la magnitud del aumento en la cota inferior de la frecuencia en déficit de capacidad seleccionada. Este algoritmo disminuye drásticamente el número de iteraciones, pero es posible que en algunos casos aumente la capacidad de algún servicio más allá de lo estrictamente necesario. El algoritmo HC_1 al seleccionar un servicio crítico y aumentar su frecuencia gradualmente puede llegar a un momento en el que el arco crítico del sistema cambie de lugar, posiblemente en otro servicio. En este momento el algoritmo dirige su atención a este nuevo punto de conflicto, asegurando que en todo momento se está entregando capacidad en el lugar donde es más necesario. El algoritmo HC_2 acá planteado no detecta esta situación, sino que realiza el aumento de frecuencia en un paso fijo que puede exceder el punto recién descrito.

Sin embargo, se debe tomar en cuenta que ambos procedimientos se basan en criterios heurísticos que no aseguran optimalidad lo que hace difícil predecir cuál de los dos llegará a una mejor solución en un caso particular, y que la reducción en el tiempo de ejecución del algoritmo gracias a la simplificación de HC_2 suele ser drástica. En el Capítulo 3 se muestra una comparación de la calidad de ambos métodos en términos de la solución que encuentran y del número de iteraciones realizadas para alcanzarla, mostrando que efectivamente la reducción en el tiempo de ejecución es considerable, mientras que el efecto en la calidad de la solución encontrada es poco predecible.

Vale la pena notar que el paso 3 propuesto para esta variante de la heurística no asegura que el problema de capacidad de la línea se vea resuelto en esa misma iteración, incluso aunque el arco crítico no cambiase. Esto sucede debido a que al aumentar la frecuencia de un servicio también cambian las preferencias de los usuarios: la existencia de un servicio más frecuente se puede traducir en que los usuarios descarten algunos de los servicios más lentos para esperar el servicio más frecuente si es que esto resulta en una disminución de su tiempo esperado de viaje. Cuando esto sucede, el aumento de frecuencia no será suficiente para cubrir la demanda del servicio, y serán necesarias nuevas iteraciones para satisfacer la demanda.

Una debilidad que presenta tanto HC_1 como HC_2 y que pudo ser apreciada durante pruebas preliminares consiste cuando el punto de mayor déficit ocurre en el servicio regular, ambas heurísticas elevan la frecuencia de este servicio ignorando la posibilidad de aplacar este déficit entregando frecuencia a algún servicio expreso que

acarree parte del flujo de dicho arco crítico, a menor costo para el operador, y entregando mejores tiempos de viaje a los usuarios.

Una forma de prevenir este problema consiste en priorizar en la heurística de selección a los servicios expresos sobre los servicios regulares. Esta variante del criterio de selección será aplicada sobre la versión HC_2 de la heurística, ya que la versión HC_1 realiza demasiadas iteraciones y, como se mostrará en el próximo capítulo, no entrega resultados significativamente mejores que HC_2 . Esta tercera versión de la heurística se presenta a continuación.

Heurística para el ajuste de capacidad HC_3 :

1. Inicializar las cotas en $\check{f}_l = 0$.
2. Resolver el problema (2.12) - (2.15) más la restricción (2.19). Si la solución encontrada no presenta arcos con insuficiencia de capacidad, detenerse.
3. Para la solución actual, escoger una línea \hat{l} donde exista un déficit de capacidad, es decir donde la inecuación (2.18) no es satisfecha, utilizando la fórmula (2.70). Notar que la diferencia entre esta fórmula y (2.20) es que en este caso se excluye el servicio regular (l_a) para la elección de c . Sólo si no hay servicios en déficit salvo el regular, escoger este servicio como \hat{l} .

$$\hat{l} = \operatorname{argmax}_{l \in \mathcal{L} - \{l_a\}} \left[\max_{a \in \mathcal{A}} \left(\sum_{w \in \mathcal{W}} \sum_{s \in \mathcal{S}} V_s^w \zeta_a^s \frac{f_l^s}{\sum_{i \in \mathcal{L}} f_i^s} - cap_l f_l \right) \right] \quad (2.70)$$

4. Aumentar la cota mínima para la frecuencia de la línea \hat{l} hasta el valor donde la frecuencia sea la suficiente para transportar la carga crítica del servicio, es

decir, asignar a la cota inferior el valor $\check{f}_i = P_i^{max} / cap_i$, donde P_i^{max} corresponde a la carga máxima del servicio \hat{l} .

5. Volver al paso 2.

Las variantes propuestas para el algoritmo de búsqueda de solución para el problema con capacidad ofrecen una reducción importante en el tiempo de ejecución del algoritmo. Sin embargo, el problema de la suboptimalidad sigue presente. Una forma de resolver el problema con capacidad a optimalidad consiste en reducir el conjunto de servicios a considerar, de forma de que el problema resultante sea analíticamente tratable. Este tipo de algoritmos son estudiados en el próximo capítulo como heurísticas simultáneas de ajuste de capacidad y de diseño de servicios.

- **Otras restricciones de capacidad**

Como se menciona al presentar el método para el ajuste de capacidad de los vehículos para el modelo de Leiva, existen otras limitaciones físicas en el problema que podría ser necesario considerar en algunas ocasiones. En particular, estas otras formas de capacidad del sistema son:

- Para cada paradero existe un número máximo de buses que pueden ser atendidos simultáneamente.
- Los paraderos deben tener capacidad suficiente para albergar a los pasajeros que lo utilizan.
- Es necesario que el tiempo que los buses permanecen en los paraderos sea suficiente para que los pasajeros alcancen a abordar y descender de éstos.

- La capacidad de las vías debe ser suficiente que los buses puedan circular. Asumiendo que los buses circulan por vías exclusivas y que no compiten por capacidad con otros vehículos, el cuello de botella en términos de capacidad vial, cuando no se encuentre en los paraderos, se encontrará en las intersecciones semaforizadas.

Estas restricciones, hasta donde hemos podido verificar, no han sido abordadas en los trabajos existentes relativos al diseño de servicios expresos. Sin embargo, en los corredores de demanda muy alta, estas restricciones pueden ser muy importantes, e incluso centrales para el problema. Esto es particularmente crítico en casos donde la demanda es tal que si no se utilizan servicios expresos, la frecuencia mínima del servicio regular es tan alta que algunos paraderos del corredor simplemente no darían abasto (o que obligaría a sobredimensionar estaciones de poca confluencia). En este caso, los servicios expresos no sólo pueden ser una alternativa más económica para atender el corredor, sino que se transforman en la única alternativa viable, y el problema adquiere una nueva dimensión de complejidad como problema de factibilidad, dado que ahora no siempre es trivial encontrar una solución factible.

A continuación se discute como se podría abordar cada uno de los casos propuestos.

a) Capacidad de atención de buses en un paradero

Una forma de modelar la primera restricción de capacidad antes mencionada sería incorporar en el modelo base una restricción como la siguiente:

$$\sum_{l \in \mathcal{L}} \pi_l^i f_l \leq f_i^{max}, \quad \forall i \in \mathcal{N} \quad (2.71)$$

En esta expresión π_l^i es el parámetro binario que indica si la línea l atiende al nodo i . El parámetro f_i^{max} corresponde a la máxima frecuencia de buses que el paradero es capaz de atender. Esta frecuencia máxima se puede estimar como el inverso del tiempo promedio de detención de los vehículos en el paradero (medido entre la entrada de dos buses seguidos, para considerar tiempos de maniobra) multiplicado por el número de espacios para buses existentes en el paradero. Sería conveniente además amplificar este valor para dar un margen de seguridad, debido al conocido resultado de teoría de colas que indica que si la tasa de llegada se acerca mucho a la tasa de atención se comienzan a formar colas en el sistema, producto de la aleatoriedad existente en los tiempos de servicio.

La solución de un problema con este tipo de restricciones de capacidad puede ser muy complicada. Esto sucede porque, a diferencia de las restricciones de capacidad de los vehículos, la capacidad de los paraderos impone una cota máxima para las frecuencias en lugar de mínima, lo que implica que no se puede asegurar la existencia de solución. Además, estas restricciones de capacidad, al igual que las de los vehículos, entran en conflicto con el comportamiento de los usuarios si se incluyen como restricción en un problema de optimización de costo social, ya que el modelo puede obligar a algunos usuarios a desviarse de su ruta ideal de forma de no sobrepasar la capacidad de algún paradero en particular. En un principio, se podría estudiar esta variante del problema relajando la condición de comportamiento de usuarios de forma de no tener que realizar un ajuste de capacidad a la solución que se obtenga, ya que las heurísticas de capacidad antes presentadas trabajan con cotas mínimas que entrarían en conflicto con las restricciones de capacidad en paradero.

b) Capacidad de pasajeros en paraderos

En principio, el problema de capacidad de albergue de pasajeros en paraderos podría ser modelado como una restricción sobre las frecuencias. Esta restricción no es sencilla de modelar, ya que implica estudiar la formación de colas de pasajeros en los paraderos. Esto obliga a estudiar el funcionamiento del paradero como un proceso estocástico. Si se considera que distintos usuarios desean abordar distintos buses dentro de un conjunto de líneas atractivas, la complejidad de este tipo de restricciones crece considerablemente. Por esto, el enfoque que parece más sensato en este caso es simplemente diseñar los paraderos ex post, en función de los resultados del modelo, ya que no vale la pena limitar el dominio ni complejizar la formulación del problema por incorporar este efecto.

c) Capacidad de abordaje y descenso de pasajeros en buses

Una opción que parece natural en este caso sería modelar el tiempo de detención de los vehículos como un tiempo variable, proporcional al flujo que aborda y desciende del bus en cada paradero. Sin embargo, como se muestra en Larrain y Muñoz (2008), este enfoque puede presentar casos donde, para que el equilibrio exista, se debe permitir que en una misma sección coexistan distintos tipos de comportamiento (es decir, que existan usuarios, iguales entre sí, con distintos conjuntos de líneas atractivas), lo que complica enormemente la búsqueda del equilibrio de usuarios de este tipo de sistema.

Una simplificación que permite modelar esta restricción consiste en definir el tiempo de detención de cada línea en cada paradero como un parámetro y chequear a mano

que la solución obtenida no viole la capacidad, aumentando iterativamente los tiempos de detención en los paraderos y líneas donde resulte necesario. Definiendo τ_i^l como el tiempo de detención en el paradero i del servicio l , y a_1 y a_2 como los tiempos de abordaje y descenso por pasajero, la condición de capacidad sería la siguiente:

$$a_1 \sum_{j \in \mathcal{N}} \frac{f_l^{(i,j)} \sum_{w \in \mathcal{W}} V_{(i,j)}^w}{\sum_{l \in \mathcal{L}} f_l^{(i,j)}} + a_2 \sum_{j \in \mathcal{N}} \frac{f_l^{(j,i)} \sum_{w \in \mathcal{W}} V_{(j,i)}^w}{\sum_{l \in \mathcal{L}} f_l^{(j,i)}} \leq \beta \tau_i^l, \quad \forall i \in \mathcal{N} \quad (2.72)$$

En la expresión anterior β corresponde a un factor de seguridad para la capacidad. Naturalmente, la formulación de los tiempos de viaje y costos de operación debe también ser redefinida para incluir adecuadamente estos nuevos tiempos de parada.

Esta forma iterativa de abordar la capacidad de carga y descarga de los buses tiene la ventaja de que no entra en conflicto con la asignación de los usuarios, ya que en ningún momento restringe activamente el acceso a las diversas opciones.

Cabe notar que en la condición (2.72) el tiempo de carga y descarga de los buses es modelado como dependiente de la suma de los pasajeros que abordan y descienden del bus. Sin embargo, existen casos donde este tiempo es mejor descrito como el valor máximo entre estas dos variables, como sucede en la mayoría de los sistemas BRT donde el descenso y acenso se simultáneamente hace por diferentes puertas. Para modelar el problema de esta manera se debe restringir por separado la capacidad de subida y de bajada como se muestra en las desigualdades (2.73) y (2.74).

$$a_1 \sum_{j \in \mathcal{N}} \frac{f_l^{(i,j)} \sum_{w \in \mathcal{W}} V_{(i,j)}^w}{\sum_{l \in \mathcal{L}} f_l^{(i,j)}} \leq \beta \tau_i^l, \quad \forall i \in \mathcal{N} \quad (2.73)$$

$$a_2 \sum_{j \in \mathcal{N}} \frac{f_l^{(j,i)} \sum_{w \in \mathcal{W}} V_{(j,i)}^w}{\sum_{l \in \mathcal{L}} f_l^{(j,i)}} \leq \beta \tau_i^l, \quad \forall i \in \mathcal{N} \quad (2.74)$$

d) Capacidad de las intersecciones semaforizadas

Este tipo de restricciones tiene una naturaleza similar a las de capacidad de atención de buses en los paraderos. Así como los servicios expresos pueden evitar ciertos paraderos que se encuentran operando a capacidad, los servicios expresos podrían también evitar pasar por ciertas intersecciones más complicadas, suponiendo que la omisión de algunos paraderos abre la posibilidad de utilizar vías alternativas.

La restricción de capacidad asociada a semáforos podría ser incorporada al modelo con restricciones del tipo:

$$\sum_{l \in \mathcal{L}} \sigma_i^{tl} f_l \leq f_{tl}^{max}, \quad \forall tl \in \mathcal{T}l \quad (2.75)$$

En la última ecuación $\mathcal{T}l$ representa el conjunto de intersecciones semaforizadas del corredor. El parámetro binario σ_i^{tl} indica si la línea l pasa por la intersección tl , y f_{tl}^{max} corresponde al flujo máximo de buses que puede ser atendido por el semáforo sin que se formen colas.

La naturaleza de estas últimas restricciones es muy similar a la de las de capacidad de buses en paraderos discutida antes. Esta restricción también implica cotas superiores para las frecuencias, lo que acarrea los mismos problemas que en primer caso, haciendo compleja su incorporación al modelo sobre todo si la capacidad de los buses también se desea tomar en consideración.

3. GENERACIÓN DE SERVICIOS EXPRESOS PARA UN CORREDOR

Como se explicó en el Capítulo 2, una limitación importante del modelo de Leiva es que necesita que el modelador proporcione un conjunto de servicios candidatos para ser optimizado, pero no entrega ningún tipo de procedimiento o criterio para generar dichos servicios.

En esta sección se presentan diferentes algoritmos para la generación de servicios para alimentar al modelo. Estos algoritmos se pueden clasificar dentro de dos grandes categorías. Los primeros establecen su foco en la generación de servicios nuevos sin tomar en consideración la restricción de capacidad del problema, y aplican el modelo y las heurísticas de capacidad presentadas en el capítulo anterior para determinar si estos servicios son o no beneficiosos. La segunda categoría corresponde a heurísticas que generan uno o dos servicios de un formato preestablecido con el fin de mejorar la operación para un escenario con congestión. Las heurísticas de esta segunda categoría poseen elementos tanto de generación de servicios como de ajuste de capacidad.

Los métodos acá presentados se obtienen en general mediante la siguiente estrategia: se propone una forma de generar un nuevo servicio (que puede ser, por ejemplo, agregar paraderos a un expreso existente, o trabajar sobre un formato previamente impuesto, como por ejemplo un servicio corto) y se evalúa el efecto que posee su implementación (esto puede ser en términos absolutos, utilizando la función de costos sociales, o en términos relativos respecto a la solución con un único servicio regular) para escoger la mejor opción disponible.

El primer método que se propone construye servicios a partir del efecto sobre los costos del sistema que posee la omisión de un paradero. Esto involucra identificar los costos de los pasajeros beneficiados (pasajeros que pasan por el paradero omitido a lo largo de su viaje) versus los pasajeros que se ven afectados negativamente (pasajeros que tienen como origen o destino el paradero omitido). Los dos métodos que siguen trabajan sobre un formato establecido de servicio (servicios cortos y zonales) y evalúan los costos involucrados para determinar cuáles de estos servicios pueden resultar beneficiosos para el sistema. El cuarto enfoque propuesto analiza los ahorros potenciales de generar servicios expresos en función de los ahorros que ofrecen con respecto a una solución ya existente.

Los métodos que se proponen al final de esta sección, que caen dentro de la segunda categoría antes mencionada, evalúan los mismos formatos predefinidos pero estudiando el efecto de éstos cuando el corredor se encuentra operando a capacidad, centrando el análisis en el arco crítico del corredor.

En esta sección se asumirá, a menos que se indique lo contrario, que el modelo de optimización y la codificación de la red se encuentra separada por sentidos (es decir, se relaja la imposición de servicios cíclicos), y que la notación a utilizar corresponde a la introducida en la sección anterior al proponer esta variante del modelo. Más aún, para simplificar el análisis de los servicios cortos, se asumirá en general que los paraderos del corredor coinciden entre ambos sentidos. De todas formas, las heurísticas que se obtienen podrían ser aplicadas a casos más generales, modificando algunos detalles en la codificación de los corredores.

Para las heurísticas de diseño que acá se presentan, se asume que en general los pasajeros no realizan trasbordos, lo que permite simplificar la formulación del problema y facilitar el

análisis. En todo caso, los servicios generados son evaluados por el modelo de optimización de frecuencias, de forma que sus beneficios sí consideran trasbordos o el tipo de comportamiento que se estime conveniente.

3.1 Heurísticas para la generación de servicios

- **Generación de servicios expresos iniciales**

La primera heurística que se propone consiste en tomar un servicio regular y, primero, definir un indicador de qué tan atractivo resulta cada paradero para ser omitido, para luego usar este indicador para construir un conjunto de servicios expresos a utilizar como punto de partida para el problema.

Para obtener este indicador, consideremos el problema del corredor donde se desea diseñar un servicio expreso para acompañar a un servicio regular existente, sin restricción de capacidad. Supondremos que el corredor inicialmente opera con un único servicio regular con una frecuencia de f_t , y que se quiere destinar parte de esta frecuencia, f_e , a un servicio expreso que omita únicamente el paradero i (no extremo). En pasos posteriores se extiende el análisis para analizar la omisión de un paradero i en un servicio expreso que ya omita otros paraderos en su trazado. La variación en los costos totales debido a este cambio se puede analizar para las tres componentes de los costos totales: costos del operador, costos por espera, y costos por tiempo de viaje en vehículo. En la figura a continuación se ilustra la situación a estudiar.

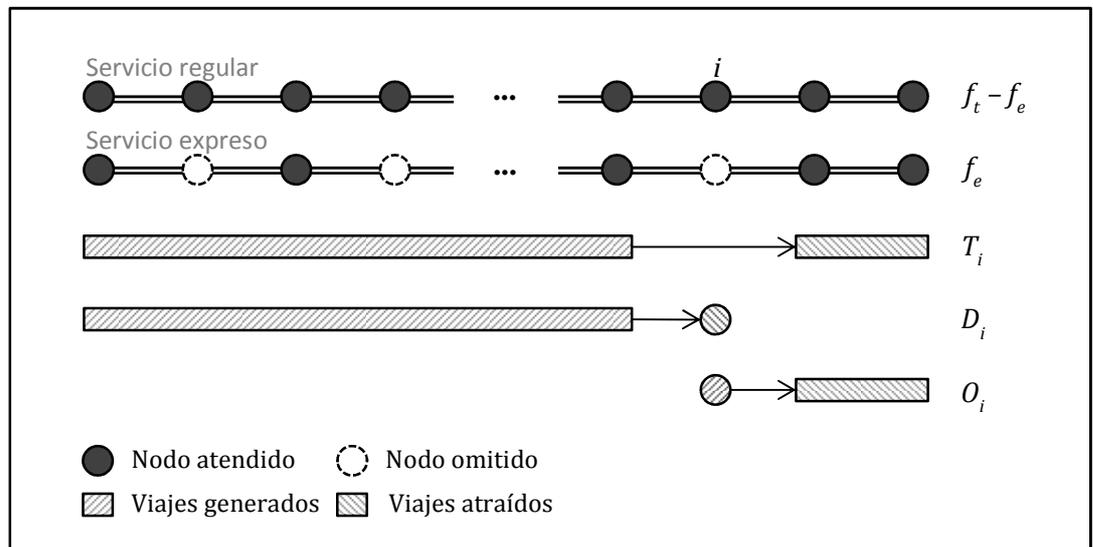


Figura 3-1: Generación de servicios iniciales

En la figura se muestran dos servicios, uno regular, y uno expreso (que en este caso ya omite algunos paraderos en su trazado) para el que se está considerando omitir, adicionalmente, el paradero i . La figura muestra también los movimientos afectados por esta medida, detallados más adelante.

Respecto a los costos del operador, estos disminuirán en función de la diferencia de costos de operar el servicio regular versus el servicio expreso. Considerando que los costos de operación están conformados por una parte dependiente del tiempo de ciclo y otra del largo de éste, como se muestra en la ecuación (2.3), el efecto de introducir una línea que omita un paradero corresponderá al ahorro de tiempo obtenido por esta omisión, es decir:

$$\Delta OC = -f_e c_T \tau \quad (3.1)$$

El beneficio por ahorro de tiempo proviene principalmente de una disminución en el tamaño de flota, ya que una de las principales componentes de c_T es la depreciación de los vehículos.

Para calcular costos del usuario se realizará una simplificación en su comportamiento, consistente en asumir que siempre el usuario utilizará el primer bus que pase y que lo lleve directo a su destino, sin considerar conjuntos de líneas atractivas ni trasbordos. Bajo este supuesto, el aumento en el tiempo de espera que experimenta un usuario que se ve forzado a esperar al servicio regular corresponde a:

$$\Delta WT = \frac{\lambda}{f_t - f_e} - \frac{\lambda}{f_t} = \frac{\lambda f_e}{(f_t - f_e) f_t} \quad (3.2)$$

Los usuarios que se ven obligados a esperar al servicio regular son los que su viaje comienza o termina en el paradero omitido i . Si O_i y D_i corresponden a los viajes que se originan y terminan en i (ilustrados en la Figura 3-1), y definimos OD_i como la suma de estos dos términos, la variación en los costos de espera será:

$$\Delta WTC = \theta_{wt} OD_i \frac{\lambda f_t}{(f_t - f_e) f_t} \quad (3.3)$$

Los tiempos de viaje se verán reducidos para los usuarios que su viaje pasa por el nodo i . Llamemos a este número de viajes T_i (ilustrados en la Figura 3-1). El ahorro de tiempo que perciben estos usuarios corresponderá a τ por la probabilidad de que el servicio expreso sea el primero en pasar. Con esto, la variación en costos por tiempos de viaje se pueden modelar como:

$$\Delta TTC = -T_i \theta_{tt} \tau \frac{f_e}{f_t} \quad (3.4)$$

Por lo tanto, la función de ahorros del sistema (que corresponde a la variación en los costos en negativo) producto de la omisión del paradero i será de:

$$S_i = f_e c_T \tau + T_i \theta_{tt} \tau \frac{f_e}{f_t} - OD_i \theta_{wt} \frac{\lambda f_e}{(f_t - f_e) f_t} \quad (3.5)$$

Podemos esperar que existan ahorros en el sistema cuando esta última función sea positiva. Dado que se desea conocer cuándo estos ahorros son mayores, lo lógico sería derivar esta función con respecto a f_e y f_t , y resolver el sistema de ecuaciones resultante. Sin embargo, este procedimiento no lleva a una solución explícita, por lo que se diseñó un método alternativo que lleva a una solución más sencilla. Calculamos en primer lugar la derivada de los ahorros con respecto a f_e :

$$\frac{\partial S_i}{\partial f_e} = c_T \tau + \frac{T_i \theta_{tt} \tau}{f_t} - \frac{OD_i \theta_{wt} \lambda}{(f_t - f_e)^2} \quad (3.6)$$

Con el propósito de estudiar la convexidad de la función, calculamos la segunda derivada de la función de ahorros:

$$\frac{\partial^2 S_i}{\partial f_e^2} = -\frac{2\lambda OD_i \theta_{wt}}{(f_t - f_e)^3} \quad (3.7)$$

Dado que por definición $f_t \geq f_e$, podemos concluir que la función es cóncava, es decir, que si la función de ahorros presenta un punto estacionario, este debe ser un máximo. Además, la función de ahorro vale cero en $f_e = 0$ y se hace infinitamente negativa cuando $f_e \rightarrow f_t$, debido a que las esperas de usuarios que no tienen acceso al servicio expreso se hacen infinitas. Combinando estos hechos, podemos concluir que la función de ahorro tiene una de las dos siguientes formas:

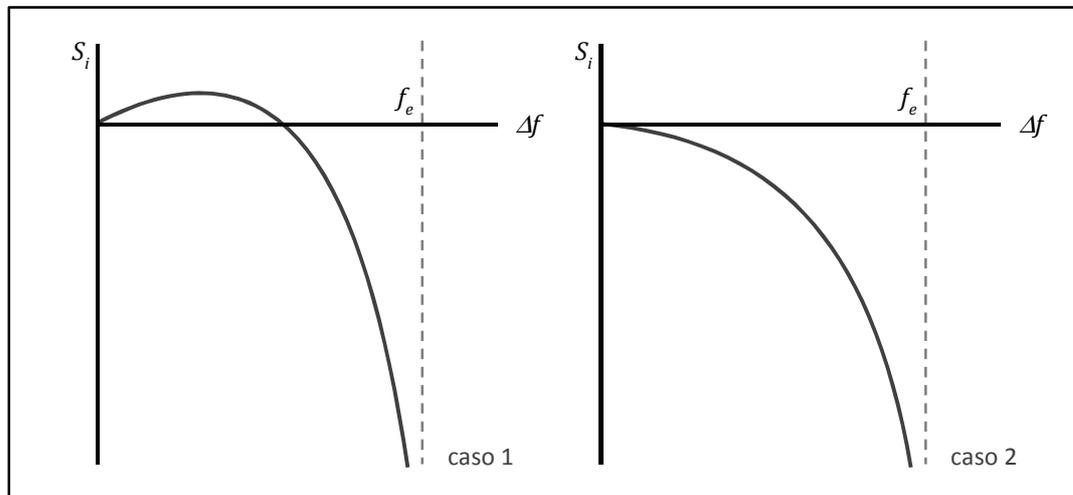


Figura 3-2: Función de ahorros

Para determinar a cuál de las dos situaciones corresponde un problema con f_t e i determinados, basta con evaluar la derivada de la función en el punto $f_e = 0$:

$$\left. \frac{\partial S_i}{\partial f_e} \right|_{f_e=0} = c_T \tau + \frac{T_i \theta_{tt} \tau}{f_t} - \frac{OD_i \theta_{wt} \lambda}{f_t^2} \quad (3.8)$$

Este indicador sirve para obtener una idea del potencial de cada paradero como para ser omitido por un servicio expreso: un mayor valor de este indicador implica una mayor pendiente en el gráfico de la Figura 3-2, lo que debiese corresponder a un mayor ahorro máximo. Como se puede apreciar, la expresión (3.8) depende de parámetros conocidos del corredor salvo por el valor de la frecuencia total f_t . Para evaluar el indicador se hace necesario escoger un valor de referencia para esta variable. Parece una opción sensata utilizar la frecuencia óptima para un único

servicio regular (obtenida a partir de la fórmula de la raíz de Mohring). Con esto, podemos definir el primer indicador para el diseño de servicios como:

$$\text{Ind1}_i = c_T \tau + \frac{T_i \theta_{tt} \tau}{f_t^*} - \frac{OD_i \theta_{wt} \lambda}{f_t^{*2}} \quad (3.9)$$

$$f_t^* = \sqrt{\frac{\theta_{wt} \lambda \sum_w T_w}{c_a}} \quad (3.10)$$

En esta expresión c_a corresponde al costo de operación del servicio regular y se puede obtener a partir de la ecuación (2.3). Este indicador representa sólo una aproximación del efecto de omitir un paradero por al menos tres motivos. En primer lugar, se ha simplificado el comportamiento de los usuarios, lo que implica que hay casos donde la fórmula (3.5) subestima los ahorros. Estos casos son los que el servicio expreso es suficientemente atractivo como para que los usuarios dejen de utilizar el servicio regular en algunos pares. Lo segundo es que el efecto de un paradero por separado no es el mismo que el de ese mismo paradero en conjunto con otros paraderos, ya que ciertas particularidades de la matriz de viajes pueden favorecer ciertas combinaciones de paraderos que un indicador como éste no percibe. Finalmente, como se explica al comienzo de esta sección, el cálculo de este indicador desconoce la posible existencia de trasbordos en el sistema.

Para abordar el segundo problema mencionado construiremos una nueva versión del indicador que considera el efecto de un conjunto de paraderos a omitir. Con este fin hacemos antes las siguientes definiciones:

- \mathcal{K} : Conjunto de paraderos omitidos por el servicio expreso.

- $N_{\mathcal{K}}$: Número de paraderos omitidos en el trazado del servicio expreso ($N_{\mathcal{K}} = |\mathcal{K}|$).
- $OD_{\mathcal{K}}$: Número de viajes con origen o destino dentro de \mathcal{K} .
- $Sk_w^{\mathcal{K}}$: Número de paraderos omitidos por el servicio expreso a lo largo del par w (sin contar extremos) dado un conjunto \mathcal{K} .
- $\delta_w^{\mathcal{K}}$: Variable binaria que indica si el par w está conectado por el servicio expreso definido por el conjunto \mathcal{K} , es decir, si sus paraderos de origen y de destino son atendidos por dicho servicio.
- $T_{\mathcal{K}}$: Número de saltos de parada que se realizarían si todos los usuarios que tienen acceso al servicio expreso lo utilizaran, independiente de la espera involucrada. Este valor se puede calcular de la siguiente manera:

$$T_{\mathcal{K}} = \sum_w \delta_w^{\mathcal{K}} T_w S k_w^{\mathcal{K}} \quad (3.11)$$

Con esto, la función de ahorros se puede generalizar para un conjunto de paraderos como:

$$\Delta SC_{\mathcal{K}} = f_e c_T N_{\mathcal{K}} \tau + T_{\mathcal{K}} \theta_{tt} \tau \frac{f_e}{f_t^*} - OD_{\mathcal{K}} \theta_{wt} \frac{\lambda f_e}{(f_t^* - f_e) f_t^*} \quad (3.12)$$

Notar que esta función tiene la misma estructura que la de la ecuación (3.5). El primer término, correspondiente al ahorro del operador, se ve modificado para considerar que el ahorro en tiempo de viaje sube de τ a $N_{\mathcal{K}} \tau$. El segundo término calcula el ahorro total de tiempo que se obtendría si todos los pasajeros que les sirve el servicio expreso lo utilizaran, $T_{\mathcal{K}} \theta_{tt} \tau$, y lo multiplica por la probabilidad de que el servicio expreso sea el primero en pasar. El último término sólo se ve modificado en

que el número de pasajeros afectados debe ser recalculado en función de los paraderos omitidos.

A partir de la última expresión obtenemos un nuevo indicador derivando y evaluando en $f_e = 0$:

$$\left. \frac{\delta \Delta SC_{\mathcal{K}}}{\delta f} \right|_{f_e=0} = c_T N_{\mathcal{K}} \tau + \frac{T_{\mathcal{K}} \theta_{tt} \tau}{f_t^*} - \frac{OD_{\mathcal{K}} \theta_{wt} \lambda}{f_t^{*2}} \quad (3.13)$$

A partir de (3.13) podemos definir una versión generalizada del indicador como:

$$\text{Ind1}_{\mathcal{K}} = c_T N_{\mathcal{K}} \tau + \frac{T_{\mathcal{K}} \theta_{tt} \tau}{f_t^*} - \frac{OD_{\mathcal{K}} \theta_{wt} \lambda}{f_t^{*2}} \quad (3.14)$$

Para comparar Ind1_i e $\text{Ind1}_{\mathcal{K}}$, tomamos la expresión (3.9) y sumamos para todos los paraderos del conjunto \mathcal{K} :

$$\sum_{i \in \mathcal{K}} \text{Ind1}_i = \sum_{i \in \mathcal{K}} \left[f_e c_T \tau + T_i \theta_{tt} \tau \frac{f_e}{f_t^*} - OD_i \theta_{wt} \frac{\lambda f_e}{(f_t^* - f_e) f_t^*} \right] \quad (3.15)$$

$$= \sum_{i \in \mathcal{K}} [f_e c_T \tau] + \sum_{i \in \mathcal{K}} \left[T_i \theta_{tt} \tau \frac{f_e}{f_t^*} \right] - \sum_{i \in \mathcal{K}} \left[OD_i \theta_{wt} \frac{\lambda f_e}{(f_t^* - f_e) f_t^*} \right] \quad (3.16)$$

$$= f_e c_T N_{\mathcal{K}} \tau + \theta_{tt} \tau \frac{f_e}{f_t^*} \sum_{i \in \mathcal{K}} T_i - OD_{\mathcal{K}} \theta_{wt} \frac{\lambda f_e}{(f_t^* - f_e) f_t^*} \quad (3.17)$$

Esta suma difiere del ahorro de la fórmula (3.12) tan sólo en el término $\sum_{i \in \mathcal{K}} T_i$ que aparece como $T_{\mathcal{K}}$ en la fórmula del efecto conjunto. Es fácil demostrar que $\sum_{i \in \mathcal{K}} T_i \geq T_{\mathcal{K}}$; basta con considerar que un viaje en un par w que pasa por un paradero omitido $i \in \mathcal{K}$ siempre es considerado en T_i , pero podría ser omitido en $T_{\mathcal{K}}$ porque w puede tener su origen o destino en un paradero omitido. Esto implica que si deseamos estudiar el efecto de la omisión de un paradero cuando ya existen otros paraderos omitidos en el servicio expreso, Ind1_i tenderá a sobreestimar estos beneficios. Esto

es importante de tener en cuenta al momento de construir una heurística de generación.

Como se mencionaba antes, los resultados acá expuestos son válidos siempre y cuando los usuarios que realizan un viaje donde el servicio expreso está disponible optan por tomar el primer servicio que pasa, sea éste regular o expreso. Esto descarta la posibilidad de que el usuario prefiera esperar el servicio expreso y dejar pasar el servicio regular. Esta última opción puede mejorar el tiempo esperado de viaje del usuario cuando el servicio expreso ofrece ahorros considerables de tiempo. De hecho, es posible determinar los casos en que esto sucede, comparando el tiempo total esperado de viaje de ambas opciones. Esto sucederá en los pares donde el número de paraderos omitidos supere el umbral X_s presentado en el Lema 3, ecuación (2.24).

Para conocer la magnitud del error que se comete al asumir que los usuarios usan siempre ambos servicios, compararemos los tiempos esperados de viaje para cada tipo de comportamiento en un determinado par (ecuaciones (2.31) y (2.32)):

$$\varepsilon_w = \frac{f_a t_a^w + f_e t_e^w + \lambda}{f_a + f_e} - \left(t_e^w + \frac{\lambda}{f_e} \right) \quad (3.18)$$

Desarrollando la expresión se obtiene que:

$$\varepsilon_w = (t_a^w - t_e^w) \left(1 - \frac{f_e}{f_t} \right) + \frac{\lambda}{f_t} - \frac{\lambda}{f_e} \quad (3.19)$$

La diferencia de tiempos para ambos servicios corresponde al número de paraderos omitidos, $Sk_w^{\mathcal{K}}$, por el tiempo de detención. De esta forma se tiene que:

$$\varepsilon_w = Sk_w^{\mathcal{K}}\tau \left(1 - \frac{f_e}{f_t}\right) + \frac{\lambda}{f_t} - \frac{\lambda}{f_e} \quad (3.20)$$

Tomando f_t fijo, este error se maximiza cuando:

$$\frac{\delta \varepsilon_w}{\delta f_e} = 0 \rightarrow -\frac{Sk_w^{\mathcal{K}}\tau}{f_t} + \frac{\lambda}{f_e^2} = 0 \rightarrow f_e^* = \sqrt{\frac{f_t \lambda}{Sk_w^{\mathcal{K}}\tau}} \quad (3.21)$$

Reemplazando f_e^* en ε y desarrollando:

$$\varepsilon_w|_{f_e=f_e^*} = \varepsilon_w^{max} = Sk_w^{\mathcal{K}}\tau \left(1 - \sqrt{\frac{\lambda}{Sk_w^{\mathcal{K}}\tau f_t}}\right) + \frac{\lambda}{f_t} - \sqrt{\frac{Sk_w^{\mathcal{K}}\tau \lambda}{f_t}} \quad (3.22)$$

$$\varepsilon_w^{max} = \left(\sqrt{Sk_w^{\mathcal{K}}\tau} - \sqrt{\frac{\lambda}{f_t}}\right)^2 \quad (3.23)$$

Este último valor representa una cota máxima para el error cometido por la simplificación del comportamiento de usuario en determinado par origen destino. Naturalmente, este error es mayor mientras más paradas sean omitidas o mientras mayor sea la frecuencia total de los buses sobre el corredor.

Con los indicadores propuestos en las fórmulas (3.9) y (3.14) se pueden construir servicios expresos para el corredor a partir de una heurística sencilla de tipo *greedy* (ver, por ejemplo, Leiserson et al., 2001). Si definimos como $\mathcal{L}g$ al conjunto donde registraremos los servicios generados, la heurística paso a paso se detalla a continuación:

Heurística para la generación de servicios HG_{1a} :

1. Considerar un servicio que inicialmente atiende todos los paraderos entre dos paraderos extremos dados, que llamaremos l . Llamaremos \mathcal{K} al conjunto de paraderos omitidos por este servicio entre sus extremos, inicialmente vacío.
2. Evaluar para cada paradero i donde el servicio se detiene (excluyendo sus extremos) el indicador $\text{Ind}1_i$, y determinar para cuál paradero \hat{i} este valor es máximo, como se muestra en la expresión (3.24). El valor de f_t^* se obtiene de la fórmula (3.10).

$$\hat{i} = \underset{i \in \mathcal{N}}{\text{argmax}} \left[c_T \tau + \frac{T_i \theta_{tt} \tau}{f_t^*} - \frac{OD_i \theta_{wt} \lambda}{f_t^{*2}} \right] \quad (3.24)$$

3. Hacer $\mathcal{K} \leftarrow \mathcal{K} \cup \{\hat{i}\}$, $\mathcal{L}g \leftarrow \mathcal{L}g \cup \{l\}$ y volver al paso 2 hasta que queden sólo los paraderos extremos del corredor.

Al final de la ejecución del algoritmo se obtiene un conjunto $\mathcal{L}g$ de servicios para entregar al modelo de optimización de frecuencias. Esta primera versión del algoritmo utiliza la versión sencilla del indicador, a pesar de que, como se demostró anteriormente, entrega una sobreestimación de los ahorros cuando el expreso omite otros paraderos. Sin embargo, dado que esta sobreestimación afecta a todos los paraderos y que la heurística trabaja jerarquizando los paraderos a partir de este indicador, este problema no debiese ser grave si esta sobreestimación es similar entre distintos paraderos. De todas formas, el algoritmo también puede ser adaptado para considerar el indicador para un conjunto de paraderos, cambiando el indicador en el paso 2. Ambas versiones del algoritmo serán comparadas más adelante para verificar si la simplificación antes hecha es razonable o no.

Dado que la heurística trabaja a partir de un ranking construido a partir de los indicadores, la parte del indicador correspondiente a los costos del operador, que es constante, puede ser omitida en el análisis.

Heurística para la generación de servicios HG_{1b} :

1. Considerar un servicio l que inicialmente atiende todos los paraderos entre dos paraderos extremos dados. Llamaremos \mathcal{K} al conjunto de paraderos omitidos por este servicio entre sus extremos, inicialmente vacío.
2. Evaluar para cada paradero i donde el servicio se detiene (excluyendo sus extremos) el indicador $\text{Ind}1_{\mathcal{K} \cup \{i\}}$, y determinar para cuál paradero \hat{i} este valor es máximo.
3. Hacer $\mathcal{K} \leftarrow \mathcal{K} \cup \{\hat{i}\}$, $\mathcal{L}g \leftarrow \mathcal{L}g \cup \{l\}$ y volver al paso 2 hasta que queden sólo los paraderos extremos del corredor.

Este último indicador en principio debería entregar mejores resultados, aunque implica algunos cálculos adicionales respecto a la heurística anterior, ya que en ésta cada vez que se agrega un paradero a \mathcal{K} se debe volver a calcular los valores de $T_{\mathcal{K}}$ y $OD_{\mathcal{K}}$.

Como se mencionó anteriormente, las heurísticas acá presentadas siguen una lógica tipo *greedy*, donde en cada etapa el algoritmo elimina el paradero cuyo indicador lo indica como más promisorio. Esta heurística podría ser fácilmente adaptada para explorar nuevas soluciones por medio de, por ejemplo, la implementación de una

heurística tipo GRASP que corresponde a una heurística *greedy* con un componente aleatorio (ver Kochenberger, 2003).

- **Generación de servicios cortos**

Un tipo específico de servicio que suele ser utilizado para aumentar la capacidad y la frecuencia en tramos más cargados de un corredor corresponde a los servicios cortos, también conocidos como bucles. A continuación desarrollaremos un criterio para identificar qué servicios cortos pueden ser beneficiosos para implementar en un corredor, de forma de poder entregárselos al modelo base de optimización de frecuencias.

Definiremos al conjunto de los servicios cortos factibles para el corredor como \mathcal{L}_C . Debido a que este tipo de servicios tienen un efecto en ambos sentidos del corredor, asumiremos que los servicios se encuentran definidos como ciclos completos, sin separación por sentido. Supongamos que en un corredor opera un servicio regular con frecuencia f_a y costos de operación c_a , y un servicio corto $c \in \mathcal{L}_C$ de frecuencia f_c y costos de operación c_c . Este servicio corto opera entre los paraderos i_1 y j_1 en el sentido de ida, e i_2 y j_2 en el sentido de vuelta. Los costos de operación c_c , que se obtienen de acuerdo a la fórmula (2.1), pueden incorporar los costos asociados a las maniobras de retorno dependientes de los nodos extremos del servicio.

Es posible clasificar a los usuarios de este corredor en dos categorías: los usuarios cuyo viaje se encuentra contenido en el servicio corto y que utilizarán ambos sentidos, cuyo número denotaremos como T_{AC}^c , y el resto los usuarios, que representaremos por T_A^c y que están obligados a utilizar sólo el servicio regular. En lo

que sigue, se utilizarán los índices de servicios en mayúsculas cuando se refieran a un tipo de comportamiento de usuarios, como en este caso, donde el índice AC quiere decir “utiliza el servicio a o el servicio c ”. Definiremos como \mathcal{W}_{AC}^c el conjunto de pares correspondientes al primero de estos comportamientos, es decir, $w \in \mathcal{W}_{AC}^c \Leftrightarrow (O_w \geq i_1 \wedge D_w \leq j_1) \vee (O_w \geq i_2 \wedge D_w \leq j_2)$. El conjunto de los pares asociados al otro comportamiento, \mathcal{W}_A^c , se puede definir como el complemento del conjunto \mathcal{W}_{AC}^c , es decir, $\mathcal{W}_A^c = \mathcal{W} - \mathcal{W}_{AC}^c$. Las fórmulas (3.25) y (3.26) muestran cómo calcular las demandas T_{AC}^c y T_A^c . Notar que estos valores dependen del servicio corto c en consideración.

$$T_{AC}^c = \sum_{w \in \mathcal{W}_{AC}^c} T_w \quad (3.25)$$

$$T_A^c = \sum_{w \in \mathcal{W}_A^c} T_w \quad (3.26)$$

Un esquema de los servicios a optimizar y una representación gráfica de los parámetros de demanda recién definidos se pueden ver en la figura a continuación.

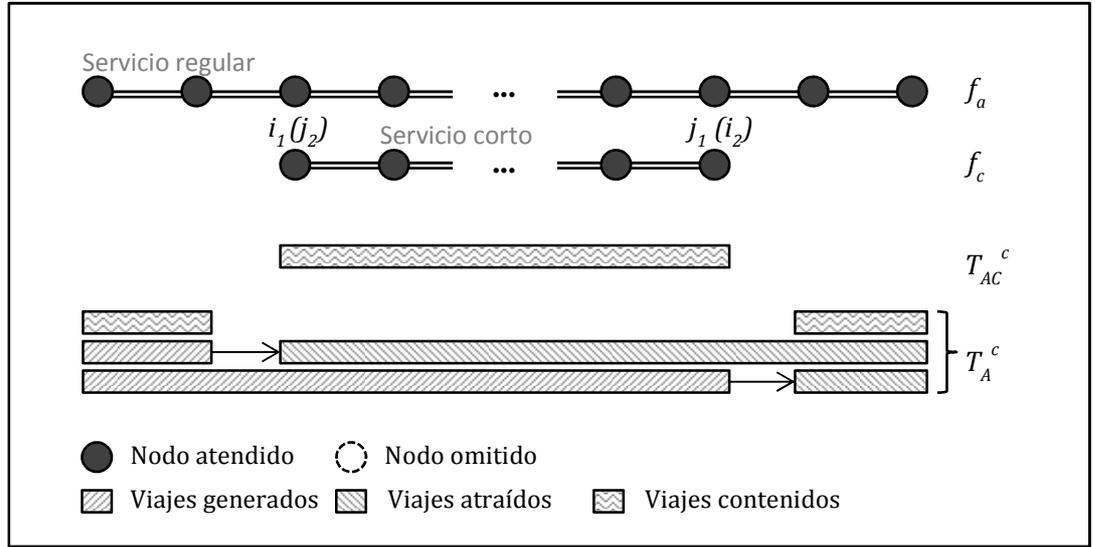


Figura 3-3: Generación de servicios cortos

En la figura anterior los bloques marcados como viajes contenidos representan todos los viajes que poseen tanto su origen como su destino dentro de las regiones pintadas.

Debido a que ambos servicios entregan los mismos tiempos de viaje en los pares que atienden, los costos totales del sistema por concepto de tiempos de viaje son constantes, y podrán ser omitidos del problema. A partir de las definiciones hechas podemos escribir la función de costos a optimizar como:

$$SC = f_a c_a + f_c c_c + \frac{\theta_{wt} \lambda T_A^c}{f_a} + \frac{\theta_{wt} \lambda T_{AC}^c}{f_a + f_c} \quad (3.27)$$

Para minimizar estos costos calcularemos las derivadas respecto a f_a y f_c e igualaremos a cero:

$$\frac{\delta SC}{\delta f_a} = 0 \rightarrow c_a - \frac{\theta_{wt} \lambda T_A^c}{f_a^2} - \frac{\theta_{wt} \lambda T_{AC}^c}{(f_a + f_c)^2} = 0 \quad (3.28)$$

$$\frac{\delta SC}{\delta f_c} = 0 \rightarrow c_e - \frac{\theta_{wt}\lambda T_{AC}^c}{(f_a + f_c)^2} = 0 \quad (3.29)$$

Resolviendo el sistema de ecuaciones para (3.28) y (3.29) se obtiene que:

$$f_a^* = \sqrt{\frac{\theta_{wt}T_A^c\lambda}{c_a - c_c}} \quad (3.30)$$

$$f_c^* = \sqrt{\frac{\theta_{wt}T_{AC}^c\lambda}{c_c}} - \sqrt{\frac{\theta_{wt}T_A^c\lambda}{c_a - c_c}} \quad (3.31)$$

Para que este resultado tenga sentido, se debe imponer que $f_c^* > 0$. Si esta última condición no se cumple, entonces el servicio corto no puede ser beneficioso para el sistema.

Reemplazando f_a^* y f_c^* en (3.27) podemos obtener una expresión para los costos óptimos del sistema operando con el servicio corto en evaluación:

$$SC_c^* = 2\sqrt{(c_a - c_c)\lambda\theta_{wt}T_A^c} + 2\sqrt{c_c\lambda\theta_{wt}T_{AC}^c} \quad (3.32)$$

Lo que nos interesa no es diseñar un único servicio corto, sino que generar un conjunto de servicios para alimentar al modelo de optimización de frecuencias. Sin embargo, la condición $f_c^* > 0$ nos entrega un buen criterio para decidir cuáles servicios parecen más interesantes de considerar para la optimización. Desarrollando esta desigualdad, se obtiene el siguiente indicador para la generación de servicios cortos:

$$\text{Ind}2_c = \frac{T_{AC}^c}{T_A^c} - \frac{c_c}{c_a - c_c} > 0 \quad (3.33)$$

El procedimiento para definir qué servicios cortos entregarle al modelo base consistiría entonces en calcular $\text{Ind}2_c$ para cada servicio corto posible del corredor, e

incluir en la modelación los servicios donde este indicador arroje un valor positivo. Adicionalmente, estos servicios cortos generados se pueden usar para construir servicios expresos cortos, aplicando sobre ellos la heurística de generación de servicios detallada para el primer indicador.

Heurística para la generación de servicios HG_2 :

1. Para cada servicio corto $c \in \mathcal{L}_c$ evaluar el indicador $\text{Ind}2_c$.
2. Si $\text{Ind}2_c > 0$, hacer $\mathcal{L}g \leftarrow \mathcal{L}g \cup \{c\}$.

El conjunto $\mathcal{L}g$ resultante contendrá los servicios cortos que parezcan relevantes para el corredor en estudio.

- **Generación de servicios zonales**

Un formato de servicio que parece promisorio en muchos escenarios consiste en el servicio que atiende ciertos paraderos al principio y al final de su trazado, omitiendo un número de paraderos seguidos en su tramo intermedio. Este tipo de servicio entrega importantes ahorros en tiempo de viaje para un cierto número de usuarios, y es más barato de operar que un servicio regular debido a su menor tiempo de ciclo, aunque destinar parte de la flota de buses a este tipo de servicios puede conllevar un aumento en los tiempos de espera de algunos usuarios. En el contexto de esta tesis nos referiremos a este tipo de servicio en particular como servicio zonal.

Supongamos que se conoce el conjunto de servicios zonales factibles para el corredor, \mathcal{L}_E . Un servicio zonal por definición atenderá una serie de paraderos al

comienzo del corredor, después omitirá N_e paraderos en el tramo intermedio del corredor, y después atenderá todos los paraderos restantes. Denotaremos como i_e y j_e al primero y al último de estos paraderos omitidos. Si suponemos que este servicio opera en paralelo a un servicio regular, podemos clasificar los tipos de viaje dentro de tres categorías. Llamaremos T_A^e la demanda que sólo puede ser atendida por el servicio regular (es decir, la que se origina o termina en alguno de los paraderos omitidos por el expreso), y T_E^e a la demanda por viajes que utilizarán el servicio expreso comenzando en uno de los paraderos del comienzo del corredor y terminando en el tramo final (asumiremos que el servicio expreso es suficientemente atractivo como para que en estos pares el servicio regular no resulte conveniente de utilizar). Finalmente, llamaremos T_{AE}^e a la demanda que puede ser atendida indistintamente por ambos servicios, que corresponde a la demanda contenida por los tramos inicial y final del servicio expreso. Las demandas T_A^e , T_E^e y T_{AE}^e se pueden calcular mediante de las siguientes fórmulas:

$$T_E^e = \sum_{w \in \mathcal{W}_E^e} T_w \quad (3.34)$$

$$T_{AE}^e = \sum_{w \in \mathcal{W}_{AE}^e} T_w \quad (3.35)$$

$$T_A^e = \sum_{w \in \mathcal{W}_A^e} T_w \quad (3.36)$$

Los conjuntos utilizados en las fórmulas anteriores contienen a los pares de la matriz que caen dentro de cada tipo de comportamiento de usuario. Específicamente, el conjunto \mathcal{W}_E^e se puede definir como $w \in \mathcal{W}_E^e \Leftrightarrow (O_w < i_e \wedge D_w > j_e)$, el conjunto

\mathcal{W}_{AE}^e como $w \in \mathcal{W}_{AE}^e \Leftrightarrow (D_w < i_e \vee O_w > j_e)$, y el conjunto \mathcal{W}_A^e como $\mathcal{W}_A^e = \mathcal{W} - \mathcal{W}_E^e - \mathcal{W}_{AE}^e$. Las demandas y tipos de servicio se resumen en la figura que sigue.

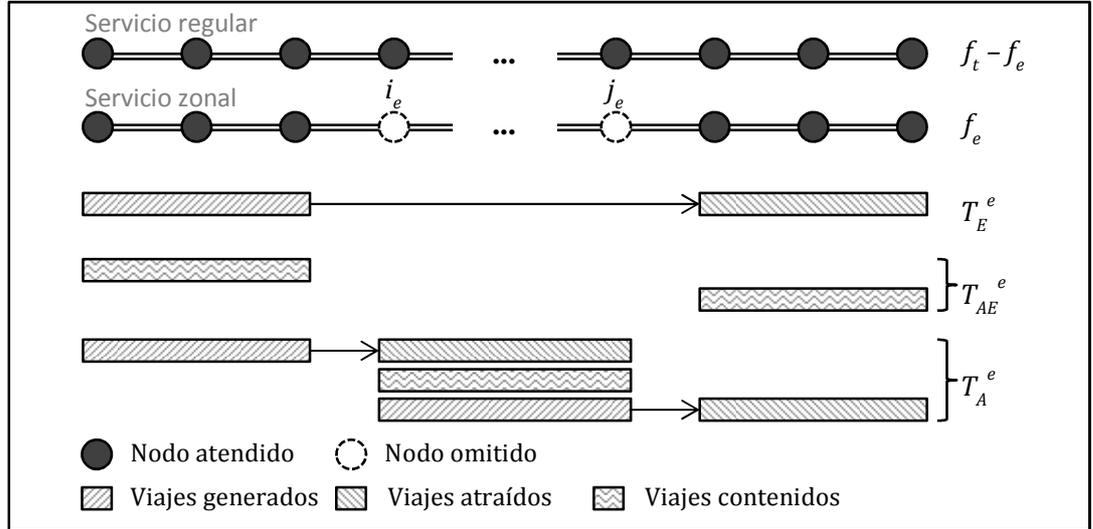


Figura 3-4: Generación de servicios zonales

Minimizar los costos por tiempo de viaje equivale en este caso a maximizar el ahorro en tiempo por los paraderos omitidos. Considerando que estos ahorros corresponden a la omisión de N_e paradas de duración τ , los ahorros en tiempos de viaje que genera esta medida son de:

$$\Delta TTC = \theta_{tt} T_E^e N_e \tau \quad (3.37)$$

La función de costos sociales corregida a optimizar en este caso corresponde a:

$$SC = f_a c_a + f_e c_e + \lambda \theta_{wt} \left(\frac{T_A^e}{f_a} + \frac{T_E^e}{f_e} + \frac{T_{AE}^e}{f_a + f_e} \right) - \theta_{tt} T_E^e N_e \tau \quad (3.38)$$

En principio, ya que deseamos minimizar esta función de costos sociales, podríamos evaluar sus derivadas parciales, igualar a cero, y resolver el sistema de ecuaciones resultante. Sin embargo, esta metodología no lleva a una solución explícita, de forma

que se optó por trabajar con una versión simplificada de esta función. Dado que la principal justificación para la implementación de un servicio zonal es beneficiar a los usuarios que lo utilizan para viajar de extremo a extremo del corredor, asumiremos que el beneficio de los usuarios T_{AE}^e por utilizar el servicio zonal puede ser despreciado. En otras palabras, asumiremos que estos usuarios utilizarán sólo el servicio regular, lo que implica que formarían parte de T_A^e . Este supuesto de comportamiento corresponde a una subestimación de los beneficios de los zonales ya que estos usuarios podrían disminuir sus tiempos de espera utilizando este último servicio. Con esto, la nueva función objetivo del problema es:

$$SC = f_a c_a + f_e c_e + \frac{\lambda \theta_{wt} (T_A^e + T_{AE}^e)}{f_a} + \frac{\lambda \theta_{wt} T_E^e}{f_e} - \theta_{tt} T_E^e N_e \tau \quad (3.39)$$

Derivando por las frecuencias e igualando a cero podemos obtener las condiciones de optimalidad del problema:

$$\frac{\delta SC}{\delta f_a} = 0 \rightarrow f_a^* = \sqrt{\frac{\lambda \theta_{wt} (T_A^e + T_{AE}^e)}{c_a}} \quad (3.40)$$

$$\frac{\delta SC}{\delta f_e} = 0 \rightarrow f_e^* = \sqrt{\frac{\lambda \theta_{wt} T_E^e}{c_e}} \quad (3.41)$$

Reemplazando estos valores en la función objetivo obtenemos el valor óptimo del problema para una determinada configuración del servicio zonal:

$$SC_e^* = 2\sqrt{\lambda \theta_{wt} (T_A^e + T_{AE}^e) c_a} + 2\sqrt{\lambda \theta_{wt} T_E^e c_e} - \theta_{tt} T_E^e N_e \tau \quad (3.42)$$

Deseamos conocer si el servicio zonal en el óptimo representa o no una mejora con respecto a la solución base donde sobre el corredor opera un único servicio regular.

De la fórmula de la raíz (3.10) se desprende que los costos óptimos de este corredor operado por un servicio regular corresponderán a:

$$SC_a^* = 2 \sqrt{\lambda \theta_{wt} c_a \sum_{w \in \mathcal{W}} T_w} \quad (3.43)$$

Con esto, el criterio para determinar qué servicios zonales resultarían beneficiosos al operar en conjunto con un servicio regular consistiría en calcular SC_e^* para diferentes configuraciones de servicios zonales, y seleccionar los que sean menores que SC_a^* . Para hacer completo este análisis sería necesario tomar en cuenta que suponer que el servicio zonal es atractivo por sí solo implica una cota mínima para f_e^* , que se puede desprender de (2.30); **Error! No se encuentra el origen de la referencia.** Sin embargo, lo que se desea obtener con este indicador no es un único servicio zonal para operar en conjunto con un servicio regular, sino que un conjunto de servicios que puedan ser beneficiosos para incorporar dentro del modelo de optimización de frecuencias. Es por esto que no se hace necesario imponer una frecuencia mínima para el expreso en el análisis ni compensar por los viajes despreciados en el análisis, ya que el modelo de optimización de frecuencias determinará cuáles servicios son efectivamente los que vale la pena utilizar.

De acuerdo a esto, el indicador asociado a la generación de servicios zonales es el siguiente:

$$\begin{aligned} \text{Ind}3_c = & 2\sqrt{\lambda \theta_{wt} (T_A^e + T_{AE}^e) c_a} + 2\sqrt{\lambda \theta_{wt} T_E^e c_e} - \theta_{tt} T_E^e N_e \tau \\ & - 2 \sqrt{\lambda \theta_{wt} c_a \sum_{w \in \mathcal{W}} T_w} \end{aligned} \quad (3.44)$$

La heurística para generar servicios zonales para alimentar al modelo corresponderá entonces a la siguiente:

Heurística para la generación de servicios HG_3 :

1. Para cada servicio zonal $e \in \mathcal{L}_E$ evaluar el indicador $\text{Ind}3_c$.
2. Si $\text{Ind}3_c < 0$, hacer $\mathcal{L}g \leftarrow \mathcal{L}g \cup \{e\}$.

- **Generación de servicios adicionales**

Los indicadores anteriores sirven para construir servicios que se le pueden entregar al modelo de optimización de frecuencias para una primera etapa del diseño. Una vez entregados estos servicios al modelo, resultaría interesante contar con otro tipo de criterio que permitiera diseñar nuevos servicios tomando de alguna forma en cuenta la mejor solución vigente. El indicador que se presenta en esta sección hace esto buscando, a partir de los tiempos de viaje de una solución de referencia, qué paraderos se pueden agregar a un servicio de referencia de forma de generar nuevas conexiones expresas que entreguen ahorros respecto de los tiempos de viaje existentes.

Con este fin, supondremos que conocemos una solución actual para el corredor, es decir, un conjunto de servicios con sus frecuencias, y con su correspondiente asignación de usuarios y tiempos de viaje para cada par. Aparte del conjunto de servicios actuales, definiremos un servicio base de referencia sobre el cual se irá agregando gradualmente nuevos paraderos. Este servicio no necesariamente debe ser

parte de la solución actual, y como punto de partida asumiremos que es el servicio que va de extremo a extremo del corredor sin realizar detenciones intermedias. Este servicio de referencia, que denotaremos como r , será modificado en cada iteración hasta encontrar un servicio que represente una mejora para el sistema.

Para poder en cada iteración escoger qué paradero agregar al servicio de referencia, mediremos el efecto marginal de incorporar un nuevo servicio que atiende los paraderos del servicio de referencia más un paradero i . Como aproximación, asumiremos que los principales afectados por esta medida son los usuarios de los pares que conectan con el nuevo paradero, que perciben los mayores ahorros en tiempos de viaje como consecuencia de esta incorporación. Debido a que el indicador será utilizado para decidir qué paradero agregar en cada iteración, y a que el costo para el operador de agregar una detención es la misma independiente de dónde, el costo del operador es excluido de la formulación de este indicador.

Consideremos en primer lugar un único par w que tiene como origen o destino el nodo i . Llamaremos tt_w al tiempo de viaje de este par para la solución actual, y ett_w^r al tiempo de viaje que ofrecería el servicio de referencia en el par w si atendiese al nodo i . El valor de ett_w^r puede ser calculado como t_r^w usando la fórmula (2.1). El valor de tt_w se puede calcular como el promedio ponderado de los tiempos de los servicios atractivos existentes, es decir:

$$tt_w = \frac{\sum_{l \in \mathcal{L}} t_l^w f_l^w}{\sum_{l \in \mathcal{L}} f_l^w} \quad (3.45)$$

El ahorro total de costos por tiempos de viaje percibidos en este par, suponiendo que los conjuntos de servicios atractivos no cambian (lo que corresponde a una aproximación conservadora de estos ahorros), será:

$$\Delta TTC_w = \theta_{tt} T_w (tt_w - ett_w^r) \frac{f_r}{f_r + f_w} \quad (3.46)$$

$$f_w = \sum_{l \in \mathcal{L}} f_l^w \quad (3.47)$$

El término f_w corresponde a la suma de las frecuencias de los servicios atractivos del par w . En la expresión (3.46) se asume en que los usuarios no dejarán de utilizar los servicios atractivos actuales por la inclusión del nuevo servicio expreso. Con respecto a los tiempos de espera, el ahorro percibido en costos de espera para usuarios de este par será de:

$$\Delta WTC_w = \lambda \theta_{wt} T_w \frac{f_r}{(f_r + f_w) f_w} \quad (3.48)$$

El ahorro total que genera incorporar el nodo i al servicio r entonces será:

$$S_i^r = \sum_{j \in \mathcal{N}, j < i} \pi_r^j (\Delta TTC_{(j,i)} + \Delta WTC_{(j,i)}) + \sum_{j \in \mathcal{N}, j > i} \pi_r^j (\Delta TTC_{(i,j)} + \Delta WTC_{(i,j)}) \quad (3.49)$$

El parámetro binario π_r^j indica si el servicio l se detiene en el paradero i . Para simplificar la notación, definimos un parámetro adicional δ_w^{ri} que indique si el par w va desde un nodo atendido por r hasta el nodo i , o desde el nodo i a un nodo atendido por r . Con esto (3.49) puede ser reescrito como:

$$S_i^r = \sum_{w \in \mathcal{W}} \delta_w^{ri} (\Delta TTC_w + \Delta WTC_w) \quad (3.50)$$

En forma similar a lo que se hizo para obtener el indicador para generar líneas iniciales, queremos determinar cuál es el servicio cuyo efecto marginal es mayor al ser incorporado, por lo que derivamos el ahorro y evaluamos en $f_r = 0$.

$$\frac{\delta S_i^r}{\delta f_r} = \sum_{w \in \mathcal{W}} \delta_w^{ri} \left(\frac{\delta \Delta TTC_w}{\delta f_r} + \frac{\delta \Delta WTC_w}{\delta f_r} \right) \quad (3.51)$$

Evaluamos en primer lugar la derivada para los tiempos de viaje:

$$\frac{\delta \Delta TTC_w}{\delta f_r} = \theta_{tt} T_w (tt_w - ett_w) \frac{f_w}{(f_r + f_w)^2} \quad (3.52)$$

$$\left. \frac{\delta \Delta TTC_w}{\delta f_r} \right|^{f_r=0} = \frac{\theta_{tt} T_w (tt_w - ett_w)}{f_w} \quad (3.53)$$

Y luego el término del tiempo de espera:

$$\frac{\delta \Delta WTC_w}{\delta f_r} = \theta_{tt} T_w \lambda \frac{1}{(f_r + f_w)^2} \quad (3.54)$$

$$\left. \frac{\delta \Delta WTC_w}{\delta f_r} \right|^{f_r=0} = \frac{\theta_{tt} T_w \lambda}{f_w^2} \quad (3.55)$$

Con esto, el indicador para generar nuevos servicios es el siguiente:

$$\text{Ind4}_i^r = \sum_{w \in \mathcal{W}} \delta_w^{ri} \left(\frac{\theta_{tt} T_w (tt_w - ett_w)}{f_w} + \frac{\theta_{tt} T_w \lambda}{f_w^2} \right) \quad (3.56)$$

Un mayor valor de este indicador implica que el paradero i es potencialmente más beneficioso al ser agregado al servicio de referencia.

Este indicador permite construir una nueva heurística para el diseño de servicios.

Para la ejecución de esta heurística, es necesario definir en primer lugar qué servicios se desea utilizar como puntos de partida. Como el indicador está diseñado para

agregar paraderos a un servicio de referencia, resulta lógico considerar como puntos de partida servicios sin detenciones. En una versión sencilla de esta heurística bastaría con considerar en este conjunto de servicios de referencia, que llamaremos \mathcal{L}_R , a los dos servicios expresos que visitan los dos paraderos extremos del corredor, en ambos sentidos. Si se desea realizar una búsqueda más exhaustiva se pueden agregar a este conjunto los servicios que visitan los paraderos extremos de los servicios cortos donde $\text{Ind}2_c > 0$. Con esto, la heurística para la generación de servicios adicionales queda como sigue:

Heurística para la generación de servicios HG_A :

1. Seleccionar como servicio de referencia un nuevo servicio $r \in \mathcal{L}_R$. Si no existen más servicios de referencia, detener el algoritmo.
2. Hacer $\mathcal{L} \leftarrow \mathcal{L} \cup \{r\}$ y optimizar las frecuencias del corredor.
3. Si se obtiene mejora en los costos sociales del sistema (considerando todos sus componentes), se ha encontrado un nuevo servicio beneficioso: por lo que se mantiene el nuevo servicio dentro de \mathcal{L} y se vuelve a comenzar el algoritmo desde el comienzo (desde el primer servicio de referencia). Si no se obtiene mejora, hacer $\mathcal{L} \leftarrow \mathcal{L} - \{r\}$ y seguir al paso siguiente.
4. Calcular para cada paradero i el valor de $\text{Ind}4_i^r$.
5. Escoger entre los paraderos omitidos por el servicio y que estén contenidos entre sus extremos el que presente un mayor valor para el indicador (llamaremos a este paradero \hat{i}), hacer $r \leftarrow r \cup \{\hat{i}\}$, y volver al paso 2. De no existir paraderos omitidos en el servicio, volver al paso 1.

Observando la fórmula (3.56) se puede ver que $\text{Ind}4_i^r$ será más alto mientras mayor sea la diferencia $(tt_w - ett_w^r)$, lo que sucede cuando el número de detenciones del expreso de referencia en el par sea menor. Esto implica que el algoritmo debería funcionar mejor cuando los servicios de referencia tienen todavía pocas detenciones. Este hecho se podría utilizar para establecer un criterio para acelerar el algoritmo, donde se abandone un servicio de referencia cuando el número de paraderos omitidos alcance cierto umbral (tal como está planteado el algoritmo, este umbral valdría 0).

Es importante notar que este método es considerablemente más lento de ejecutar que los antes propuestos, ya que para cada servicio a evaluar se debe realizar una optimización de frecuencias.

Al igual que con la heurística de generación de servicios expresos iniciales, esta heurística podría ser generalizada para incorporar aleatoriedad en el orden de selección de paraderos.

- **Generación de servicios cortos para congestión**

Los criterios hasta aquí expuestos para el diseño de servicios no consideran en el análisis la existencia de una capacidad máxima en los vehículos. Esto ignora el potencial que tienen los servicios expresos como medio de aliviar el punto crítico de un corredor, abaratando los costos del operador.

A grandes rasgos, la idea para esta y las próximas heurísticas de generación con congestión consiste en utilizar la información sobre el flujo en el arco crítico en combinación con otras condiciones de optimalidad, y así obtener un indicador de los

ahorros para diferentes configuraciones del servicio. Cabe notar que el foco de estas heurísticas está en encontrar soluciones que reduzcan los costos sociales, y no en resolver ni atenuar el problema de congestión en el corredor.

Para diseñar servicios expresos para un escenario con restricción de capacidad activa, supondremos que tenemos un corredor cuyo arco crítico, a^* tiene una carga de P_{a^*} . Cuando la restricción de capacidad es activa, la frecuencia óptima para el diseño inicial es la mínima que permita acarrear la carga del arco crítico:

$$f_0 = \frac{P_{a^*}}{cap} \quad (3.57)$$

Supongamos que este corredor está siendo atendido por un único servicio regular. Este servicio, dependiendo de cómo está distribuido el perfil de carga a lo largo del corredor, puede llegar a ser muy ineficiente en algunos tramos, debido a que entrega a lo largo de todo el corredor la capacidad necesaria para llevar el total de la carga del tramo crítico. Una forma de aliviar la carga del servicio regular en el arco crítico consistiría en implementar un servicio expreso que tome parte de esta carga y permita disminuir la frecuencia del servicio regular. Existen diferentes formatos de servicio expreso que podrían ayudar a mover la carga del arco crítico. En esta sección y en las siguientes evaluaremos los dos tipos de formato estudiados anteriormente, es decir, servicios cortos y servicios zonales en escenarios congestionados.

Para evaluar el primero de estos formatos, supondremos que un corredor opera con dos servicios, uno regular de frecuencia f_a y costo c_a , y un servicio corto de frecuencia f_c y costo c_c . Al igual que en el caso sin congestión, supondremos que los

servicios se encuentran definidos como ciclos completos, de ida y vuelta, y sus costos deben ser calculados en consecuencia a esto. Definiremos como i_1 y j_1 los paraderos inicial y final del ciclo en el sentido de ida, y como i_2 y j_2 los paraderos inicial y final del sentido de vuelta. Para el análisis consideraremos únicamente los servicios cortos que pasan por el arco crítico (es decir, definiremos como conjunto \mathcal{L}_c al conjunto de los servicios cortos que pasan por el arco crítico).

Al igual que para el caso sin congestión, podemos clasificar a los usuarios de acuerdo a si utilizan sólo el servicio regular o ambos servicios. Al igual que con la demanda total, la carga del perfil del arco crítico se puede descomponer entre usuarios que forman parte de T_A^c , y usuarios que forman parte de T_{AC}^c . Llamaremos estas componentes PM_A^c y PM_{AC}^c , y se pueden calcular de acuerdo a las expresiones (3.58) y (3.59).

$$PM_{AC}^c = \sum_{w \in \mathcal{W}_{AC}^c} \zeta_{a^*}^w T_w \quad (3.58)$$

$$PM_A^c = P_{a^*} - PM_{AC}^c \quad (3.59)$$

En la expresión (3.58) el parámetro $\zeta_{a^*}^w$ toma valor 1 cuando el arco crítico a^* está contenido en el arco w . La carga máxima de cada bus del servicio corto y del servicio regular en el arco crítico se puede calcular como:

$$LM_c^c = \frac{1}{f_c} \left(\frac{f_c}{f_c + f_a} PM_{AC}^c \right) = \frac{PM_{AC}^c}{f_c + f_a} \quad (3.60)$$

$$LM_a^c = \frac{1}{f_a} \left(\frac{f_a}{f_c + f_a} PM_{AC}^c \right) + \frac{PM_A^c}{f_a} = \frac{PM_{AC}^c}{f_c + f_a} + \frac{PM_A^c}{f_a} \quad (3.61)$$

De las ecuaciones anteriores es posible concluir que, si asumimos que el arco crítico de ambos servicios coincide con el arco crítico original del corredor, la carga por bus

del servicio regular siempre va a ser mayor que la del servicio corto, independiente de las frecuencias que se entreguen. Esto implica que al considerar restricción de capacidad de los buses, se deberá imponer solamente que el servicio regular funcione a capacidad, es decir:

$$\frac{PM_{AC}^c}{f_c + f_a} + \frac{PM_A^c}{f_a} = cap \quad (3.62)$$

De lo anterior también se podría concluir que los servicios cortos son siempre ineficientes en este tipo de escenarios, y que nunca operarían a capacidad. La evidencia empírica muestra en todo caso que esto no es necesariamente así: los servicios cortos se suelen aplicar como medida para aumentar la oferta en tramos de mayor demanda de un corredor, cosa que sucede, por ejemplo, en el Metro de Santiago. Esta diferencia se explica por el hecho de que el modelo de comportamiento que se utiliza en este trabajo asume que el diseño siempre entrega capacidad suficiente para que todos los usuarios puedan utilizar la opción que desean, o, en otras palabras, que los usuarios nunca dejan pasar un bus lleno para esperar el siguiente bus con capacidad disponible en pasar, cosa que sí sucede en la práctica. De todas formas, estas limitaciones del modelo no son críticas ya que no impiden que la heurística desarrollada permita encontrar combinaciones atractivas de servicios, que es su finalidad última.

Dado que los tiempos de viaje son los mismos en ambos servicios para los pares que atienden, podemos obviar los costos por tiempo de viaje de los costos sociales al momento de optimizar las frecuencias. Con esto, la función de costos sociales (modificada) de este problema será la siguiente:

$$SC = f_a c_a + f_c c_c + \frac{\lambda \theta_{wt} T_A^c}{f_a} + \frac{\lambda \theta_{wt} T_{AC}^c}{f_a + f_c} \quad (3.63)$$

En la función anterior los tiempos de espera son calculados tomando en cuenta que para los viajes contenidos por el servicio corto los usuarios utilizarán indistintamente ambos servicios disponibles. Construimos el lagrangiano del problema correspondiente a minimizar la función (3.63) sujeta a la restricción impuesta por (3.62):

$$\mathfrak{L} = f_a c_a + f_c c_c + \frac{\lambda \theta_{wt} T_A^c}{f_a} + \frac{\lambda \theta_{wt} T_{AC}^c}{f_a + f_c} + \alpha \left(\frac{PM_{AC}^c}{f_c + f_a} + \frac{PM_A^c}{f_a} - cap \right) \quad (3.64)$$

Derivando respecto a f_a , f_c y α e igualando a cero obtenemos las condiciones de primer orden del problema:

$$\frac{\delta \mathfrak{L}}{\delta f_a} = c_a - \frac{\lambda \theta_{wt} T_A^c}{f_a^2} - \frac{\lambda \theta_{wt} T_{AC}^c}{(f_a + f_c)^2} - \alpha \frac{PM_{AC}^c}{(f_c + f_a)^2} - \alpha \frac{PM_A^c}{f_a^2} = 0 \quad (3.65)$$

$$\frac{\delta \mathfrak{L}}{\delta f_c} = c_c - \frac{\lambda \theta_{wt} T_{AC}^c}{(f_a + f_c)^2} - \alpha \frac{PM_{AC}^c}{(f_c + f_a)^2} = 0 \quad (3.66)$$

$$\frac{\delta \mathfrak{L}}{\delta \alpha} = \frac{PM_{AC}^c}{f_c + f_a} + \frac{PM_A^c}{f_a} - cap = 0 \quad (3.67)$$

Restando las expresiones (3.65) y (3.66) y despejando se obtiene que:

$$f_a = \sqrt{\frac{\lambda \theta_{tw} T_A^c + \alpha PM_A^c}{c_a - c_c}} \quad (3.68)$$

Y de (3.66) se obtiene:

$$f_a + f_c = \sqrt{\frac{\lambda \theta_{tw} T_{AC}^c + \alpha PM_{AC}^c}{c_c}} \quad (3.69)$$

Reemplazando (3.68) y (3.69) en (3.67) se obtiene que:

$$PM_{AC}^c \sqrt{\frac{c_c}{\lambda\theta_{tw}T_{AC}^c + \alpha PM_{AC}^c}} + PM_A^c \sqrt{\frac{c_a - c_c}{\lambda\theta_{tw}T_A^c + \alpha PM_A^c}} = cap \quad (3.70)$$

Esta última ecuación puede ser resuelta numéricamente para la variable α . Se observa que la función a mano izquierda de la igualdad es decreciente con α , y que tiende a cero cuando α tiende a infinito. Cabe notar también que por interpretación de α como multiplicador de Lagrange, podemos saber que para que la solución tenga sentido debemos exigir que $\alpha > 0$ ya que de ser negativo implicaría que un aumento en la capacidad de los vehículos (sin cambiar sus costos de operación) resultaría en un aumento de los costos sociales, lo que no tendría sentido. Esta última condición nos permite descartar algunos servicios a partir de la fórmula (3.70) evaluada en $\alpha = 0$, es decir, podemos descartar para el análisis todo servicio corto que no cumpla que:

$$Ind5_c = P_{a^*}^{AC} \sqrt{\frac{c_c}{T_c \lambda \theta_{tw}}} + P_{a^*}^A \sqrt{\frac{c_a - c_c}{T_a \lambda \theta_{tw}}} - cap \geq 0 \quad (3.71)$$

Si queremos conocer cuál es el servicio óptimo que entrega más beneficios al operar en conjunto con el servicio regular, debemos en primer lugar calcular numéricamente el valor de α^* a partir de la ecuación (3.70) para cada servicio que cumpla la condición (3.71). El hecho de que el lado izquierdo (3.70) sea decreciente en α y que tienda a cero en infinito asegura la existencia de solución en este paso, además de facilitar su búsqueda por algoritmos sencillos como por ejemplo el método de bisección. Conocido α^* se pueden obtener las frecuencias óptimas de acuerdo a las fórmulas (3.68) y (3.69), para finalmente evaluar los costos sociales usando la fórmula (3.63). Esto hace posible obtener para cada servicio corto un estimador

comparativo de su costo social óptimo, lo que permite escoger entre todos ellos el que aparece como más beneficioso.

Dado que la heurística para generar servicios cortos en escenarios congestionados posee elementos de generación de servicios y de ajuste de capacidad, se ha optado por clasificarla en una nueva categoría de heurísticas.

Heurística para generación con capacidad HGC_1 :

1. Para cada servicio corto $c \in \mathcal{L}_c$ con $\text{Ind}5_c \geq 0$ determinar numéricamente el valor de α^* a partir de la ecuación (3.70).
2. Para cada servicio corto $c \in \mathcal{L}_c$ reemplazar α^* en (3.68) y (3.69) para encontrar el valor de f_a^* y f_c^* . Reemplazar estos valores en (3.63) para obtener el costo social de cada alternativa.
3. Escoger el servicio corto que minimice el costo social y agregarlo a \mathcal{L} en caso de que todavía no se encuentre en este conjunto.
4. En el sentido crítico del corredor, fijar la frecuencia del servicio corto en f_c^* , la del regular en f_a^* , y todas las demás frecuencias de servicios en este sentido en 0. En el sentido contrario del corredor dejar libres todas las frecuencias.
5. Optimizar las frecuencias del problema.
6. Aplicar la heurística para capacidad HC_2 descrita en el Capítulo 2.

Este algoritmo saca partido de la separación de sentidos del problema, utilizando la solución encontrada para el sentido crítico del corredor y optimizando el sentido no crítico ignorando la restricción de capacidad. Es posible que al resolver el paso 5

algún servicio del sentido menos cargado se vea sobrepasado en capacidad. A pesar de que es poco probable que esto suceda, se agrega en el paso 6 un ajuste de capacidad para prevenir esta situación.

- **Generación de servicios zonales para congestión**

El segundo formato de servicio expreso a evaluar en esta sección será el servicio zonal, que, al igual que en el caso sin congestión, pareciera ser una alternativa de interés para el análisis.

Si suponemos que este servicio opera en conjunto con un único servicio regular y que los buses son todos de la misma capacidad, la suma de frecuencias que atraviesan el arco crítico debe ser al menos f_0 para que exista capacidad suficiente. Por otra parte, dado que por definición en un escenario congestionado las frecuencias óptimas son las determinadas por la restricción de capacidad, podemos asumir que en general (salvo casos donde el nivel de congestión es muy bajo) esta desigualdad es activa. De acuerdo a esto, si designamos como f_e a la frecuencia a destinar para el servicio zonal, y como f_a la nueva frecuencia del servicio regular que lo acompaña, tenemos que:

$$f_a + f_e = f_0 \quad (3.72)$$

La implementación de un servicio expreso se puede entender como una reducción en los costos de operación promedio de los servicios del sistema, que a su vez, de acuerdo con la fórmula (3.10), puede traducirse en un aumento en la frecuencia total óptima del sistema, lo que en algunos casos podría hacer que la frecuencia total óptima supere la estrictamente necesaria por capacidad. Sin embargo, para poder

proseguir con el análisis que lleve a una solución factible para el problema y que de todas formas representa una disminución en los costos del sistema, asumiremos que esta igualdad se cumple en forma estricta.

Podemos conocer, para determinado servicio zonal e , de qué forma la carga máxima en el arco crítico P_{a^*} se descompone entre usuarios que sólo utilizan el expreso, usuarios que sólo usan el servicio regular, y usuarios que utilizan cualquiera de los dos servicios. Llamaremos a estos valores respectivamente PM_E^e , PM_A^e , y PM_{AE}^e , y se pueden obtener a partir de las siguientes fórmulas que utilizan los mismos conjuntos definidos cuando se analizó este tipo de servicios en un corredor sin cogestión:

$$PM_E^e = \sum_{w \in \mathcal{W}_E^e} \zeta_{a^*}^w T_w \quad (3.73)$$

$$PM_{AE}^e = \sum_{w \in \mathcal{W}_{AE}^e} \zeta_{a^*}^w T_w \quad (3.74)$$

$$PM_A^e = \sum_{w \in \mathcal{W}_A^e} \zeta_{a^*}^w T_w \quad (3.75)$$

A partir de estos parámetros, que dependen de la configuración del servicio zonal, es posible determinar la carga en el arco máximo en los buses de cada uno de los servicios.

$$LM_a^e = \frac{PM_A^e}{f_a} + \frac{PM_{AE}^e}{f_0} \quad (3.76)$$

$$LM_e^e = \frac{PM_E^e}{f_e} + \frac{PM_{AE}^e}{f_0} \quad (3.77)$$

En el óptimo, cada uno de estos servicios debe operar a máxima capacidad de forma de logren llevar toda la carga del arco crítico. Haciendo $LM_a^e = cap$ y $LM_e^e = cap$ y

utilizando la ecuación (3.57) se obtienen las siguientes expresiones para las frecuencias óptimas con capacidad:

$$f_a^* = \frac{f_0 PM_A^e}{PM_A^e + PM_E^e} \quad (3.78)$$

$$f_e^* = \frac{f_0 PM_E^e}{PM_A^e + PM_E^e} \quad (3.79)$$

Para que esta solución tenga validez, es necesario confirmar que efectivamente los usuarios considerados como parte de \mathcal{W}_E^e tengan el incentivo para esperar al servicio expreso en forma exclusiva. De acuerdo la fórmula (2.30) (Lema 3) se desprende que esto sucederá cuando:

$$f_e^* \geq \frac{\theta_{wt}\lambda}{\theta_{tt}Nt_{stop}} \quad (3.80)$$

En caso de no cumplirse esta última condición, los usuarios de los viajes largos (conjunto \mathcal{W}_E^e) se repartirán entre el servicio zonal y el servicio regular, invalidando los supuestos hechos para obtener el indicador. Debido a esto, un servicio donde esta última condición no se cumple debe ser excluido del análisis.

Analicemos el efecto sobre los costos sociales de pasar de la situación inicial a la con un servicio zonal. El ahorro en costos de operación es el que se muestra a continuación:

$$\Delta OC = f_0 c_a - (f_a c_a + f_e c_e) \quad (3.81)$$

Los parámetros c_a y c_e corresponden a los costos de operación por unidad de frecuencia de cada servicio. Aplicando la fórmula (2.3) para los costos de operación, y dado que por construcción se debe satisfacer (3.72) y $lc_a = lc_e$, podemos reducir la expresión a:

$$\Delta OC = f_e c_T (tc_a - tc_e) \quad (3.82)$$

Si N_e es el número de paraderos que el servicio zonal omite en su trazado, podemos reescribir esta última expresión como:

$$\Delta OC = f_e c_T N_e \tau \quad (3.83)$$

Con respecto a los ahorros en tiempo de viaje, el análisis del caso sin restricción de capacidad no pierde validez, y responden a la fórmula (3.37).

Los pasajeros que perciben un cambio en su tiempo de espera son los que no son atendidos por el servicio expreso, y los que desean usar el expreso para ir del comienzo al final del corredor. Los pasajeros que realizan viajes cortos en el comienzo o el final del corredor no ven alterados sus tiempos de espera. Entonces, la variación de costos por espera corresponderá a:

$$\Delta TWC = \theta_{wt} T_A^e \left(\frac{\lambda}{f_0} - \frac{\lambda}{f_a} \right) + \theta_{wt} T_E^e \left(\frac{\lambda}{f_0} - \frac{\lambda}{f_e} \right) \quad (3.84)$$

Es fácil ver que esta expresión es negativa (basta con notar que $f_0 > f_a$ y $f_0 > f_e$). Esto implica que los tiempos de espera del sistema necesariamente aumentan al destinar parte de la frecuencia del servicio regular a un servicio expreso.

Sumando los tres términos obtenemos el ahorro total para el sistema producto de la incorporación del servicio expreso:

$$S_e = f_e c_T N_e \tau + \theta_{tt} T_E^e N_e \tau + \theta_{wt} T_A^e \left(\frac{\lambda}{f_0} - \frac{\lambda}{f_a} \right) + \theta_{wt} T_E^e \left(\frac{\lambda}{f_0} - \frac{\lambda}{f_e} \right) \quad (3.85)$$

Evaluando esta última función en f_a^* y f_e^* (y cuidando de que se cumpla la condición (3.80)) podemos conocer el beneficio máximo que se puede obtener con un servicio expreso del formato impuesto. Esto implica que podemos hacer una lista de los

servicios expresos que posean esta forma y escoger entre ellos el que presente un mejor indicador para obtener así una mejora en los costos sociales del sistema.

Definiendo como \mathcal{L}_E al conjunto de servicios zonales factibles para un corredor, la heurística para ajustar la capacidad utilizando un servicio zonal es la que sigue:

Heurística para generación con capacidad HGC₂:

1. Para cada servicio zonal $e \in \mathcal{L}_E$ calcular el valor de f_a^* y f_e^* a partir de las fórmulas (3.78) y (3.79). Reemplazar estos valores en (3.85) para obtener el ahorro que proporciona cada alternativa.
2. Escoger el servicio zonal que maximice el ahorro y agregarlo a \mathcal{L} en caso de que todavía no se encuentre en este conjunto.
3. En el sentido crítico del corredor, fijar la frecuencia del sentido zonal en f_e^* , la del regular en f_a^* , y todas las demás frecuencias de servicios en este sentido en 0. En el sentido contrario del corredor dejar libres todas las frecuencias.
4. Optimizar las frecuencias del problema.
5. Aplicar la heurística para capacidad HC₂.

Este algoritmo funciona en forma similar al de los servicios cortos para capacidad: se diseña un servicio para acompañar al servicio regular en el sentido más cargado, y se deja libre para optimizar el sentido menos cargado. Es posible, aunque poco probable, que alguna línea del sentido menos cargado viole alguna restricción de capacidad al ser optimizado en forma irrestricta, es por esto que al final del algoritmo se hace preventivamente ajuste de capacidad.

- **Generación de servicios combinados para congestión**

Hasta este punto hemos propuesto métodos analíticos para optimizar dos formatos de servicio para apoyar a un servicio regular en el sentido crítico de un corredor. Resulta natural entonces preguntarse por la posibilidad de combinar estos dos tipos de servicio para mejorar la solución obtenida. Un posible enfoque para resolver este problema consistiría en formular el problema de optimización para tres servicios simultáneos e intentar obtener una expresión para las frecuencias óptimas, de manera de poder escoger la combinación de servicios que entregue un mayor beneficio. Este problema es considerablemente más complejo que los problemas estudiados en las secciones anteriores. Además de corresponder a un problema en tres variables en lugar de dos, el comportamiento de los usuarios se hace menos predecible debido a distintas combinaciones de servicios disponibles que se generan, y la función objetivo ya deja de presentar una componente constante (recordemos que en el caso de los servicios cortos los tiempos de viaje no dependen de las frecuencias, y que con los servicios zonales las esperas eran constantes para el análisis). Sin embargo, haciendo algunas simplificaciones es posible encontrar una buena solución utilizando una combinación de las técnicas descritas.

Gran parte de la dificultad de este problema se ve reducida si suponemos que el arco crítico del corredor debe estar contenido en el tramo omitido por el servicio zonal, además del supuesto antes hecho de que el servicio corto debe contener al arco crítico. Con esto, podemos asegurar que la carga del arco crítico se puede clasificar dentro de tres grupos de comportamiento: viajes que utilizan sólo el servicio expreso, PM_E^{ec} , viajes que utilizan el servicio corto o el regular PM_{AC}^{ec} , y viajes que sólo

utilizan en servicio regular, PM_A^{ec} . Esto implica que la carga máxima del servicio expreso será de:

$$LM_e^{ec} = \frac{PM_E^{ec}}{f_e} \quad (3.86)$$

Si imponemos que este servicio opere a capacidad, obtenemos que la frecuencia del servicio expreso se puede calcular como:

$$f_e = \frac{PM_E^{ec}}{cap} \quad (3.87)$$

Podrían existir dos tipos de usuarios que utilizarán el servicio zonal: los que lo utilizan de extremo a extremo pasando por el arco crítico (que antes fueron referidos como conjunto \mathcal{W}_E^e , que ahora llamaremos \mathcal{W}_E^{ec}), y los que lo utilizan tal como si fuese un servicio regular (que antes correspondían a \mathcal{W}_{AE}^e y que ahora podrían pertenecer a \mathcal{W}_{AE}^{ec} o a \mathcal{W}_{ACE}^{ec}). Asumiremos que no existen usuarios del segundo tipo (lo cual constituye una subestimación de los beneficios de los usuarios). De esta forma, podemos clasificar a los usuarios dentro de tres grupos: \mathcal{W}_A^{ec} , \mathcal{W}_{AC}^{ec} y \mathcal{W}_E^{ec} . Estos tres conjuntos, y un esquema de los formatos de servicio a utilizar, se presentan a continuación en la siguiente figura.

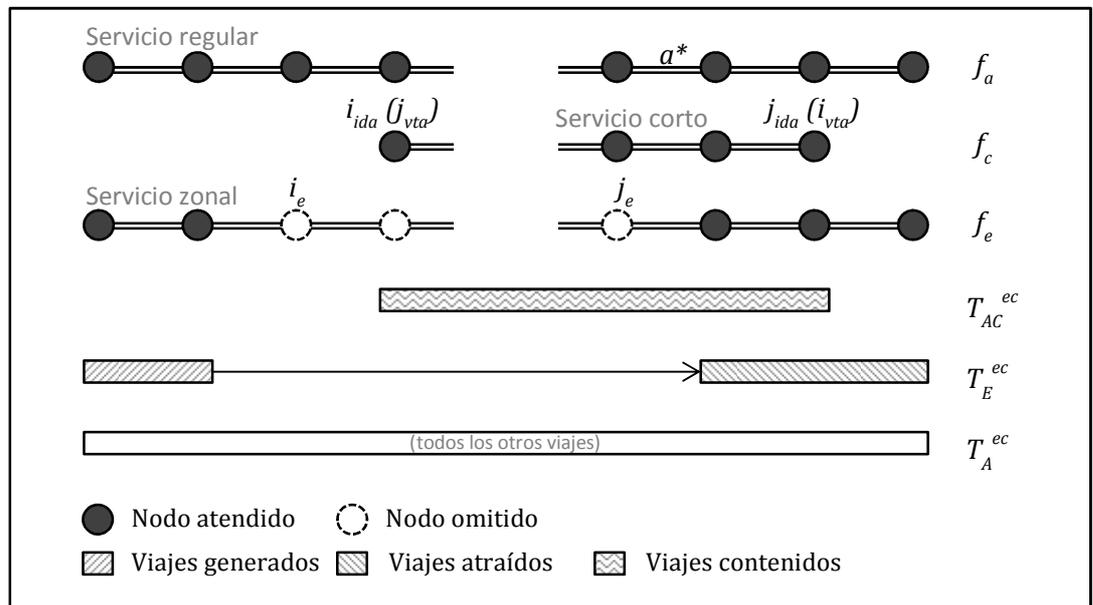


Figura 3-5: Generación de servicios combinados para congestión

Si excluimos del problema los viajes ya atendidos por el servicio expreso ($T_w | w \in \mathcal{W}_E^{ec}$), obteniendo un problema equivalente al de generación de un servicio corto para congestión, y que se puede resolver aplicando HGC_1 . Llamaremos T'_w a la matriz que resulta de eliminar los viajes capturados por el expreso, es decir:

$$T'_w = \begin{cases} 0, & \text{si } w \in \mathcal{W}_E^{ec} \\ T_w, & \text{e. o. c.} \end{cases} \quad (3.88)$$

Esto significa que, para cada servicio zonal y servicio corto que cumplan las condiciones impuestas, es posible encontrar las frecuencias óptimas del problema. Esto implica que una posible aproximación al problema consistiría en comparar, para cada combinación posible, los costos sociales del sistema y escoger la solución que posea el menor costo social. Dado a que se está asumiendo en la formulación del problema que en el óptimo todos los servicios operan, se debe tener cuidado de descartar soluciones no factibles (es decir, que posean alguna frecuencia negativa).

De todas formas, si el óptimo del problema no fuera operar con estos tres servicios, sino con sólo dos de ellos, este óptimo sería encontrado por las heurísticas HGC_1 y HGC_2 antes descritas.

Finalmente, dado que la heurística de generación de servicios cortos consideran servicios de ida y vuelta, mientras que la de servicio zonal trabaja con solo ida, es necesario, para poder evaluar los costos sociales, definir con qué criterio se cuantificarán los costos de operación en el sentido no crítico del corredor para el servicio zonal. Una forma sencilla de resolver este problema consiste en simplemente suponer que el servicio zonal opera como un servicio regular en el sentido de vuelta al momento de obtener c_e . Con esto, la fórmula para evaluar los costos sociales es la que se presenta a continuación:

$$SC = (f_a c_a + f_c c_c + f_e c_e) + \lambda \theta_{wt} \left(\frac{T_A^{ec}}{f_a} + \frac{T_{AC}^{ec}}{f_a + f_c} + \frac{T_E^{ec}}{f_a} \right) - \theta_{tt} T_E^e N_e \tau \quad (3.89)$$

De acuerdo a lo dicho, un posible algoritmo para diseñar un servicio zonal más un servicio corto que acompañen a un servicio regular en el sentido más cargado del corredor se presenta a continuación:

Heurística para generación con capacidad $HGC_{3\alpha}$:

1. Para cada servicio zonal $e \in \mathcal{L}_E$ que contenga al arco crítico en su tramo expreso, calcular el valor de f_e^* a partir de la fórmula (3.87).
2. Construir la matriz T'_w de acuerdo a la expresión (3.88). Aplicar la heurística HGC_1 utilizando esta matriz para determinar el servicio corto óptimo $c(e)$

asociado al zonal e , y las frecuencias óptimas f_a^* y f_c^* correspondientes a esta combinación de servicios.

3. Evaluar la expresión (3.89) para el costo social para cada combinación de servicios e y $c(e)$, y escoger la combinación que minimice este costo.
4. En el sentido crítico del corredor, fijar la frecuencia del servicio zonal en f_e^* , la del regular en f_a^* , y la del servicio corto en f_c^* . Fijar todas las demás frecuencias de servicios en este sentido en 0. En el sentido contrario del corredor dejar libres todas las frecuencias.
5. Optimizar las frecuencias del problema.
6. Aplicar la heurística para capacidad HC_2 .

En el paso 6 se puede usar alternativamente la heurística HC_3 en lugar de HC_2 . La nueva heurística que se obtiene la denotaremos como HC_7 .

Heurística para generación con capacidad HGC_{3b} :

1. Aplicar los pasos 1 a 5 de HGC_{3a} .
2. Aplicar la heurística para capacidad HC_3 .

Estos métodos acá propuestos desprecian el efecto de los viajes que el servicio expreso puede quitar al servicio regular o al corto en los tramos de operación regular. Esta situación puede conllevar que las frecuencias del sentido crítico no sean las óptimas, pero permite asegurar que en términos de capacidad este resultado se encuentra por el lado conservador, de forma que no existe riesgo de generar

soluciones infactibles. Además, el paso final se encargaría de cualquier infactibilidad en caso de surgir alguna.

Hasta ahora hemos descrito cuatro métodos que permiten diseñar servicios destinados a aliviar la carga del arco crítico del corredor al operar en conjunto con un servicio regular. Todos ellos son relativamente rígidos con respecto a los servicios a utilizar en el sentido más cargado, ya que la existencia de servicios adicionales puede violar los supuestos asumidos al momento de predecir la carga de cada servicio. Sin embargo, para el caso del servicio zonal es posible hacer una pequeña modificación para abordar este problema. Todo se basa en el hecho de que un servicio que no contiene la parte omitida del zonal no puede tener efecto en la carga de dicho servicio. Esto implica que otra forma de considerar servicios cortos consiste en resolver el problema como ya se vio en generación de servicios expresos, pero fijando en cero las frecuencias sólo para los servicios que entran en conflicto con el servicio zonal. Diremos que un servicio entra en conflicto con el zonal si su nodo inicial es anterior a i_e (el primer nodo omitido del zonal) y su nodo final es posterior a j_e .

Heurística para generación con capacidad HGC_{4a} :

1. Para cada servicio zonal $e \in \mathcal{L}_E$ calcular el valor de f_a^* y f_e^* a partir de las fórmulas (3.78) y (3.79). Reemplazar estos valores en (3.85) para obtener el ahorro que proporciona cada alternativa.
2. Escoger el servicio zonal que maximice el ahorro y agregarlo a \mathcal{L} en caso de que todavía no se encuentre en este conjunto.

3. En el sentido crítico del corredor, fijar la frecuencia del sentido zonal en f_e^* , y la de los servicios que entran en conflicto con el expreso (salvo el servicio regular) en 0. En el sentido contrario del corredor dejar libres todas las frecuencias.
4. Optimizar las frecuencias del problema.
5. Aplicar la heurística para capacidad HC_2 .

En esta heurística el paso 5 ya no sólo es una medida preventiva para el caso de que el sentido menos cargado presente alguna violación de capacidad: ahora, dado que el único servicio con frecuencia fija y con cumplimiento de capacidad asegurada es el zonal, este paso corrige las frecuencias del servicio regular y los servicios cortos para ajustarlas también a la restricción de capacidad. Al igual que con HGC_{3a} y HGC_{3b} definiremos la versión alternativa de este algoritmo como la que aplica el algoritmo HC_3 en lugar de HC_2 para ajustar la capacidad:

Heurística para generación con capacidad HGC_{4b} :

1. Aplicar los pasos 1 a 4 de HGC_{4a} .
2. Aplicar la heurística para capacidad HC_3 .

3.2 Evaluación de las heurísticas

Los métodos de generación de líneas, en conjunto con el modelo de optimización de frecuencias, permiten resolver el problema completo de diseño de servicios expresos

para un corredor. En la sección anterior se presentan varias heurísticas de generación, que en principio podrían ser utilizadas todas en forma simultánea, en conjunto con el modelo de optimización de frecuencias, para construir una heurística combinada. Sin embargo, el esfuerzo de ejecutar todas las heurísticas en conjunto puede ser injustificado, dado que es posible que algunas de ellas tiendan a entregar resultados similares, o a que existan heurísticas que dominen a otras.

Con el fin de determinar el verdadero aporte de cada heurística de generación propuesta, se diseñaron doce escenarios con diferentes tipos de perfil de carga y niveles de congestión (es decir, escenarios donde la restricción de capacidad es activa y donde no lo es), y a partir de los resultados obtenidos en cada uno de ellos se evaluó el aporte de las distintas heurísticas tanto en términos de su capacidad de descubrir nuevos servicios que generan ahorros en los costos sociales, como en la magnitud de dichos ahorros.

En la Tabla 3-1 a continuación se presenta un resumen con las 5 heurísticas para la generación de servicios presentadas en este trabajo.

Tabla 3-1: Heurísticas para la generación de servicios sin considerar capacidad

Heurística	Descripción	Formato Impuesto	Líneas Existentes
<i>HG_{1a}</i>	Generación de servicios iniciales	No	No
<i>HG_{1b}</i>	Generación de servicios iniciales	No	No
<i>HG₂</i>	Generación de servicios cortos	Sí	No
<i>HG₃</i>	Generación de servicios zonales	Sí	No
<i>HG₄</i>	Generación de servicios adicionales	No	Sí

En la tabla anterior se denomina como generación de servicios iniciales a las heurísticas que en la sección anterior fueron llamadas de generación de servicios expresos iniciales. Al igual que los servicios adicionales generados por HG_4 , estos servicios no poseen una estructura o formato impuesto, a diferencia de los servicios construidos por HG_2 y HG_3 . Como se mencionó en la descripción de las heurísticas, la diferencia entre HG_{1a} y HG_{1b} es que la segunda heurística utiliza una versión más compleja del indicador por lo que su cálculo es un poco más complicado. Las heurísticas desde HG_{1a} hasta HG_3 construyen servicios que son independientes de los ya existentes en el conjunto \mathcal{L} . La heurística HG_4 , por su parte, calcula el efecto de los servicios y cambios a evaluar en función de la asignación, servicios, y frecuencias existentes. Esta heurística es, de las cinco acá presentadas, la que consume más tiempo, debido a que cada vez que encuentra un nuevo servicio beneficioso comienza su ejecución desde el comienzo, y a que cada evaluación de un nuevo servicio implica la resolución del problema de optimización de frecuencias. A continuación se presenta una tabla resumen de las heurísticas para el problema con capacidad limitada descritas en este trabajo.

Tabla 3-2: Heurísticas para ajuste de capacidad

Heurística	Descripción
HC_1	Ajuste de capacidad por paso fijo
HC_2	Ajuste de capacidad por paso por perfil máximo
HC_3	Priorización de servicios expresos

La primera de estas heurísticas corresponde a la presentada en Leiva et al. (2010), mientras que las otras dos corresponden a mejoras propuestas en esta tesis para dicha heurística. Un último tipo de heurísticas corresponde a las que sugieren nuevos servicios para escenarios en que la capacidad está activa en el arco crítico. En este trabajo se proponen cuatro heurísticas de este tipo, que se muestran en la Tabla 3-3:

Tabla 3-3: Heurísticas de generación con restricción de capacidad activa

Heurística	Descripción
<i>HGC₁</i>	Servicio corto para capacidad
<i>HGC₂</i>	Servicio zonal para capacidad
<i>HGC_{3a}</i>	Servicio corto más servicio expreso
<i>HGC_{3b}</i>	Servicio corto más servicio expreso

Para poner a prueba las heurísticas propuestas se diseñaron doce escenarios distintos. Estos escenarios están basados en las matrices de demanda de dos corredores de la ciudad de Santiago. El primer corredor corresponde al corredor de Avenida Pajaritos en horario punta mañana; denominaremos P1 a esta matriz. Este corredor posee diez paradas por sentido y un perfil de carga creciente. Esta es la misma matriz de viajes con la que se trabaja en Leiva et al. (2010). La otra matriz de demanda corresponde a la operación en horario punta mañana del corredor de Avenida Grecia, que llamaremos G1. El perfil de carga de este corredor posee forma de cerro, con su máximo ocurriendo cerca de la mitad del corredor. La demanda de esta matriz fue agregada para quedar con diez paraderos por sentido.

A partir de los dos escenarios base se generaron dos otros escenarios haciendo algunas alteraciones puntuales en las matrices de demanda. La idea detrás de esto era

generar singularidades que podrían justificar el diseño de servicios expresos especiales, y así probar si el modelo es capaz de identificar estas oportunidades y aprovecharlas. Llamaremos a estos otros dos perfiles P2 y G2. Las matrices de viajes, perfiles de carga, y otros indicadores de interés de cada uno de estos cuatro casos base se presentan en las figuras 3-6 a 3-9. En cada una de estas figuras se presentan ambos sentidos del corredor (de 1 a 10 y de 10 a 1).

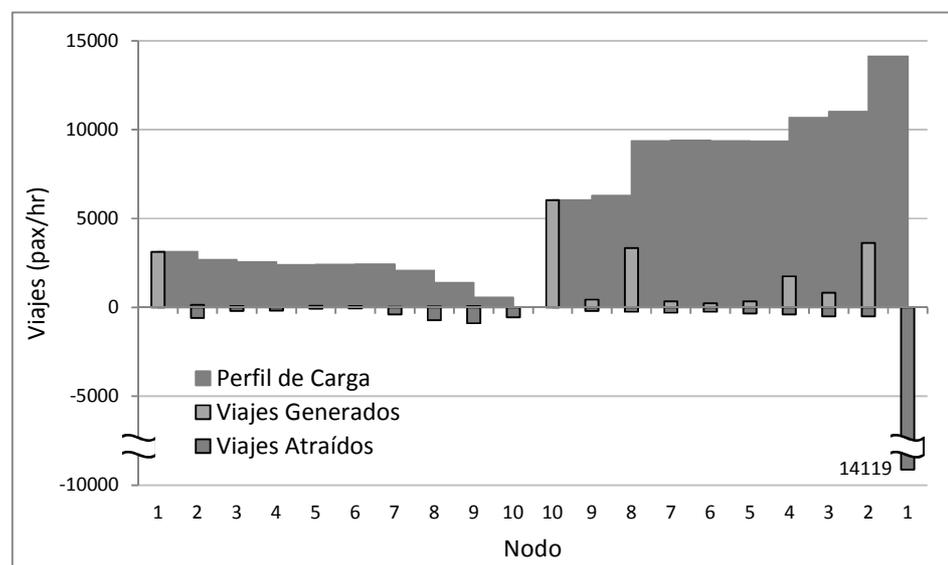


Figura 3-6: Perfil de carga matriz P1

La primera matriz corresponde a un corredor de perfil creciente en su sentido crítico. Una gran parte de los viajes de esta matriz ocurren entre sus nodos extremos, en especial su nodo final, como sucede en corredores que alimentan otro tipo de servicios, como por ejemplo el metro.

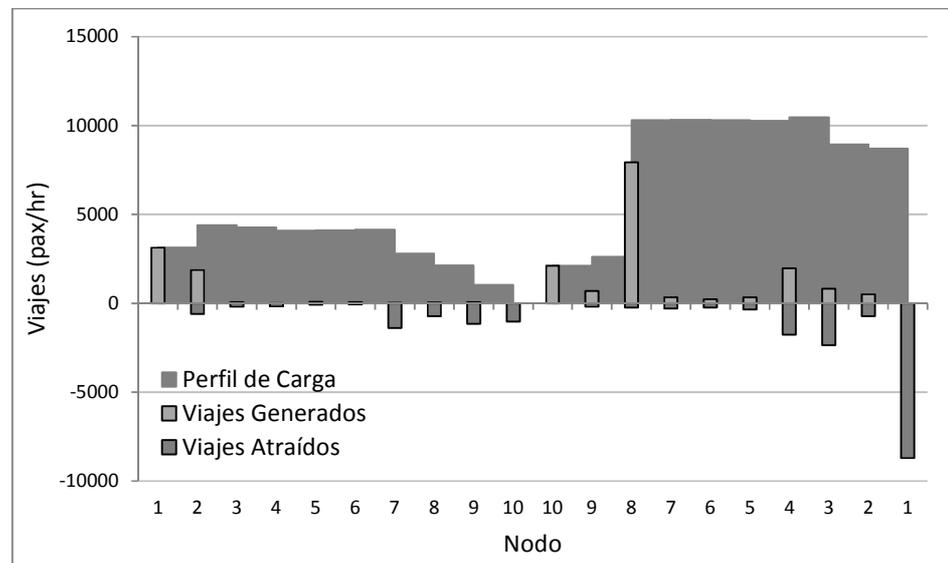


Figura 3-7: Perfil de carga matriz P2

La segunda matriz a considerar posee un perfil similar al de la primera, ya que fue construida a partir de ésta realizando algunas modificaciones locales, que consistieron principalmente en eliminar viajes en los pares 10 - 1 (para hacer menos evidente el trazado de los servicios óptimos) y 2 - 1 (para eliminar un exceso de viajes cortos, como se puede apreciar en el perfil de la matriz P1), y generar viajes originados en los paraderos 2 hacia 10, y 8 hacia 1, repartidos entre los posibles destinos, para que la matriz presentara más actividad en sus tramos intermedios.

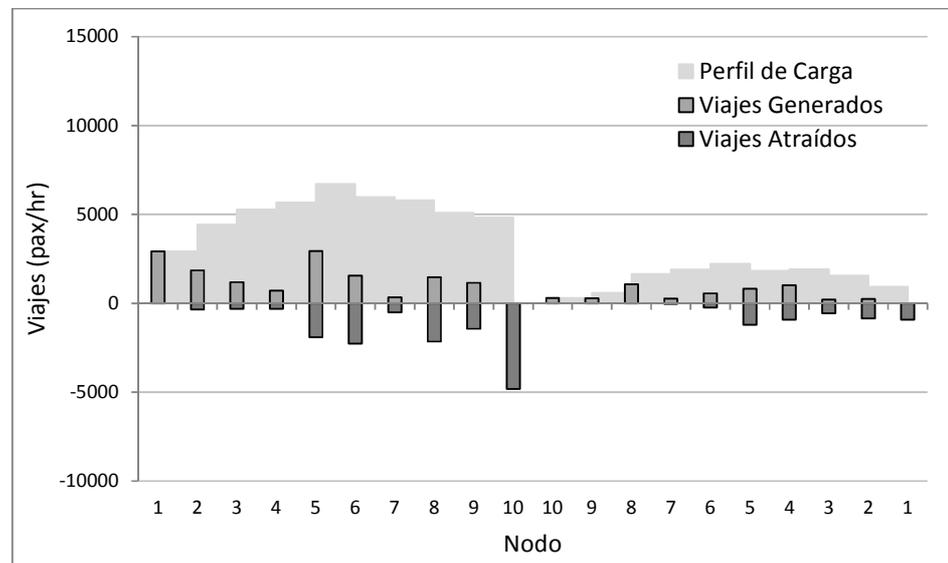


Figura 3-8: Perfil de carga matriz G1

El perfil de demanda de G1 corresponde al de un corredor con su mayor carga concentrada en el centro, y con sus viajes distribuidos en toda su extensión. La concentración de orígenes y destinos presenta las irregularidades propias de un corredor real.

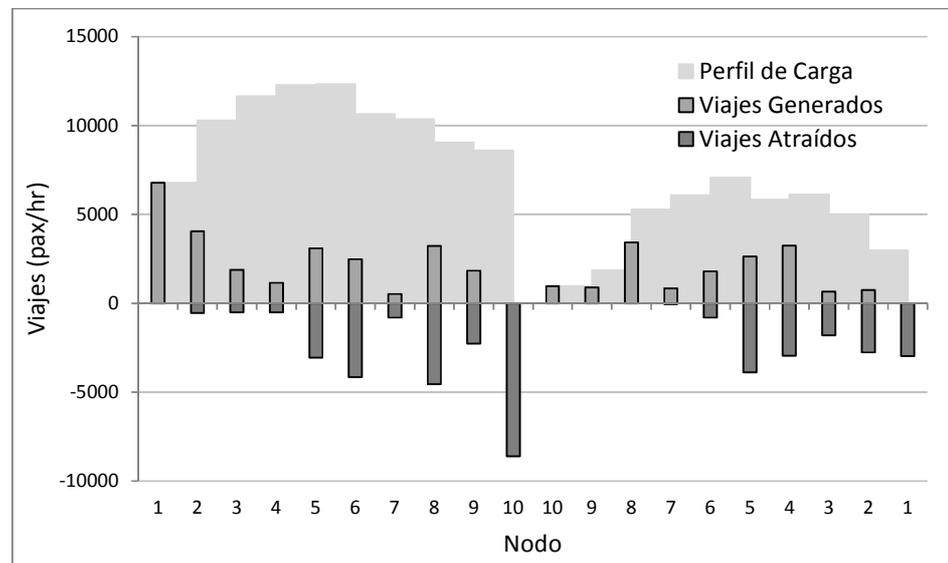


Figura 3-9: Perfil de carga matriz G2

El perfil de carga G2 fue construido a partir de G1 amplificando la matriz en un 60% en su sentido más cargado, y en un 220% en su sentido menos congestionado, y modificando algunos pares específicos.

Para cada uno de estos escenarios se definió adicionalmente tres niveles de capacidad para los buses: irrestricta, activa, y ajustada. Pruebas preliminares del modelo mostraron que valores de 200, 120 y 80 *pax/veh* cumplían las condiciones deseadas para cada caso.

Los otros parámetros de los escenarios utilizados para el experimento se presentan en la tabla que sigue:

Tabla 3-4: Parámetros de los escenarios

Parámetro	Valor	Unidades
λ	1	
θ_{tr}	75	\$
θ_{wt}	30	\$/min
θ_{tt}	15	\$/min
Velocidad	25	km/hr
Distancia entre paradas	0.8	Km
τ	1	min
c_L	500	\$/bus-Km
c_T	5.300	\$/bus-hr

Para el cálculo de tiempos de viajes y costos de operación de los servicios en este experimento se utilizaron las fórmulas (2.1) a (2.5), asumiendo que los tiempos y distancias asociados a maniobras de retorno de los buses son despreciables.

Con esto, podemos resumir los 12 escenarios como se muestra en la tabla a continuación:

Tabla 3-5: Escenarios para el experimento

Escenario	Matriz Base	Capacidad
P1A	Pajaritos I	Irrestringida
P1M	Pajaritos I	Activa
P1B	Pajaritos I	Ajustada
P2A	Pajaritos II	Irrestringida
P2M	Pajaritos II	Activa
P2B	Pajaritos II	Ajustada
G1A	Grecia I	Irrestringida
G1M	Grecia I	Activa
G1B	Grecia I	Ajustada
G2A	Grecia II	Irrestringida
G2M	Grecia II	Activa
G2B	Grecia II	Ajustada

El experimento a realizar para cada escenario consistió en aplicar en cada uno de ellos diferentes estrategias para la generación de servicios y para el ajuste para capacidad. Estas estrategias fueron agrupadas en etapas de solución, y se indican a continuación en la Tabla 3-6.

Tabla 3-6: Etapas de solución

Etapa	Descripción	Heurísticas	Cap.
St_1	Regular sin capacidad	-	No
St_2	Regular con capacidad	-	Sí
St_3	Líneas iniciales, versión simple	HG_{1a}, HG_2	No
St_4	Líneas iniciales, versión compleja	HG_{1b}, HG_2	No
St_5	Líneas iniciales, versión combinada	HG_{1a}, HG_{1b}, HG_2	No
St_6	Líneas iniciales más zonales	$HG_{1a}, HG_{1b}, HG_2, HG_3$	No
St_7	Líneas adicionales.	$HG_{1a}, HG_{1b}, HG_2, HG_3, HG_4$	No
St_8	Óptimo social con capacidad	HG_{1a-4}	Sí*
St_9	Ajuste de capacidad por paso fijo	HG_{1a-4}, HC_1	Sí
St_{10}	Ajuste de capacidad por paso por perfil máximo	HG_{1a-4}, HC_2	Sí
St_{11}	Priorización de servicios expresos	HG_{1a-4}, HC_3	Sí
St_{12}	Servicio corto para capacidad	HG_{1a-4}, HGC_1	Sí
St_{13}	Servicio zonal para capacidad	HG_{1a-4}, HGC_2	Sí
St_{14}	Servicio corto más servicio expreso	HG_{1a-4}, HGC_{3a}	Sí
St_{15}	Servicio corto más servicio expreso	HG_{1a-4}, HGC_{3b}	Sí

Las etapas St_1 y St_2 corresponden a los casos base donde se resuelve el problema utilizando solamente un servicio regular en ambos sentidos. Ambas soluciones óptimas se pueden obtener fácilmente en forma analítica: en el primer caso el óptimo se puede obtener mediante la fórmula de la raíz cuadrada (3.10), y en el segundo caso la frecuencia óptima el máximo entre la de la primera etapa y la impuesta por el perfil máximo (3.57). Estos dos escenarios se utilizarán como referencia para calcular los ahorros obtenidos en los otros escenarios.

Las etapas St_3 y St_4 consisten en ejecutar la heurística HG_2 para identificar los servicios cortos atractivos, y después aplicar HG_{1a} y HG_{1b} (respectivamente) sobre cada servicio corto atractivo (incluyendo los servicios regulares que cubren el corredor completo, por supuesto). Para el conjunto de líneas obtenido se resuelve el problema de optimización de frecuencias, ignorando las restricciones de capacidad, y se eliminan las líneas con frecuencias nulas de forma de alivianar el problema para los próximos pasos. La etapa St_5 consiste simplemente en hacer la unión de los conjuntos de líneas generados por St_3 y St_4 y luego buscar las frecuencias óptimas, sin considerar capacidad.

En la etapa St_6 se toma como punto de partida la solución entregada por St_5 , y se aplica sobre ella la heurística de generación de servicios zonales, HG_3 . La etapa St_7 parte de la solución entregada por St_6 , y aplica sobre ella la heurística de generación de servicios adicionales HG_4 .

Para hacer más completo el análisis se optó por incorporar una etapa que entregue la solución del problema de óptimo social con restricciones de capacidad. La solución alcanzada en esta etapa sólo respeta la restricción de capacidad en forma aparente,

debido a que fuerza un comportamiento en los usuarios que no necesariamente corresponde con el que adoptarían en forma espontánea. Esta etapa fue incorporada para entregar una noción de la brecha que puede existir entre una solución de este tipo y una solución óptima que sí considera comportamiento de usuarios. Adicionalmente, esta etapa podría ser utilizada como cota para la solución de las próximas etapas, y podría servir para evitar perder el tiempo iterando en escenarios donde no hay margen de mejora para la solución de un único servicio regular. Esta etapa, que llamaremos St_8 , toma la solución de St_7 y resuelve el problema con capacidad como un problema de óptimo social. Este problema equivale a resolver el problema base separado por sentidos planteado en el Capítulo 2 agregando restricciones de capacidad como la que se muestra en (2.18). Por motivos de capacidad de la herramienta utilizada para la solución (solver MINOS, programado en AMPL) se diseñó un algoritmo en el que las restricciones se inicializaban en un principio todas relajadas, y se activaba por cada iteración la restricción que representaba un mayor déficit de capacidad.

En las etapas restantes, de St_9 a St_{15} , se aplican las heurísticas desde HC_1 a HGC_{3b} sobre la solución obtenida en St_7 .

Para representar adecuadamente el impacto de cada metodología, se opta por concentrar el análisis de los beneficios en aquella parte del costo social que realmente puede ser reducido. Así, se decide tomar como medida de desempeño el costo social corregido, definido como sigue:

$$CSC = SC - FTTC \quad (3.90)$$

La expresión anterior corrige los costos sociales totales restando la parte fija de los tiempos de viaje en vehículo que es imposible reducir. Esto equivale a considerar los costos por operación más espera y restar los ahorros en tiempo de viaje de los usuarios respecto a la solución del servicio regular. El término de tiempos de viaje fijo $FTTC$ se puede calcular como se muestra en la fórmula (3.91):

$$FTTC = \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} \theta_{tt} T_w (t_a^w - (j - i + 1)\tau) \quad (3.91)$$

En la expresión anterior se calcula el tiempo de viaje sin detención tomando el tiempo del servicio regular y restando los tiempos de detención en los paraderos intermedios.

En primer lugar se evaluaron las etapas correspondientes a casos sin capacidad activa. Los costos sociales corregidos obtenidos para las cuatro matrices base aplicando cada una de las etapas que no consideran capacidad se presentan en las tablas a continuación. Los resultados se presentan para cada de matriz de viajes en lugar de escenario debido a que en esta fase el nivel de capacidad no tiene efecto sobre ellos.

Tabla 3-7: Costos sociales corregidos (MM \$/hr), sin capacidad activa

Esc.	P1	P2	G1	G2
<i>St</i>₁	2,59	2,57	1,88	3,25
<i>St</i>₃	2,16	2,17	1,85	3,10
<i>St</i>₄	2,16	2,17	1,85	3,10
<i>St</i>₅	2,16	2,17	1,85	3,10
<i>St</i>₆	2,16	2,17	1,85	3,10
<i>St</i>₇	2,16	2,17	1,85	3,10

En esta tabla se puede apreciar que los costos sociales que se obtienen utilizando las estrategias St_3 a St_7 son los mismos. Esto significa que el indicador sencillo (implementado en St_3), utilizado en conjunto con el algoritmo de generación de servicios cortos, es una buena herramienta para generar los servicios expresos en el caso en que no se alcanza la capacidad y que no se justifica utilizar su versión más elaborada (St_4). Cabe notar que los valores obtenidos para las diferentes etapas del algoritmo pueden tener diferencias en los decimales que no se muestran en la tabla, y, lo que es más importante, pueden utilizar diferentes servicios en sus soluciones, lo que puede implicar un efecto perceptible en la solución que se obtendría en la solución con capacidad.

Los valores destacados en negrita indican el mínimo dentro de cada columna. Como las etapas desde St_5 en adelante incluyen las líneas generadas por las etapas anteriores, resulta natural que el menor valor se alcance siempre en St_7 (a pesar de que esta diferencia es apenas perceptible). En el caso de la matriz G1 el óptimo es alcanzado en St_3 y no es mejorado en las etapas sucesivas.

A continuación se presenta un análisis a nivel de línea para cada matriz de viajes, de forma de poder identificar, más allá del aporte a la reducción de los costos sociales, cuáles son los algoritmos que efectivamente generan las líneas que aparecen en la solución óptima de cada matriz. Para esto se presenta, para cada matriz origen destino, una tabla resumen de los servicios provistos en el óptimo, junto a la frecuencia de cada uno de estos, la carga máxima que llevan, y la etapa del algoritmo responsable de la generación de dicho servicio.

Tabla 3-8: Servicios óptimos para P1 y P2 sin capacidad

Matriz	Línea	Ida	Vuelta	Frecuencia (bus/hr)	Carga (pax/hr)	Primera Aparición (Etapa)
P1	1	1111111111	0000000000	21,3	31,3	St1
P1	2	0000000000	1111111111	28,2	132,6	St1
P1	3	1100011111	0000000000	8,5	55,4	St3
P1	4	1000001111	0000000000	30,8	38,5	St3
P1	5	0000000000	1111011111	3,5	126,0	St3
P1	6	0000000000	1010000001	37,7	184,8	St3
P1	7	1100000000	0000000000	29,9	8,2	St3
P1	8	0000000000	0000000011	29,9	47,3	St3
P1	9	1111000000	0000000000	13,6	17,8	St3
P1	10	0000000000	0000011111	14,9	98,4	St3
P1	11	0000000000	1110000001	0,5	195,2	St6
P1	12	1011000000	0000000000	1,3	9,6	St7
P1	13	1000010111	0000000000	5,7	31,8	St7
P1	14	1000101011	0000000000	2,2	34,6	St7
P1	15	1010001011	0000000000	1,4	36,9	St7
P2	1	1111111111	0000000000	17,7	30,4	St1
P2	2	0000000000	1111111111	11,9	91,8	St1
P2	3	1100011111	0000000000	40,3	78,3	St3
P2	4	1100010111	0000000000	9,2	63,6	St3
P2	5	0000000000	1110001111	14,2	189,7	St3
P2	6	0000000000	1110001101	21,8	171,8	St3
P2	7	0000000000	1110000101	7,3	143,0	St4
P2	8	0000000000	1110000111	2,9	153,2	St6
P2	9	1110000111	0000000000	2,9	60,0	St7
P2	10	1101000111	0000000000	0,3	59,9	St7
P2	11	1111110111	0000000000	2,0	30,4	St7
P2	12	0000000000	1011011001	5,9	134,7	St7
P2	13	0000000000	1010111001	2,8	133,8	St7
P2	14	0000000000	1011111111	5,7	91,3	St7

En primer lugar se observa que el algoritmo ha logrado obtener, para las dos matrices acá expuestas, soluciones óptimas que involucran la utilización de hasta 15 servicios diferentes en cada caso, todos ellos diseñados por las heurísticas propuestas. Se

puede observar la existencia de servicios con frecuencias muy bajas y cargas máximas altas (el servicio 11 de P1, por ejemplo). Una frecuencia así de baja no se condice con los supuestos de comportamiento realizados respecto a los tiempos de espera, pero no se descartará este tipo de servicios porque sirven para alimentar las heurísticas de la etapa con congestión. De hecho, un servicio que es atractivo para tantos usuarios a pesar de su poca frecuencia es claramente un muy buen candidato para ayudar a aliviar la carga del servicio regular en un escenario congestionado.

Cabe aclarar que estos servicios de baja frecuencia que aparecen en la solución, a pesar de no tener mucho sentido para ser implementados en la práctica, pueden perfectamente pertenecer a una solución localmente óptima. Un ejemplo sencillo donde esto mismo ocurre fue entregado en el Capítulo 2, en el apartado sobre la multiplicidad de óptimos locales del problema.

Se puede apreciar también que en este caso los servicios que resultan relevantes para la solución (aparte del regular) son generados en las etapas 3, 4, 6, y 7, lo que en principio indicaría que todas las heurísticas son capaces de encontrar servicios beneficiosos en alguna instancia. También se puede apreciar en la solución de P1 la existencia de cinco servicios cortos, lo que implica que a pesar de no estar asociada a una etapa en particular, la heurística para la identificación de servicios cortos resultó de gran utilidad para la obtención de esta solución.

Los servicios óptimos para las matrices G1 y G2 sin capacidad se presentan la Tabla 3-9 a continuación.

Tabla 3-9: Servicios óptimos para G1 y G2 sin capacidad

Matriz	Línea	Ida	Vuelta	Frecuencia (bus/hr)	Carga (pax/hr)	Primera Aparición (Etapa)
G1	1	1111111111	0000000000	30,6	129,7	St1
G1	2	0000000000	1111111111	30,6	42,9	St1
G1	3	1111110111	0000000000	7,4	119,8	St3
G1	4	1110110111	0000000000	17,1	104,7	St3
G1	5	1100110111	0000000000	0,7	90,0	St3
G1	6	0000000000	1110111111	25,0	36,0	St3
G1	7	0000000000	1110011111	0,2	30,9	St3
G2	1	1111111111	0000000000	32,3	131,9	St1
G2	2	0000000000	1111111111	39,8	93,3	St1
G2	3	1110110111	0000000000	18,0	165,2	St3
G2	4	1100110111	0000000000	30,9	135,8	St3
G2	5	0000000000	1110111111	20,3	76,5	St3
G2	6	0000000000	1110011111	14,9	65,0	St3
G2	7	1101010101	0000000000	8,1	118,9	St7
G2	8	1110110101	0000000000	0,1	159,1	St7
G2	9	0000000000	1110111011	14,5	70,9	St7

En el caso de las matrices de Av. Grecia se puede apreciar que el algoritmo de generación de líneas iniciales proporciona casi todos los servicios beneficiosos. En estos escenarios no hay presencia de servicios cortos en las soluciones óptimas, posiblemente como consecuencia de la forma en particular del perfil de carga de esta matriz y de los supuestos de comportamiento que no permiten que los usuarios dejen pasar buses que pasan llenos para esperar uno con capacidad. En G1, como ya se mencionaba cuando se presentaron los valores óptimos por escenario, el óptimo se alcanzó en St_3 y las etapas sucesivas no significaron un nuevo aporte. En el caso de G2 la etapa St_7 permitió la incorporación de tres nuevos servicios de alta participación en la solución.

A continuación se presentan los costos sociales corregidos para las etapas del algoritmo que sí consideran capacidad limitada en los vehículos.

Tabla 3-10: Costos sociales corregidos (MM \$/hr), con capacidad activa

Esc.	P1A	P1M	P1B	P2A	P2M	P2B	G1A	G1M	G1B	G2A	G2M	G2B
<i>St</i> ₂	2,63	2,98	3,57	2,57	2,71	3,09	1,88	1,88	2,01	3,25	3,31	3,69
<i>St</i> ₈ *	2,16	2,24	2,49	2,17	2,23	2,46	1,85	1,85	1,96	3,10	3,12	3,38
<i>St</i> ₉	2,16	2,29	2,83	2,17	2,31	2,62	1,85	1,86	2,00	3,10	3,19	3,46
<i>St</i> ₁₀	2,16	2,28	2,83	2,17	2,31	2,61	1,85	1,86	2,00	3,10	3,18	3,46
<i>St</i> ₁₁	2,16	2,28	2,62	2,17	2,31	2,61	1,85	1,86	2,04	3,10	3,18	3,56
<i>St</i> ₁₂	2,19	2,38	2,83	2,58	2,32	2,56	2,02	1,88	2,01	3,30	3,24	3,48
<i>St</i> ₁₃	2,52	2,79	3,31	2,58	2,71	2,66	2,02	1,88	2,01	3,30	3,31	3,69
<i>St</i> ₁₄	2,19	2,28	2,83			2,56						
<i>St</i> ₁₅	2,19	2,28	2,62			2,58						

* La etapa *St*₈ corresponde a óptimo social y no asegura un comportamiento racional de usuarios.

En esta tabla se presentan resultados para las etapas *St*₁₄ y *St*₁₅ solamente cuando dichas etapas del algoritmo eran capaces de encontrar una combinación válida de servicios para la optimización. Los números destacados en negrita corresponden a los valores óptimos alcanzados en cada escenario, descartando la solución no factible entregada por *St*₈. Notar que estos valores en negrita coinciden para los escenarios de capacidad más alta con los óptimos de la Tabla 3-7: Costos sociales corregidos (MM \$/hr), sin capacidad activa.

Con respecto a las soluciones obtenidas para la etapa *St*₈ se observa en primer lugar que cuando el nivel de capacidad es alto, el valor óptimo de esta etapa coincide con el valor óptimo del escenario correspondiente (es decir, el menor costo obtenido dentro de las otras etapas para el mismo escenario). Esto se debe a que cuando la capacidad no es activa, no existe diferencia entre el comportamiento óptimo de los

usuarios con el óptimo social, como ya fue explicado anteriormente. También se puede observar que la brecha entre ambos tipos de comportamiento crece a medida que la capacidad es más restrictiva.

Con respecto al algoritmo St_9 , que corresponde a la heurística de ajuste de capacidad propuesta en Leiva et al. (2010), se puede ver que a pesar de que en más de una ocasión alcanza el valor óptimo encontrado para cada escenario, siempre se ve igualado o superado por el algoritmo St_{10} , que corresponde a la mejora propuesta para dicho algoritmo en el Capítulo 2. En la tabla a continuación se muestra el número de iteraciones realizadas por cada etapa para la obtención de la solución presentada:

Tabla 3-11: Número de iteraciones del algoritmo para cada etapa y escenario

Esc.	P1A	P1M	P1B	P2A	P2M	P2B	G1A	G1M	G1B	G2A	G2M	G2B
St_2	2	2	2	1	2	2	1	2	2	1	2	2
St_8	1	5	8	1	7	12	1	2	4	1	8	9
St_9	1	98	177	1	79	119	1	48	85	1	79	155
St_{10}	1	23	13	1	7	5	1	14	4	1	10	8
St_{11}	1	23	27	1	7	5	1	14	18	1	10	6
St_{12}	1	23	27	1	7	5	1	14	18	1	10	6
St_{13}	1	23	27	1	7	5	1	14	18	1	10	6
St_{14}	1	11	6			5						
St_{15}	1	11	20			19						

En esta tabla se puede ver que el número de iteraciones efectuadas por St_9 es considerablemente más alto que el de cualquiera de los otros métodos de solución. Este hecho, en conjunto de que no representa una mejora en la solución respecto a los otros métodos, justifica la eliminación de esta etapa dentro del algoritmo de solución a implementar.

Cabe recordar que St_8 , que resuelve un problema de óptimo social con restricciones de capacidad, fue implementado también en forma iterativa debido a que el gran número de restricciones involucradas en el problema disparaba su nivel de complejidad hasta un punto no abordable para el solver utilizado, y que por esto se implementó un método iterativo en el que las restricciones se añadían al modelo en forma iterativa, seleccionando en cada paso la restricción cuya brecha fuera más crítica.

Los servicios utilizados en los escenarios con capacidad activa de la matriz P1 se presentan en la tabla a continuación.

Tabla 3-12: Servicios óptimos para P1 con capacidad

Esc.	Etapas Ópt.	Línea	Ida	Vuelta	Frec. (bus/hr)	Carga (Pax/bus)	Primera Aparición
P1M	St14	1	1111111111	0000000000	27,3	21,1	St1
P1M	St14	2	0000000000	1111111111	34,2	120,5	St1
P1M	St14	3	1100001111	0000000000	5,3	33,3	St3
P1M	St14	4	1000000111	0000000000	37,3	32,0	St3
P1M	St14	5	1100000000	0000000000	30,4	8,1	St3
P1M	St14	6	0000000000	0000000011	30,4	47,4	St3
P1M	St14	7	1111000000	0000000000	10,9	17,2	St3
P1M	St14	8	0000000000	0000001111	11,8	99,1	St3
P1M	St14	9	0000000000	1110000001	61,6	120,0	St6
P1M	St14	10	1011000000	0000000000	0,9	9,0	St7
P1M	St14	11	1000001011	0000000000	25,9	28,7	St7
P1B	St15	1	1111111111	0000000000	31,7	14,8	St1
P1B	St15	2	0000000000	1111111111	44,1	80,4	St1
P1B	St15	3	1100001111	0000000000	5,8	23,8	St3
P1B	St15	4	1000000111	0000000000	58,1	20,8	St3
P1B	St15	5	1111000000	0000000000	48,1	11,4	St3
P1B	St15	6	0000000000	0000001111	49,7	63,9	St3
P1B	St15	7	0000000000	1110000001	92,4	80,0	St6
P1B	St15	8	1011000000	0000000000	1,7	4,4	St7
P1B	St15	9	1000001011	0000000000	40,9	18,6	St7

En estos dos casos se aprecia que la solución óptima se encuentra en las etapas que incorporan combinaciones de servicios expresos diseñados para congestión. La diferencia entre ambos escenarios es que en el primero el óptimo se alcanzó en la etapa St_{14} , que realiza el ajuste de capacidad utilizando el criterio de priorizar las líneas de mayor déficit de capacidad, y en el segundo escenario se obtenía mejores resultados en St_{15} , priorizando a los servicios expresos al momento de aumentar la capacidad. De todas formas no está de más notar que en el escenario P1M St_{15} también entregaba el mejor de los valores encontrados.

Lo que resulta interesante de este resultado es que, dado que todas las líneas utilizadas en el óptimo aparecen en alguna etapa de generación de líneas, la solución encontrada era factible para las etapas St_9 , St_{10} , St_{11} que son las que buscan una solución a partir del óptimo irrestricto. Esto sirve como ejemplo de un caso donde la solución de estas heurísticas iterativas es subóptima, y muestra que el enfoque analítico para el problema puede en algunos casos superar al enfoque heurístico. Finalmente, se puede notar la ausencia en estas soluciones de líneas generadas en St_4 y St_5 , lo que parece indicar que no tiene mucho sentido utilizar la versión compleja del indicador para la generación de líneas iniciales.

Tabla 3-13: Servicios óptimos para P2 con capacidad

Esc.	Etapas Ópt.	Línea	Ida	Vuelta	Frec. (bus/hr)	Carga (Pax/bus)	Primera Aparición
P2M	St10	1	1111111111	0000000000	18,8	23,2	St1
P2M	St10	2	0000000000	1111111111	11,9	69,3	St1
P2M	St10	3	1100001111	0000000000	41,6	60,0	St3
P2M	St10	4	1100001011	0000000000	22,9	47,2	St3
P2M	St10	5	0000000000	1110001111	77,6	120,2	St3
P2M	St10	6	1110000111	0000000000	5,8	46,0	St7
P2M	St10	7	1101000111	0000000000	2,5	45,8	St7
P2M	St10	8	1111110111	0000000000	5,3	23,2	St7
P2M	St10	9	0000000000	1011011001	1,8	101,4	St7
P2M	St10	10	0000000000	1011111111	5,5	69,3	St7
P2B	St14	1	1111111111	0000000000	19,8	17,9	St1
P2B	St14	2	0000000000	1111111111	67,1	80,0	St1
P2B	St14	3	1100001111	0000000000	39,3	45,8	St3
P2B	St14	4	1100001011	0000000000	45,4	34,6	St3
P2B	St14	5	1110000111	0000000000	11,0	35,0	St7
P2B	St14	6	1101000111	0000000000	6,8	34,9	St7
P2B	St14	7	1111110111	0000000000	8,6	17,9	St7
P2B	St14	8	0000000000	1110000001	63,8	80,0	St12

En los escenarios asociados a P2 volvemos a encontrar una solución obtenida por medio de la heurística analítica, y aparece una solución también obtenida por medio del método iterativo de las capacidades.

La solución obtenida en P2B incorpora y utiliza a capacidad un servicio que no resultaba atractivo en la etapa sin capacidad (servicio 8), como se discutía anteriormente. Este servicio es introducido por primera vez en una solución en la etapa St_{12} , que optimiza un servicio zonal en combinación con un servicio regular.

En estos escenarios vuelven a destacar las mismas etapas y algoritmos de generación de líneas como los más efectivos.

Los servicios obtenidos para los escenarios con capacidad asociados a las matrices G1 y G2 se entregan en las Tabla 3-14: Servicios óptimos para G1 con capacidad y Tabla 3-15: Servicios óptimos para G2 con capacidad:

Tabla 3-14: Servicios óptimos para G1 con capacidad

Esc.	Etapa Ópt.	Línea	Ida	Vuelta	Frecuencia (bus/hr)	Carga (Pax/bus)	Primera Aparición
G1M	St10	1	1111111111	0000000000	47,0	120,5	St1
G1M	St10	2	0000000000	1111111111	30,9	42,0	St1
G1M	St10	3	1110110111	0000000000	10,2	101,8	St3
G1M	St10	4	0000000000	1110111111	25,3	35,2	St3
G1M	St10	5	0000000000	1110011111	1,0	30,1	St3
G1B	St10	1	1111111111	0000000000	83,8	80,0	St1
G1B	St10	2	0000000000	1111111111	35,6	31,7	St1
G1B	St10	3	0000000000	1110111111	29,9	25,8	St3
G1B	St10	4	0000000000	1110011111	18,3	20,6	St3
G2B	St10	5	0000000000	1110111011	41,7	59,8	St7

Tabla 3-15: Servicios óptimos para G2 con capacidad

Esc.	Etapas Ópt.	Línea	Ida	Vuelta	Frecuencia (bus/hr)	Carga (Pax/bus)	Primera Aparición
G2M	St10	1	1111111111	0000000000	26,6	110,2	St1
G2M	St10	2	0000000000	1111111111	42,5	83,2	St1
G2M	St10	3	1110110111	0000000000	77,3	120,2	St3
G2M	St10	4	0000000000	1110111111	14,1	67,4	St3
G2M	St10	5	0000000000	1110011111	24,5	56,9	St3
G2M	St10	6	1101010101	0000000000	1,3	106,1	St7
G2M	St10	7	0000000000	1110111011	24,2	61,9	St7
G2B	St10	1	1111111111	0000000000	34,8	80,5	St1
G2B	St10	2	0000000000	1111111111	49,1	38,8	St1
G2B	St10	3	1110110111	0000000000	119,1	80,4	St3
G2B	St10	4	0000000000	1110011111	63,1	45,4	St3
G2B	St10	5	0000000000	1110111011	41,7	59,8	St7

Todos los escenarios asociados a las matrices G1 y G2 encuentran su mejor solución en la etapa St_{10} , que utiliza la heurística de incrementos de capacidad HC_2 . Nuevamente observamos que los servicios utilizados en el óptimo son los generados por St_3 y St_7 .

Las soluciones obtenidas por las heurísticas pueden resultar un poco complejas para poder ser aplicables a la práctica. Por una parte, puede ser deseable imponer una cota al número de servicios operando en paralelo. Además, las soluciones consideran servicios de frecuencia muy baja, que no tienen mucho sentido en la práctica. Estos dos problemas serán abordados en el Capítulo 5, donde se proponen técnicas sencillas para corregir estas características de las soluciones obtenidas inicialmente para construir un diseño aplicable y realista.

3.3 Diseño de servicios expresos en un corredor

Una vez realizadas las pruebas descritas en la sección anterior fue posible determinar cuáles elementos resultan útiles para ser incorporados en la heurística de solución del problema de diseño de servicios expresos. En particular con respecto a las heurísticas probadas las conclusiones a las que se llegó fueron:

1. La heurística de generación de servicios iniciales HG_{1a} entrega buenos servicios en forma muy sencilla de implementar, sobre todo al ser aplicada en conjunto con la heurística de generación de servicios cortos HG_2 . La versión compleja de la heurística de generación de servicios iniciales HG_{1b} no entrega mayores beneficios sobre los que se obtienen con la versión sencilla.
2. La heurística de generación de servicios zonales HG_3 resultó ser útil en algunas instancias en particular. Debido a que es de muy sencilla implementación se opta por mantenerla en la generación de servicios.
3. La heurística para la generación de servicios adicionales HG_4 probó ser de gran utilidad y los servicios generados por esta heurística suelen aparecer en las soluciones de los casos más congestionados, lo que justifica su incorporación en un algoritmo consolidado a pesar de ser menos eficiente que las otras heurísticas para la generación de servicios en términos de iteraciones.
4. La heurística de ajuste de capacidad por incrementos fijos HC_1 es considerablemente más lenta que las otras opciones probadas y no entrega resultados mejores, por lo que será dejada fuera del algoritmo.

5. La heurística de ajuste de capacidad mejorada HC_2 es capaz de encontrar buenas soluciones en un número razonable de iteraciones. La heurística HC_3 , que es similar a HC_2 pero que prioriza a los servicios expresos al momento de elegir líneas en déficit para aumentar la capacidad, también encuentra buenas soluciones en un número razonable de iteraciones. Entre estas dos heurísticas no es posible determinar si alguna es mejor que la otra a partir de las pruebas realizadas, pero dado que son simples de implementar y no consumen mucho tiempo se opta por utilizar ambas y escoger la mejor solución entre ambas.
6. Con respecto a las heurísticas de generación de servicios para el problema con capacidad limitada, la heurística que construye servicios cortos HGC_1 no mostró ser particularmente útil por si sola en las pruebas, mientras que la heurística que resuelve el problema de capacidad usando servicios zonales HGC_2 mostró mejores resultados. Sin embargo, los mejores resultados para esta categoría de algoritmos se obtuvieron mediante el uso de heurísticas que combinaban ambos formatos, HGC_{3a} y HGC_{3b} . La relación entre estas dos últimas heurísticas es similar a la ya explicada para HC_2 y HC_3 , y se dejan las dos en el algoritmo por el mismo motivo. Dado que HGC_1 y HGC_2 son pasos para la solución de HGC_{3a} y HGC_{3b} también serán consideradas en el algoritmo final.

Con todo esto es posible describir el algoritmo consolidado como sigue:

Algoritmo para el diseño de servicios expresos en un corredor:

- 1) Solución inicial servicio regular.

- a) Optimizar el corredor para el caso donde opera sobre él un único servicio regular, para el caso sin congestión (fórmula (3.10)) y para el caso con congestión ($f_a^* = P_{a^*}/cap$).
- 2) Solución del problema sin capacidad.
 - a) Generación de líneas a priori.
 - i) Aplicando HG_3 identificar el conjunto de servicios cortos que podrían ser atractivos \mathcal{L}_C^* . Incluir en \mathcal{L}_C^* también los servicios regulares de extremo a extremo para ambos sentidos. Hacer $\mathcal{L} \leftarrow \mathcal{L}_C^*$.
 - ii) Aplicar HG_{1a} a partir de ambos servicios regulares, incorporando las líneas obtenidas al conjunto \mathcal{L} .
 - iii) Aplicar HG_{1a} a partir de cada elemento del conjunto \mathcal{L}_C^* que posea tres o más paraderos, incorporando las líneas obtenidas al conjunto \mathcal{L} .
 - iv) Aplicar HG_4 incorporando los servicios nuevos generados a \mathcal{L} .
 - v) Optimizar las frecuencias y eliminar de \mathcal{L} las líneas cuya frecuencia sea nula.
 - b) Generación de líneas adicionales.
 - i) Aplicar HG_5 tomando $\mathcal{L}_R = \mathcal{L}_C^*$.
 - ii) Optimizar las frecuencias y eliminar de \mathcal{L} las líneas cuya frecuencia sea nula.
- 3) Solución del problema con capacidad.
 - a) Registrar la solución actual del problema sin restricción de capacidad.

- b) A partir de la solución sin capacidad, aplicar el algoritmo HC_2 y registrar la solución obtenida.
- c) Repetir el paso 3.b. usando HC_3 , y HGC_1 a HGC_{3b} .
- d) Comparar las soluciones obtenidas (cuidando de incluir también la del paso 1.a.) y escoger la mejor.

El algoritmo acá planteado es capaz de encontrar un conjunto de servicios y sus frecuencias óptimas para un corredor, considerando restricción de capacidad en los vehículos, y donde los usuarios se asignan a rutas mínimas.

La heurística acá planteada trabaja bajo la condición de que los paraderos de ida coincidan con los de vuelta en el corredor. Esta condición se impone para facilitar la modelación de los servicios cortos y de las restricciones de continuidad de flujo. Sin embargo, existen formas de generalizar la heurística para abarcar estos casos. Una primera opción consistiría en descartar de plano los servicios cortos y enfocarse en los otros tipos de servicio, los cuales no requieren de esta coincidencia entre nodos de ida y de vuelta. Una segunda opción consistiría en hacer una homologación (quizás un tanto forzosa) entre los paraderos de ida y de vuelta, asumiendo que el efecto de las distancias entre ellos es despreciable. En todo caso, el análisis realizado en este capítulo para deducir los indicadores que permiten construir los servicios cortos no debiese complicarse demasiado al agregar los costos por movimientos entre paradas.

El algoritmo propuesto se podría también extender para abarcar los casos planteados como variantes del problema en el Capítulo 2. Resultaría de particular interés incorporar el

enfoque binivel como alternativa de solución del problema de optimización de frecuencias, ya que posibilitaría la aplicación del modelo en corredores de mayor tamaño.

Dos de las heurísticas planteadas, la de generación de servicios expresos iniciales y la de generación de servicios adicionales, corresponden conceptualmente a una metaheurística tipo *greedy*, en la que los nodos son seleccionados en cada iteración de acuerdo a un ranking construido con los indicadores. Estas heurísticas podrían ser generalizadas implementando una metaheurística tipo GRASP (Kochenberger, 2003), que amplía la búsqueda incorporando un componente de aleatoriedad en la selección.

En los capítulos restantes se pueden ver instancias donde esta heurística es aplicada exitosamente para el diseño de servicios de diferentes corredores. En el Capítulo 4 la heurística consolidada se aplica en cerca de mil escenarios conformados por corredores de diferentes atributos para estudiar qué atributos son los predictores más relevantes del desempeño de los servicios expresos como estrategia de reducción de costos sociales. En el Capítulo 5 esta heurística es integrada en una heurística general para la solución del problema para una red.

4. ¿CUÁNDO UTILIZAR SERVICIOS EXPRESOS?

En este capítulo se presenta un ejemplo de implementación de la heurística presentada al final del Capítulo 3 sobre corredores de distintas características, con el fin de mostrar su aplicabilidad, y de estudiar el potencial de los servicios expresos como herramienta de gestión. Para esto, se construyó un gran número de escenarios de forma de identificar, analizando los resultados entregados por el modelo en cada uno de ellos, cuáles son las principales señales de que un determinado corredor representa una buena oportunidad para implementar servicios expresos.

4.1 Experiencias previas

El autor de esta tesis ya ha realizado algunos esfuerzos en esta línea de investigación, que han sido presentados en congresos y publicados en revistas (Leiva et al., 2010, Larrain et al., 2010a, Larrain et al., 2010b). Estos resultados, que se detallan en lo que sigue, han sido obtenidos mediante la aplicación del modelo en su formulación original, y en general se concentran en un tipo en particular de atributos del corredor para el análisis. El experimento que se detalla en este capítulo engloba estos esfuerzos, incluyendo todos los atributos relevantes que se han detectado hasta el momento en estos experimentos previos, y además implementa la versión del modelo que se presenta en el Capítulo 3, que es mejor y más flexible al no depender de la intuición del modelador para generar la configuración de los servicios a proveer.

- **Efecto de la demanda**

El primero de estos experimentos, publicado como *Choosing the Right Express Services for a Bus Corridor with Capacity Restrictions* en la revista *Transportation Research Record* (Larrain et al., 2010a), estudia el efecto que tienen sobre la conveniencia de implementar servicios expresos en un corredor las siguientes características de la demanda que existe sobre éste:

1. Forma del perfil de carga: En este experimento se prueba el efecto sobre los servicios expresos de tres diferentes formas de perfil; decreciente, con carga concentrada en la mitad, y con dos puntas, como se puede ver en la Figura 4-1. Se asume para la modelación que los perfiles de carga de ida y vuelta son de la misma forma.
2. Largo promedio de los viajes: Un perfil de demanda dado puede corresponder a diferentes matrices de viaje con diferentes largos promedio, ya que el mismo perfil puede ser producto de un número alto de viajes cortos, o de un número bajo de viajes largos. Se definieron para el experimento tres niveles de largo de viaje (alto, medio y bajo) para cada perfil dado.
3. Desbalance de la demanda entre sentidos: Para este estudio se asumió que el perfil de viajes del sentido de vuelta era el reflejo del perfil de ida, multiplicado por un ponderador, que para fines de este experimento adoptó los valores de 0,8, 0,5 y 0,2.
4. Nivel de demanda: El nivel completo de la demanda también fue sensibilizado en los escenarios de este experimento, multiplicando la matriz de viajes por 0,75, 1, y 1,25.

De la combinación de estos factores fueron concebidos 81 escenarios que fueron posteriormente optimizados aplicando el modelo de Leiva. Una vista esquemática del experimento se entrega en la Figura 4-1 a continuación.

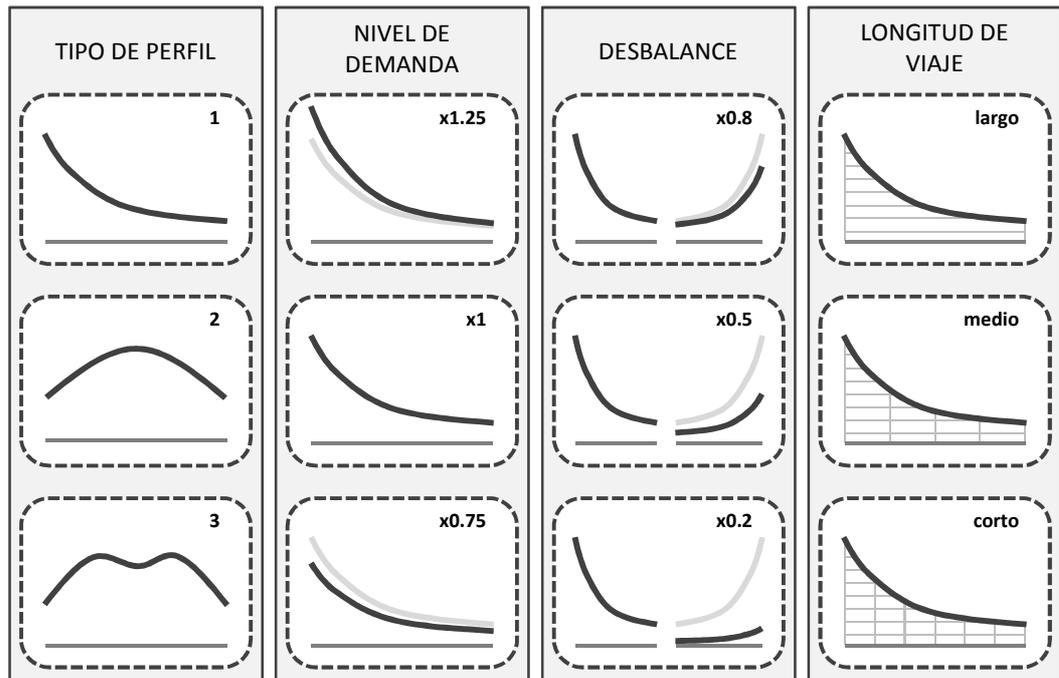


Figura 4-1: Definición de escenarios para el experimento

En la figura se pueden apreciar las tres categorías consideradas para los distintos cuatro factores que definen los 81 escenarios a evaluar del experimento.

El detalle de la metodología de construcción de los escenarios se puede encontrar en el trabajo citado, pero a grandes rasgos consiste en, para determinada forma de perfil, construir vectores de generación y atracción que entreguen el largo de viajes deseado (notar que el largo promedio dependerá únicamente de la forma del perfil y del total de viajes, por lo tanto queda determinado por estos vectores). Aplicando un modelo de máxima entropía se construyeron matrices base para cada cruce entre forma de

perfil y largo promedio, que fueron amplificadas por los factores de desbalance y de nivel de demanda para obtener las matrices de cada escenario según correspondiera. Este estudio concluye que el indicador más importante de los estudiados es el largo promedio de los viajes, y también muestra alguna evidencia de que el nivel de demanda puede ser importante, pero no es categórico en hacer esta afirmación, debido a que el rango en el que se mueven los niveles de demanda del estudio no permiten observar claramente el efecto de esta variable. Con respecto a la forma del perfil de demanda, el estudio observa que los perfiles decrecientes parecen ser una mejor oportunidad para la implementación de los expresos, pero no queda claro si este efecto es separable del hecho de que este tipo de perfiles es precisamente el que presenta normalmente los viajes más largos. Con respecto al desbalance, se destaca la posibilidad que genera para la implementación de servicios menos convencionales, pero no se observa un efecto en los beneficios obtenidos por la implementación de los servicios expresos.

Esta primera aproximación al problema de identificar los atributos más relevantes, a pesar de no entregar resultados definitivos salvo el del largo promedio de viaje, sí logra entregar valiosa información de por dónde encauzar los esfuerzos de un estudio más amplio como el que se reporta en el presente capítulo de esta tesis.

- **Efecto del nivel de concentración de los viajes**

Una continuación del trabajo anterior fue presentada en la 12^a World Conference on Transport Research realizada el año 2010 en Lisboa. Dicha presentación llevó por

título *The Effect of OD Trip Dispersion Versus Concentration in Express Service Design*, y está disponible en los registros de dicho congreso (Larrain et al. 2010b).

Este trabajo muestra el impacto de la concentración de los viajes de la matriz OD en los beneficios que se pueden obtener usando servicios expresos. Se propone como indicador de la concentración de los viajes el coeficiente de variación de la matriz de viajes, ya que una mayor varianza en esta matriz implica que algunos pares deben concentrar un mayor número de viajes.

Para este experimento se seleccionó una matriz de referencia (seleccionada entre las que mostraban tener potencial para los servicios expresos en el experimento anterior) y se derivaron a partir de ella nuevas matrices con diferentes niveles de concentración de viajes. Por construcción, una matriz generada mediante métodos de entropía tiende a distribuir lo mejor posible los viajes disponibles, lo que implica un nivel de concentración de viajes bajo. Por tanto, la estrategia adoptada para generar las 30 matrices a utilizar en este experimento consistió en partir desde la matriz base, generada mediante entropía, y sobre ésta realizar aleatoriamente permutaciones de flujo que gradualmente hiciesen crecer la variabilidad de la matriz, pero sin alterar ni la forma del perfil de viajes, ni el largo promedio ni el total de éstos. En la Figura 4-2 a continuación se presenta un gráfico de los resultados obtenidos en términos de ahorros totales versus el coeficiente de variación de las matrices construidas.

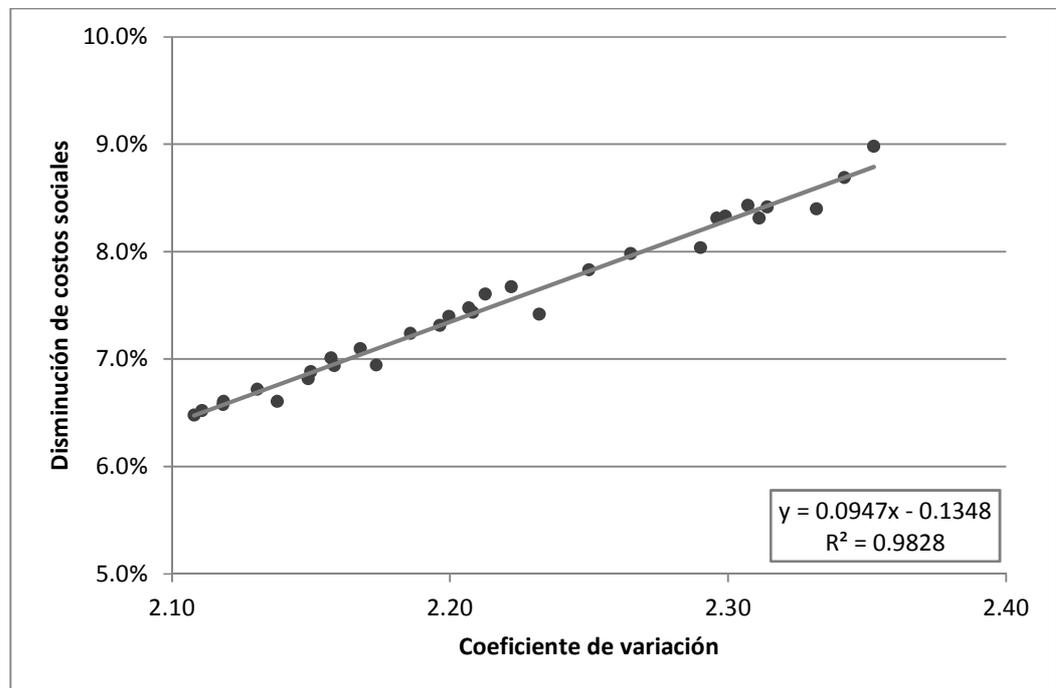


Figura 4-2: Ahorros en costos sociales v/s dispersión de los viajes

De este experimento fue posible concluir, como se aprecia en la gráfica anterior, que un mayor coeficiente de variación en las matrices, y por tanto una mayor concentración de los viajes en ciertos pares hacen aumentar (y con una alta correlación, por cierto) los beneficios a obtener por la aplicación de servicios expresos.

- **Efecto del nivel de demanda**

Un tercer experimento que se realizó utilizando la versión básica del algoritmo se presentó en el congreso Optimization, llevado a cabo en Lisboa el año 2011, en una charla que llevó como título *Choosing the Right Express Services for a Bus Corridor with Capacity Restrictions*.

En el estudio del punto anterior se presenta evidencia que hace parecer que los niveles de demanda total influyen en el desempeño de los servicios expresos como medida de gestión, pero ese estudio no obtienen resultados concluyentes al respecto. Este experimento se diseñó con el fin de estudiar este efecto, y a la vez observar si el estado de capacidad del sistema interfiere de alguna forma en estos resultados. Para esto se construyeron 20 matrices amplificando por distintos factores una matriz base, y se resolvió cada escenario primero en forma irrestricta, y después imponiendo un nivel de capacidad limitante en los vehículos.

La Figura 4-3 muestra los ahorros del sistema versus los niveles de demanda del corredor en estudio.

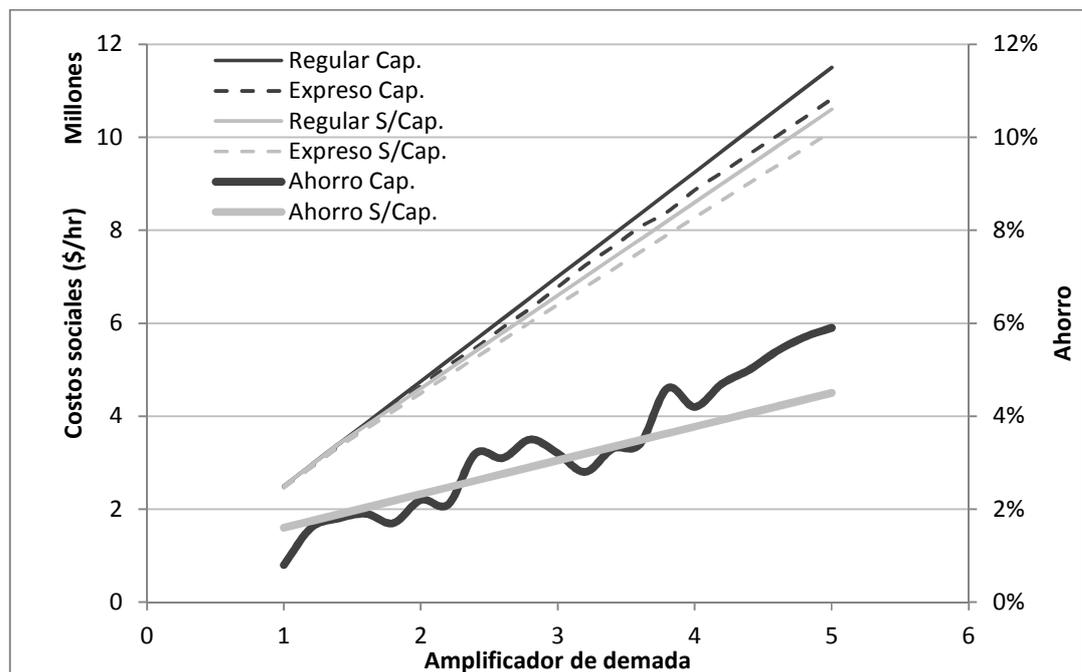


Figura 4-3: Reducción de costos sociales v/s nivel de demanda

En la figura el color gris claro representa los resultados para los escenarios sin capacidad, mientras que los resultados en gris oscuro corresponden a los escenarios con capacidad activa. Las líneas delgadas (en la parte superior del gráfico) muestran para cada caso los costos sociales del escenario optimizado utilizando servicios expresos (línea punteada) y solamente un servicio regular (línea continua). El ahorro, calculado como la diferencia porcentual entre estas dos líneas, es presentado en las líneas gruesas de la parte inferior del gráfico. De acá se concluyen dos cosas: primero, que efectivamente un aumento en el nivel de demanda redundaba en un aumento en los ahorros porcentuales que entrega el servicio expreso, y que en presencia de congestión esta reducción es aún mayor.

Llama la atención la irregularidad con la que se comportan los ahorros del caso con capacidad. Esto probablemente se debe a las limitaciones de la versión del modelo aplicada en este experimento, donde las líneas a considerar debían ser definidas a priori. Cabe notar también que en este experimento, al igual que en los dos anteriores, los costos sociales son calculados sin restar la parte fija de los costos de transporte, lo que hace menos perceptible el efecto de la reducción de costos. En los experimentos que siguen estos costos son medidos eliminando la parte fija de los costos de transporte. El principal motivo para hacer esto, como se detalla más adelante, es que la magnitud de estos costos fijos es alta en comparación con los otros costos involucrados del problema, lo que reduce el impacto comparativo de los servicios expresos.

4.2 Construcción de los escenarios

En lo que sigue, se presenta un experimento que abarcó los efectos cuantificados en los estudios antes expuestos que mostraron ser más relevantes, y se agregaron otras variables de interés no abordadas por los experimentos previos, relacionadas principalmente con la topología de la red. Además, este nuevo experimento utilizó la nueva versión del algoritmo presentada en el Capítulo 3, que no depende de la intuición del modelador al incluir la generación de servicios, y que también presenta mejoras en la solución de escenarios con capacidad activa. En particular, los efectos que se deseaba capturar con este nuevo experimento se enumeran a continuación:

1. Características de la demanda: El experimento debía considerar matrices con una variedad de volúmenes de viajes, largos promedio de viaje, y niveles de concentración de los viajes. En este experimento se optó por dejar fuera la forma del perfil como variable explicativa en sí, ya que por una parte es más bien cualitativa y no cuantitativa, y porque su efecto se puede explicar a través del largo promedio de los viajes. También se dejó fuera el desbalance, debido a que apunta a casos muy específicos que ya han sido estudiados en la literatura (*deadheading*), y porque además el desbalance también puede ser considerado como una forma de concentración de los viajes. Un nuevo factor que se incorpora al análisis es la carga del arco crítico a través de la frecuencia mínima del servicio regular, que comparada con la frecuencia óptima irrestricta sirve para determinar el estado de congestión del corredor bajo análisis.
2. Características de operación: Uno de los factores más importantes hasta ahora no considerado en los experimentos realizados es el tiempo de detención de los

vehículos en los paraderos. De este factor dependerá directamente el ahorro que potencialmente se podría obtener en los costos de tiempos de viaje de la red. Un segundo factor que se deseaba incorporar al análisis era la capacidad de los vehículos, que en estudios previos ya mostraba ser un buen predictor del potencial de los expresos. Un tercer elemento que se quiso incorporar en este análisis correspondía al largo del corredor, en términos del número total de paradas. Otros parámetros operativos y topológicos, como la velocidad de operación, las distancias entre paradas, y consecuentemente los tiempos de viaje entre paradas no son analizados en este experimento. Desde el punto de vista del usuario, los tiempos de viaje entre paradas representan una constante que no puede ser evitada, y que por lo tanto no tiene afecto en la optimización. Desde el punto de vista del operador, variar las distancias (o tiempos) entre paradas tiene un efecto similar al de modificar los parámetros de costo de operación, que ya son sensibilizados en este experimento como se explica en el siguiente punto.

3. Pesos relativos de los componentes de los costos sociales: Los costos sociales a optimizar están compuestos por cuatro elementos; los costos del operador, los costos por tiempo de viaje en vehículo, los costos por esperas, y los costos por trasbordos. Cada uno de estos elementos tiene uno o dos parámetros asociados que determinan su peso en la función objetivo. Para los costos del operador, éstos corresponden a c_L y c_T , que son las componentes por unidad de distancia y de tiempo de los costos de operar un servicio (ver ecuación (2.3) para más detalles). Los tiempos de viaje en vehículo se ponderan por el valor del tiempo

de viaje θ_{tt} , mientras que los de espera se ponderan por el valor del tiempo de espera θ_{wt} . Los tiempos de espera también son ponderados por λ , que representa la proporción del intervalo promedio entre buses que un usuario debe esperar en promedio, y calcula el efecto de la aleatoriedad en la llegada de éstos. Sin embargo, dado que este parámetro siempre aparece multiplicando a θ_{wt} , no tiene mucho sentido sensibilizarlo aparte. Por último, los trasbordos también son ponderados por el costo por trasbordo θ_{tr} .

Debido al gran número de factores que se desea tomar en cuenta para el experimento, el número de escenarios a evaluar es considerable. Para mantener el experimento realizable dentro de un tiempo razonable se optó por simplificar algunos aspectos del problema. En primer lugar, se decidió trabajar sólo con soluciones que no consideran trasbordos, lo que reduce los tiempos de solución de cada escenario en forma considerable. Esta limitación se ve compensada al tomar en cuenta que las soluciones obtenidas sin trasbordos representan una cota inferior de los costos óptimos obtenibles para cada escenario, es decir, si el análisis indica que cierta característica de un escenario es favorable para la implementación de los servicios expresos sin trasbordo, al considerar trasbordos la solución sólo puede ser igual o mejor que la observada.

Otras simplificaciones apuntaban a eliminar variables para la generación de escenarios, para mantener en un rango razonable el número de escenarios a optimizar. Además de desechar como variable el grado de desbalance y la forma de los perfiles, como ya se mencionó antes, se decidió, para los costos del operador, modificar c_L y c_T simultáneamente y no en forma independiente. Específicamente, se

definieron tres niveles para los costos del operador, el nivel alto donde estos dos parámetros toman valor alto, y nivel medio y bajo siguiendo la misma lógica. Por último, dado que para el análisis son los pesos relativos de estos parámetros los que interesa estudiar, se mantuvo θ_{tt} fijo para actuar como punto de referencia, y se hicieron variar los pesos de los otros componentes de los costos sociales.

Los escenarios a optimizar se obtuvieron como la combinación de cuatro opciones de matriz base (construidas para capturar distintos largos promedio, variabilidades, y largos de corredor, principalmente) y diferentes niveles de otros parámetros que se detallan en Tabla 4-1 a continuación:

Tabla 4-1: Parámetros para la generación de escenarios

Parámetro	Número de Opciones	Posibles Valores
Matriz Base	4	1, 2, 3 ó 4
θ_{tt}	1	15 \$/min
θ_{wt}	3	15, 13 ó 45 \$/min
c_L	3	250, 500 ó 750 \$/bus-Km
c_T	1*	$10c_L$ \$/bus-min
Amplificador de Volumen	3	0,5, 1,0, 2,0
τ	3	0,5, 1,0, 1,5 min
Capacidad	3	80, 120, 160 pax/bus

*El valor de c_T va ligado al valor de c_L .

Al igual que en el experimento del capítulo anterior, para el cálculo de los tiempos de viaje y los costos de operación se asumió que estos valores responden a las fórmulas (2.1) a (2.5). Se asumió, además, que los tiempos y distancias asociados a maniobras de retorno de los buses son despreciables.

La combinación de los tipos de matriz con los otros parámetros arroja un total de 972 escenarios a optimizar. Los cuatro tipos de matrices diseñadas para el experimento corresponden a dos tipos de perfil, inspirados en los corredores Av. Pajaritos y Av. Grecia en la ciudad de Santiago, respectivamente, y dos largos de corredor (10 y 20 paraderos por sentido). Los perfiles de carga de los cuatro tipos de matriz se muestran en las figuras a continuación.

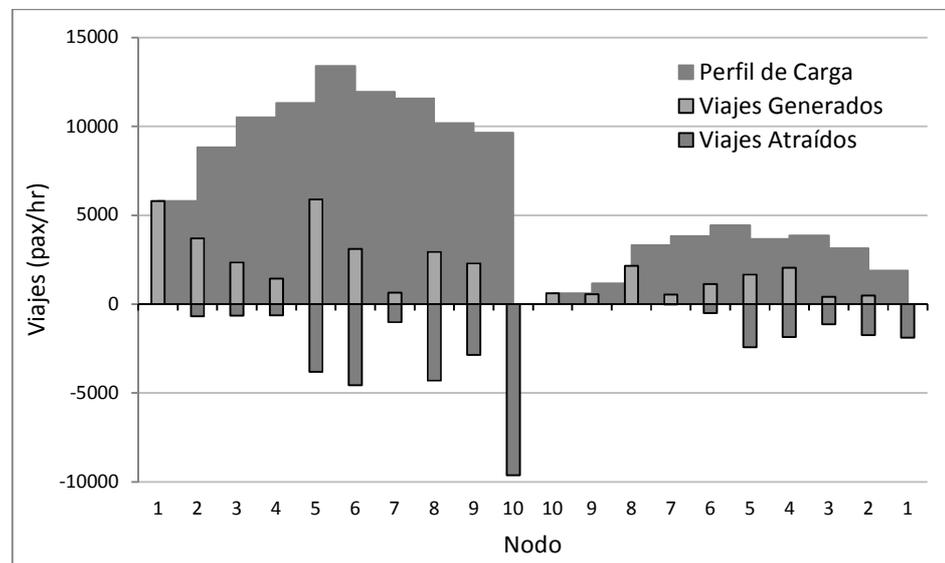


Figura 4-4: Perfil de carga matriz tipo 1

El primer perfil, ilustrado en la figura anterior, corresponde a un perfil de carga tipo Av. Grecia, que es un corredor cuya mayor actividad se concentra en el centro de éste. En el gráfico se presentan ambos sentidos del corredor. El largo del corredor completo es de 7,2 km, y los paraderos se asumen uniformemente espaciados entre sí (es decir, cada 800 m). Se asume además una velocidad de operación de los vehículos de 25 km/h.

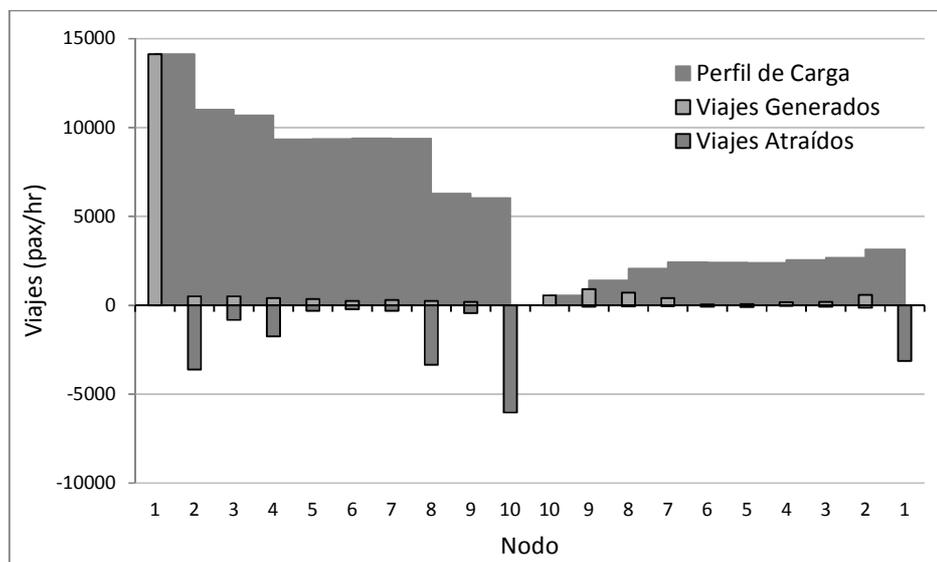


Figura 4-5: Perfil de carga matriz tipo 2

Por otra parte, la figura anterior muestra el perfil de carga de la matriz inspirada en Av. Pajaritos, antes de que funcionara la Línea 5 del Metro de Santiago sobre ese tramo. Este perfil de tipo descendente es el que se observa normalmente en corredores que se usan para ir desde o hacia el centro durante los períodos punta, en la que un gran número de pasajeros abordan el corredor en su comienzo (probablemente desde otro servicio o desde el metro) para alcanzar su destino que se encuentra en alguno de los paraderos a lo largo del corredor. Este perfil corresponde al mismo que fue utilizado en el experimento del Capítulo 3 y que se muestra en la Figura 3-6, pero invirtiendo el sentido del corredor. Para el espaciamiento entre paraderos y velocidad de operación se utilizan los mismos valores que en el perfil tipo 1.

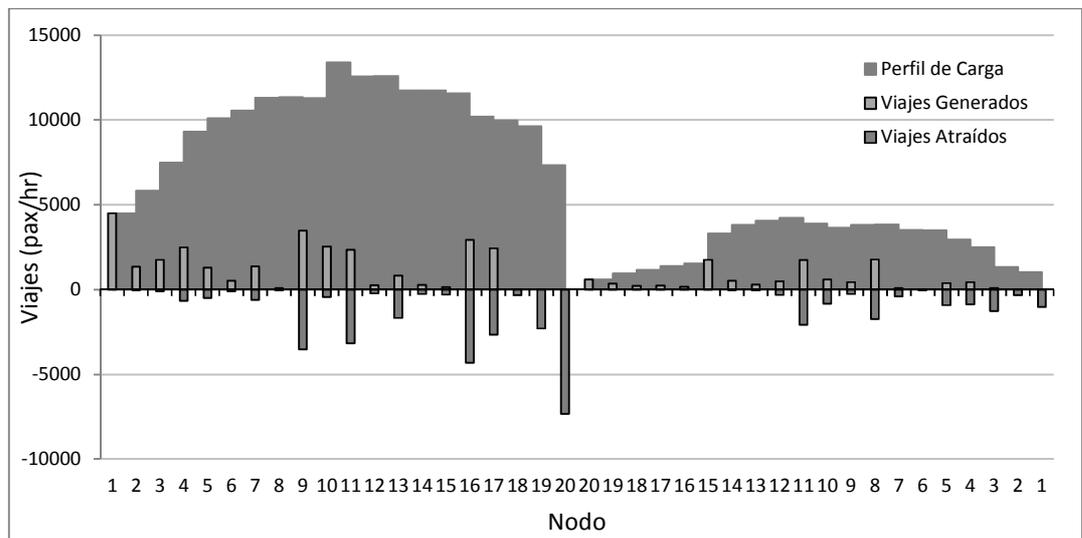


Figura 4-6: Perfil de carga matriz tipo 3

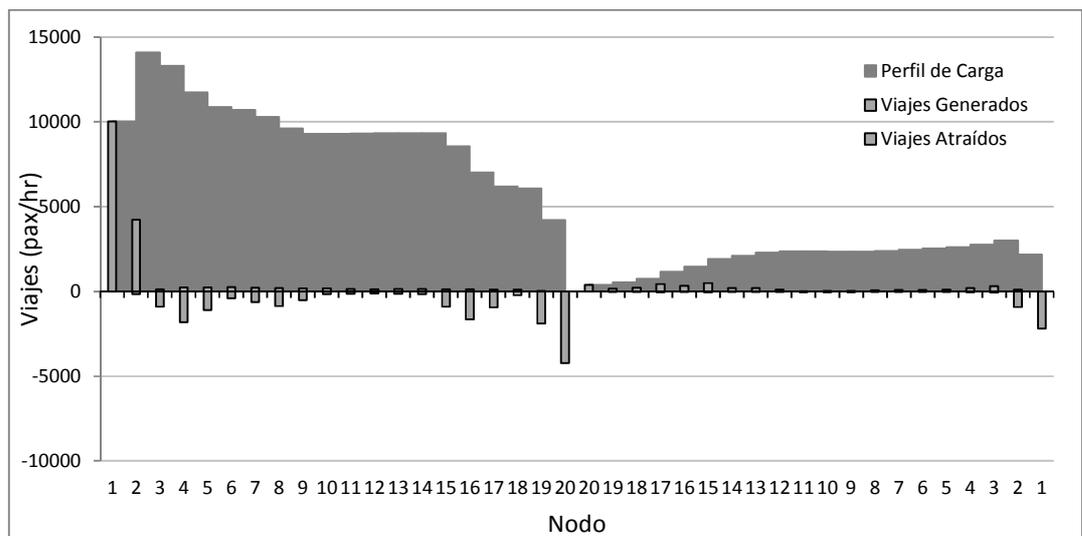


Figura 4-7: Perfil de carga matriz tipo 4

Finalmente, los gráficos presentados en las figuras 4-6 y 4-7 similares a las matrices 1 y 2 respectivamente, con sus viajes redistribuidos entre 20 paraderos por sentido en

lugar de 10. El largo del corredor para estos casos es de 7,6 km en total, con paraderos espaciados a 400 m, y una velocidad de operación de 25 km/hr.

La Tabla 4-2 a continuación muestra los diferentes atributos de cada una de estas cuatro matrices:

Tabla 4-2: Atributos de las matrices base

Matriz	Paradas	Tipo Perfil*	Total Viajes (pax/hr)	Desv. Est. (pax/hr)	Viajes Promedio (pax/hr)	Coef. Var.	Largo Promedio (paraderos)	Carga Crítica (pax/hr)
1	10	Grecia	37.728	507,0	419,2	1,2	3,2	13.392
2	10	Pajaritos	20.546	673,7	228,3	3,0	5,1	14.119
3	20	Grecia	38.744	205,9	102,0	2,0	6,3	13.400
4	20	Pajaritos	20.453	176,0	53,8	3,3	10,6	14.101

Como se puede ver, las cuatro matrices cubren una variedad de atributos de interés para el experimento. En la tabla los valores para la desviación estándar y el promedio de viajes son calculados a nivel de pares OD, y el coeficiente de variación se calcula como el cociente entre ambos. El largo promedio se presenta a nivel de paraderos, considerando un viaje entre dos paradas seguidas como de un paradero de largo. Para fines de este modelo tiene mucho más sentido considerar este indicador de distancia en lugar de la distancia física verdadera, sobre todo si los paraderos no son equidistantes, porque guarda directa relación con los ahorros potenciales de tiempo, mientras que la distancia física distorsiona los resultados al mezclar en el indicador una parte de los costos que se sabe que son fijos para el modelo. Como medida adicional de la concentración de los viajes se entrega, además del coeficiente de

variación, el valor del percentil 90 y este mismo valor normalizado con la media de los viajes.

4.3 Definición del indicador de desempeño

En esta sección se discute, en primer lugar, cuál indicador será el que se intentará explicar por medio de las distintas variables que representan cada escenario. Este indicador de desempeño definirá qué se entiende como un buen escenario para utilizar servicios expresos. Para este experimento fueron considerados cuatro indicadores de desempeño, de los cuales finalmente se optó por el que parecía más adecuado. Los tres indicadores considerados fueron los siguientes:

1. Ahorro total neto: Corresponde directamente a la diferencia entre los costos sociales de la instancia donde el corredor es atendido por un único servicio regular, y el escenario optimizado mediante la heurística propuesta.
2. Ahorro porcentual: Corresponde a la diferencia porcentual entre los términos comparados en el indicador anterior.
3. Ahorro porcentual corregido: Consiste en calcular la diferencia porcentual entre los costos corregidos de la instancia con un servicio regular, y la optimizada con servicios expresos. Como costos corregidos se entiende, al igual que en capítulos anteriores, los costos sociales menos los costos fijos por tiempos de viaje sin detenciones.

En la Figura 4-8 se grafican los tres indicadores para el total de escenarios, ordenados de acuerdo a su ahorro total neto.

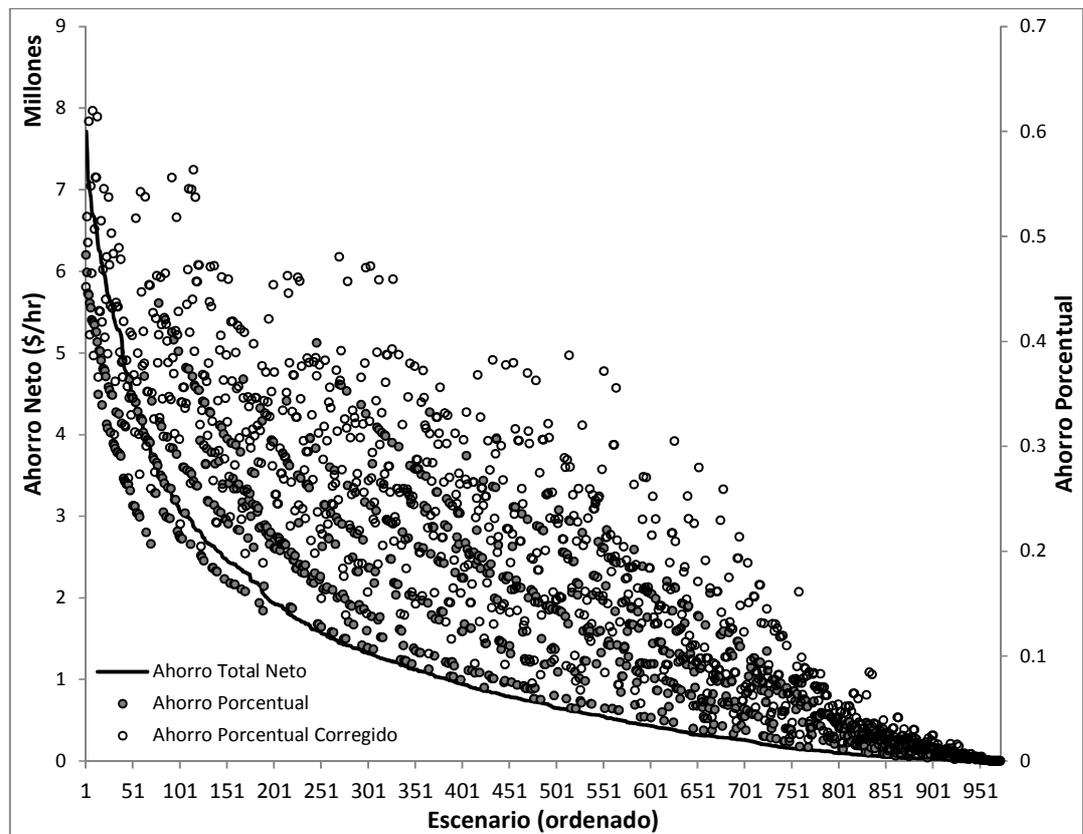


Figura 4-8: Comparación entre indicadores de ahorro social

Este gráfico permite notar que, a pesar de existir en términos generales alguna correlación entre los tres primeros indicadores (sobre todo en los escenarios donde los ahorros son menores), existe una diferencia importante entre éstos cuando los ahorros son mayores. Es por esto que se hace necesario analizar las ventajas y desventajas de cada uno de estos indicadores de desempeño para escoger el que mejor represente los beneficios de un escenario versus otro.

El primer indicador tiene la ventaja de ser muy simple y de entregar directamente una medida sobre el efecto en términos de costos sociales de la aplicación de los servicios expresos. Sin embargo, este indicador puede no ser el más indicado para

determinar el beneficio de los servicios expresos en sí ya que no da una referencia para decidir si los ahorros son importantes o no para el contexto del problema. Esto hace razonable pensar en buscar alguna forma de normalizar este indicador, usando algún tipo de referencia. Una vía para corregir este problema consiste en obtener un indicador adimensional tomando como punto de comparación los costos del caso base donde no se aplican servicios expresos. Dentro de los dos indicadores de este tipo que se presentan, el indicador corregido pareciera tener más sentido que el otro, ya que la parte constante de la función objetivo sólo introduce un costo fijo en el numerador y denominador disminuyendo el efecto real de los expresos en los escenarios en comparación.

Para entender mejor el último punto, consideremos el siguiente caso: un escenario donde los costos totales son de \$1.000 sin servicios expresos, y que se pueden hacer bajar a \$900 aplicando estos servicios. Supongamos además que los costos fijos por viaje (es decir, sin considerar detenciones) son de \$500. Esto significa que, para este escenario, el ahorro porcentual es de un 10%, y el ahorro porcentual corregido es de un 20%. Ahora supongamos que los tiempos de viaje suben al doble, lo que hace que los costos fijos suban a \$1.000, los totales sin expresos a \$1.500, y los totales con expresos a \$1.400. Con estos cambios, los ahorros porcentuales bajarían al 6,7%. Sin embargo, los ahorros corregidos se mantendrían en el 20%. Estas dos soluciones son equivalentes en términos de beneficios para los usuarios, como el indicador corregido efectivamente es capaz de identificar. Notar que tanto el ahorro neto como el ahorro per cápita también están libres de esta distorsión.

Además de los tres indicadores recién propuestos existen muchos otros indicadores parciales que pueden ser útiles para evaluar el desempeño de una red. Algunos de estos indicadores son la variación del costo de operación por kilómetro recorrido para el operador, el ahorro promedio de los usuarios, el ahorro per cápita (entendido como el ahorro neto dividido entre el número de usuarios), y la variación en el costo por kilómetro transportado, entre otros. Estos indicadores, sin embargo, no deben ser utilizados para medir el desempeño de los servicios de un corredor por sí solos ya que entregan sólo una visión parcial de lo que ocurre en este al no agregar los efectos sobre todos los involucrados. Además, poner el foco en mejorar indicadores de eficiencia puede ser engañoso: supongamos, por ejemplo, que estamos comparando dos escenarios, uno que genera un ahorro total de 1.000 beneficiando a 100 usuarios, con uno que genera un ahorro de 2.000 para 200 usuarios. Según un indicador como el ahorro promedio ambos escenarios serían equivalentes, lo que sería un error porque el beneficio social del segundo caso es mucho mayor.

4.4 Calibración del modelo de regresión

Los resultados de los 972 escenarios optimizados por el modelo fueron estudiados y ajustados a un modelo de regresión lineal, de forma de identificar la importancia de los distintos atributos de los corredores que permiten predecir el potencial de un escenario para la utilización de servicios expresos. En esta sección se muestra en secuencia las modelaciones calibradas para los datos, justificando las modificaciones efectuadas en cada etapa a la formulación propuesta.

- **Regresión 1**

La primera regresión que se aplicó a los datos de los escenarios consistió en incorporar los principales efectos ya conocidos de experiencias previas, en particular: el efecto del largo de los viajes, de la concentración de los viajes, y de la congestión. El primero de estos factores, el largo promedio de los viajes, ha mostrado ser en todos los experimentos realizados el predictor más relevante de los beneficios de incorporar servicios expresos. Esto se debe a que mientras más largos son los viajes de la red existen más paraderos por omitir, y que por lo tanto existe un mayor potencial beneficio por tiempo de viaje ahorrado. Sin embargo, dado que el ahorro potencial en tiempos de viaje puede ser conocido directamente a partir de las características de un corredor y de su matriz de demanda, tiene más sentido utilizar directamente este valor como indicador en lugar del largo promedio de viaje, que se puede calcular como:

$$PTTS = \theta_{tt} \tau \bar{S}k \sum_{w \in \mathcal{W}} T_w \quad (4.1)$$

En la expresión anterior se define $\bar{S}k$ como el número promedio de paraderos que atraviesan los viajes del sistema. Para el caso de un corredor con sus paraderos numerados secuencialmente, este valor se puede calcular como:

$$\bar{S}k = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^N T_{(i,j)} (j - i - 1)}{\sum_{w \in \mathcal{W}} T_w} \quad (4.2)$$

El indicador propuesto en (4.1) debería entregar una mejor medida del efecto del ahorro potencial de tiempos de viaje que el largo promedio de los viajes, debido a

que el tiempo de viaje entre paraderos consecutivos representa una constante para los usuarios y no tiene efecto en los ahorros, sólo en los costos de operación.

Un segundo factor que explica el efecto de los servicios expresos sobre un corredor, según se ha podido constatar en trabajos previos, es el nivel de concentración de los viajes sobre la red. El indicador propuesto en estudios previos para medir este efecto corresponde al coeficiente de variación de los viajes de la matriz, definido como:

$$CV = \frac{\sigma_w}{\overline{T_w}} \quad (4.3)$$

En la expresión anterior $\overline{T_w}$ corresponde al promedio del número de viajes en pares $w \in \mathcal{W}$, y σ_w representa la desviación estándar de estos viajes. Un mayor valor de CV implica que los viajes se encuentran menos distribuidos en la matriz, lo que equivale a decir que los viajes se encuentran concentrados en un menor número de pares.

Un tercer atributo que en las experiencias previas ha mostrado ser relevante corresponde al nivel de congestión del sistema. Una forma de determinar si un problema presenta congestión consiste en calcular su solución óptima sin imponer la restricción de capacidad, y ver si las cargas de esta solución respetan dicha restricción, y si no, por cuánto la exceden. Tomando como situación de referencia para la congestión la solución que no utiliza servicios expresos, el indicador que se propone para la congestión es el siguiente:

$$CI = \frac{P_a^*/cap}{f_a^*} \quad (4.4)$$

En la expresión anterior f_a^* es la frecuencia óptima sin congestión, que se puede obtener mediante la fórmula de la raíz (3.10). Por su parte, P_a^* corresponde a la

carga del corredor en el arco crítico, a^* . Para los análisis se asumirá que este valor es directamente proporcional al total de viajes en el sistema cuando el resto de los parámetros no varía, es decir, $P_{a^*} = k \sum_w T_w$. El valor mayor que uno del índice (4.5) implica que la restricción de capacidad define la solución para el caso sin expresos, y que por lo tanto el escenario se puede considerar como congestionado. Más aún, entre mayor este valor, mayor el nivel de congestión del escenario.

Con esto, el primer modelo a ajustar para los datos del experimento respondía al siguiente formato:

$$CPA_1 = \alpha_0 + \alpha_1 PTTS + \alpha_2 CV + \alpha_3 CI \quad (4.5)$$

En la ecuación anterior CPA_1 corresponde al ahorro porcentual corregido, como fue definido anteriormente, que el modelo 1 predice. Las unidades de $PTTS$ son \$·*pax/hr* mientras que tanto CV como CI son adimensionales. Resolviendo mediante mínimos cuadrados se obtienen los siguientes parámetros para el modelo:

Tabla 4-3: Parámetros del modelo 1

Parámetro	Valor	Estadístico t
α_0	-1,18E-01	-14,67
α_1	3,77E-08	30,96
α_2	8,54E-02	26,57
α_3	1,21E-02	5,31

Esta regresión presenta un R^2 de un 71,5%. Se aprecia que los signos de los parámetros corresponde a lo esperado de acuerdo a la experiencia previa: el beneficio crece con el ahorro potencial de los tiempos de viaje, con la concentración de los viajes, y con el nivel de congestión del sistema. Además, todos estos parámetros son

estadísticamente significativos. En el gráfico a continuación se presenta una comparación entre los valores a predecir versus los predichos por el modelo.

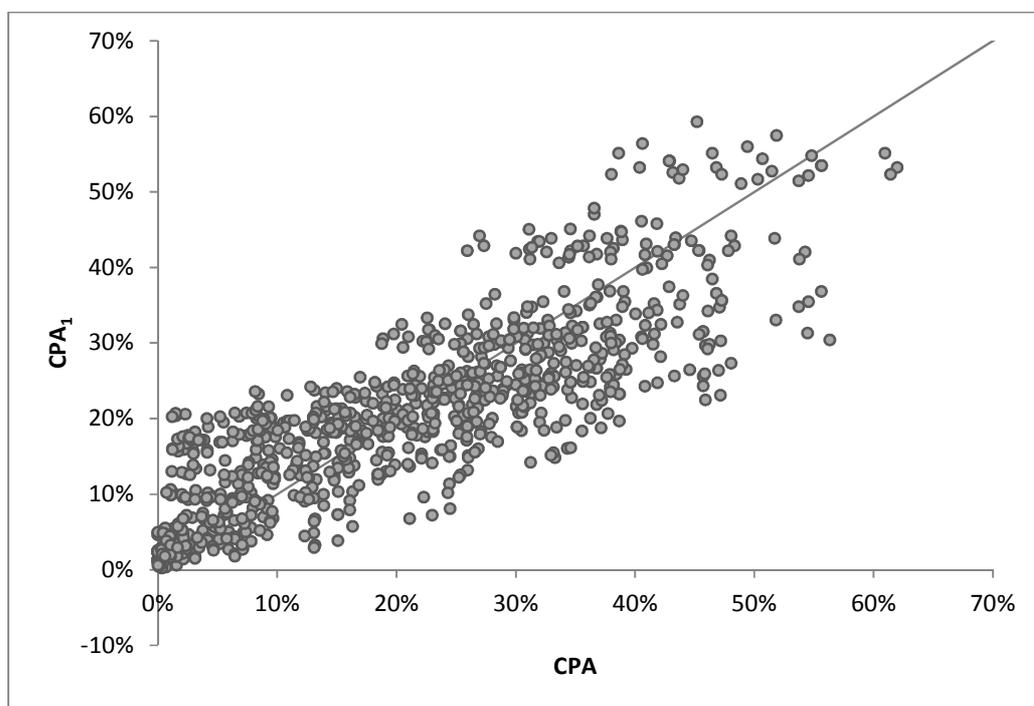


Figura 4-9: Regresión 1

Este modelo representa un buen punto de partida, al ser un modelo que utiliza pocas variables, con sentido físico, y con un buen nivel de ajuste.

- **Regresión 2**

A pesar de que el modelo anterior entrega un buen ajuste para los datos considerados, un análisis más detallado de los resultados permite ver la existencia de efectos no capturados todavía por este modelo. En particular, si graficamos estos mismos datos distinguiendo por el valor del parámetro c_L sucede lo siguiente:

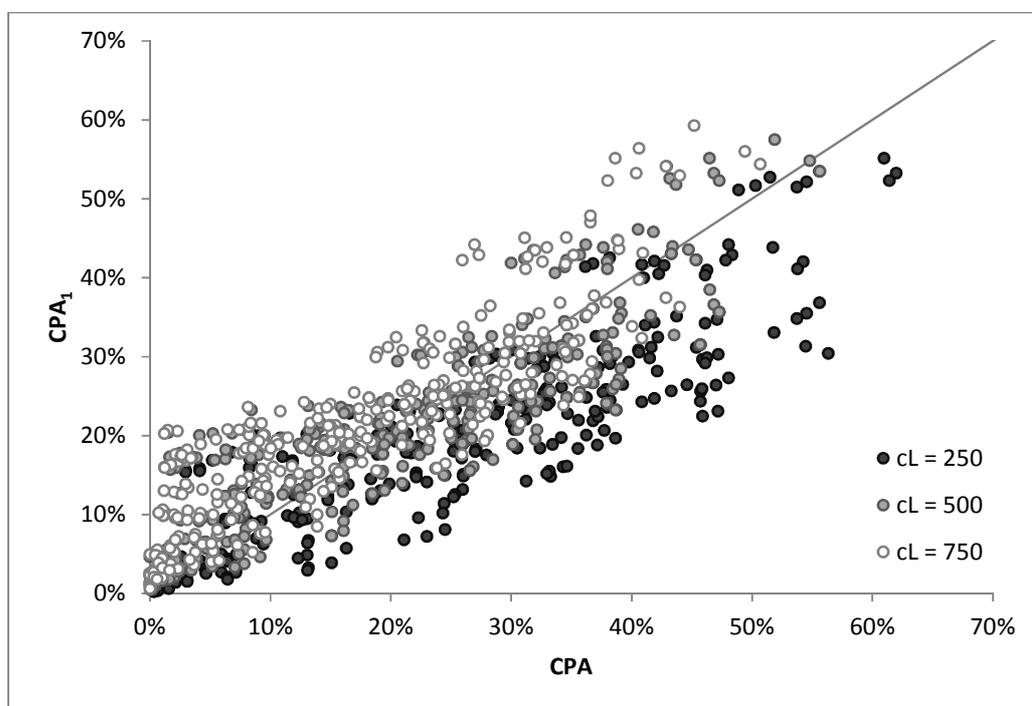


Figura 4-10: Desglose regresión 1

Como se aprecia claramente en la figura anterior, los puntos graficados en negro tienden a ser subestimados por el modelo 1 (ya que se encuentran en su mayoría por debajo de la diagonal de pendiente 1), mientras que con los puntos blancos sucede lo contrario. Esto significa que los costos de operación tienen un efecto para la predicción que debiese ser incorporado al modelo.

El parámetro c_L utilizado por simplicidad para el desglose de los valores de la Figura 4-10 representa parcialmente los costos del operador, pero es posible definir un indicador que contenga más información. Un indicador que fue probado con éxito corresponde a c_a , el costo de operación del servicio regular para el corredor. De acuerdo a (2.3), se puede observar que el costo de operación de un servicio c_l dependerá no sólo de los parámetros c_L y c_T (proporcionales entre sí para este

experimento) sino también del largo y de la duración del ciclo del servicio. De esta manera se aprovecha de incorporar en el modelo el efecto de la distancia y tiempo de viaje entre paraderos, que, como se discute anteriormente, no debiese afectar a los usuarios al representar una constante para el problema, pero sí puede afectar desde el punto de vista de la operación.

En vista de lo anterior, el nuevo modelo a calibrar tiene la siguiente estructura:

$$CPA_2 = \alpha_0 + \alpha_1 PTTS + \alpha_2 CV + \alpha_3 CI + \alpha_4 c_a \quad (4.6)$$

En esta regresión se trabaja con las mismas unidades que en la anterior. El costo de operación c_a se trabaja en \$/bus, de forma que al ser multiplicado por frecuencias de unidades de costo por tiempo. Los valores de los parámetros calibrados se muestran en la tabla a continuación:

Tabla 4-4: Parámetros del modelo 2

Parámetro	Valor	Estadístico t
α_0	-4,14E-02	-4,95
α_1	3,77E-08	35,16
α_2	8,19E-02	28,90
α_3	2,54E-02	11,84
α_4	-8,30E-06	-16,89

El ajuste de esta versión del modelo es mejor que la anterior, con un R^2 de un 78,0%, aunque podría deberse únicamente a la utilización de un mayor número de variables explicativas. Se puede corroborar también que los signos de los parámetros α_1 , α_2 y α_3 son positivos, de acuerdo a lo esperado para el modelo. El signo de α_4 podía ser predicho a partir de la Figura 4-10, donde se puede ver como el modelo 1 tendía a sobreestimar los beneficios para los escenarios con mayores costos de operación. Los

estadísticos t corroboran además que todas las cuatro variables explicativas son significativas. Un gráfico de los valores predichos por el modelo 2 se presenta en la Figura 4-11.

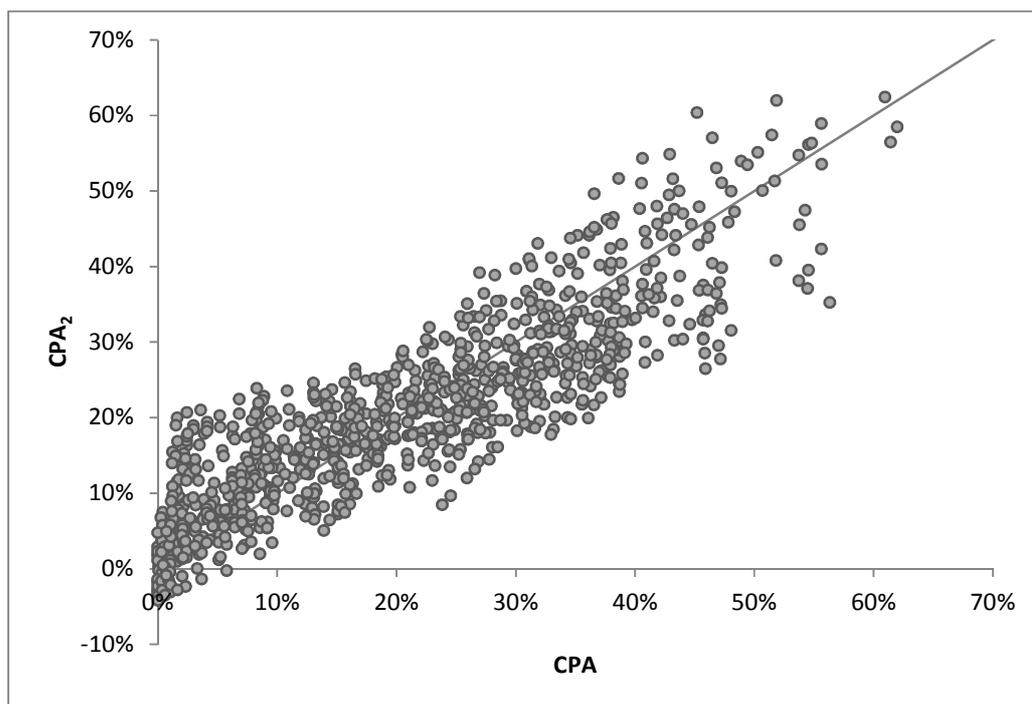


Figura 4-11: Regresión 2

Este modelo entrega una buena predicción de los beneficios para los escenarios, ya que no sólo entrega un buen ajuste, sino también abarca toda la información relevante de los escenarios (directa o indirectamente) en sus variables explicativas. Cabe notar que cuando los beneficios son bajos, a veces el modelo predice valores negativos, lo que no tiene mayor importancia, debido a que efectivamente son escenarios con poco potencial para los servicios expresos.

- **Regresión 3**

A partir del modelo anterior, se efectuó una revisión de las diferentes variables explicativas de los escenarios para detectar posibles mejoras a la modelación. De esta forma se pudo observar que todavía existe una tendencia asociada al tiempo de detención τ que la regresión no capta adecuadamente, como se muestra en el gráfico a continuación.

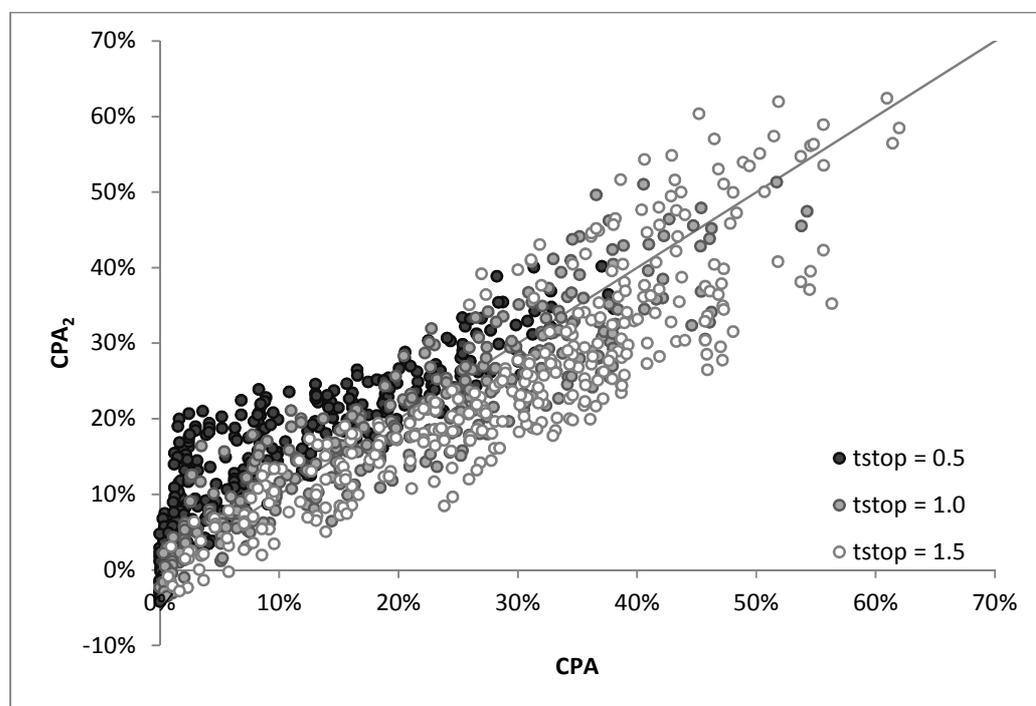


Figura 4-12: Desglose regresión 2

Del gráfico se desprende que, a pesar de que los tiempos de detención son utilizados para determinar los ahorros potenciales por tiempos de viaje, su efecto predictivo es más fuerte, y pueden ser considerados por separado en el modelo. De acuerdo a esto, se definió el tercer modelo a calibrar como:

$$CPA_3 = \alpha_0 + \alpha_1 PPTS + \alpha_2 CV + \alpha_3 CI + \alpha_4 c_a + \alpha_5 \tau \quad (4.7)$$

En la expresión anterior el parámetro τ se entrega en minutos, y los demás parámetros conservan las unidades introducidas anteriormente. Los parámetros ajustados para el modelo se presentan en la tabla que sigue.

Tabla 4-5: Parámetros del modelo 3

Parámetro	Valor	Estadístico t
α_0	-1,41E-01	-18,67
α_1	2,58E-08	27,15
α_2	8,41E-02	38,39
α_3	3,24E-02	19,24
α_4	-9,35E-06	-24,48
α_5	1,21E-01	25,56

El ajuste de este tercer modelo es de un 86,9%. Los signos de todos los parámetros se conciden con los resultados anteriores y son estadísticamente significativos. El gráfico a continuación muestra los resultados de esta tercera regresión.

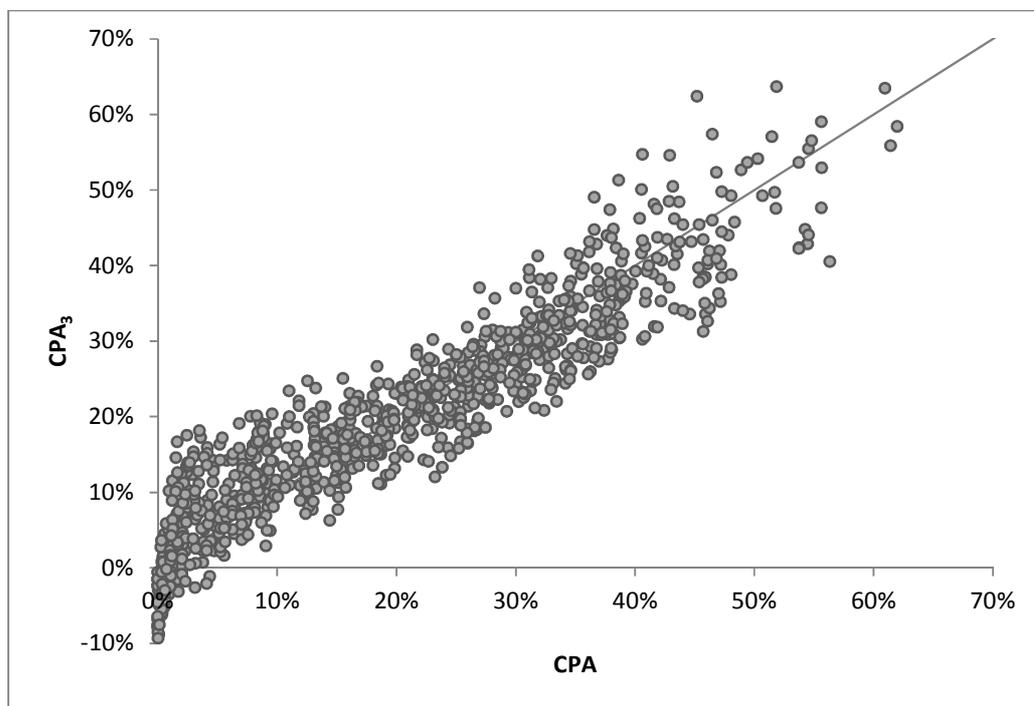


Figura 4-13: Regresión 3

4.5 Análisis de los resultados

A partir de la última regresión se puede observar que el beneficio que los servicios expresos pueden entregar en un determinado escenario puede ser predicho a partir de cinco variables explicativas:

1. El ahorro potencial en tiempos de viaje del corredor.
2. La concentración de los viajes.
3. El nivel de congestión del sistema.
4. Los costos de operación del servicio regular.
5. El tiempo de detención en paradero de los buses.

Estos factores pueden ser además descompuestos para determinar el efecto de otras variables del problema.

El modelo 3 nos permite concluir que los servicios expresos reportan mayores beneficios cuando el ahorro potencial en tiempos de viaje es mayor. Observando la fórmula (4.1) se puede concluir que estos ahorros son a su vez mayores cuando:

- El valor del tiempo de viaje en vehículo es mayor.
- El tiempo de detención en paradero de los buses es mayor. Este factor también influye por sí sólo en la regresión, pero su efecto es el mismo en ambos lados: entre mayor el tiempo de detención, mayor el incentivo a usar servicios expresos.
- El número de paraderos por viaje es mayor. Esto se puede también aproximar en algunos casos como el largo promedio de los viajes, ya que la correlación entre estas dos variables es alta. Otra variable que también se encuentra normalmente correlacionada con este valor es el número total de paraderos a lo largo del corredor, ya que un mayor número de éstos suele implicar que los viajes también atraviesan más paraderos.
- El total de viajes del sistema es mayor. Este factor, en todo caso, también tiene un efecto en el índice de congestión en sentido contrario a éste, por lo que su efecto neto será discutido más adelante.

Con respecto al nivel de concentración de los viajes, se pudo confirmar que el coeficiente de variación captura adecuadamente este efecto y que es un buen predictor del desempeño de los servicios expresos.

Con respecto al nivel de congestión, se pudo mostrar que mientras más congestionado un escenario, mayores los beneficios potenciales que ofrecen los servicios expresos. Tomando el indicador de congestión (4.4), asumiendo que el

perfil máximo es proporcional a los flujos, y reemplazando en él la fórmula de la raíz (3.10) obtenemos que:

$$CI = \frac{k}{cap} \sqrt{\frac{c_a \sum_w T_w}{\theta_{wt} \lambda}} \quad (4.8)$$

De esta expresión podemos concluir que los beneficios asociados a operar con servicios expresos son mayores cuando:

- La carga del arco crítico con respecto al total de viajes es mayor.
- La capacidad de los buses es menor, lo que claramente equivale a un mayor nivel de congestión en los buses.
- El valor del tiempo de espera multiplicado por λ es menor.
- Los costos del operador son mayores. Este efecto contradice al que poseen los costos de operador como variable explicativa por sí sola en la regresión, por lo que su efecto neto será analizado más adelante.
- El total de viajes es menor.

Debido a que el efecto del número total de viajes en el factor de congestión es contrario al que se percibe en relación al ahorro potencial, mediremos el efecto neto de esta variable sobre el beneficio predicho por la regresión. Con este fin, derivamos la expresión (4.7) con respecto al flujo total, que llamaremos T , para simplificar la notación.

$$\frac{\delta CPA_3}{\delta T} = \alpha_1 \frac{\delta PTTS}{\delta T} + \alpha_2 \frac{\delta CV}{\delta T} + \alpha_3 \frac{\delta CI}{\delta T} + \alpha_4 \frac{\delta c_a}{\delta T} + \alpha_5 \frac{\delta \tau}{\delta T} \quad (4.9)$$

En la expresión anterior los términos asociados a α_2 , α_4 y α_5 son independientes de T y por lo tanto sus derivadas son nulas. De la expresión (4.1) se puede desprender que:

$$\frac{\delta PTTs}{\delta T} = \theta_{tt} \tau \bar{S} k \quad (4.10)$$

Por su parte, derivando la expresión (4.8) se obtiene que:

$$\frac{\delta CI}{\delta T} = \frac{k}{2cap} \sqrt{\frac{c_a}{\theta_{wt} \lambda T}} \quad (4.11)$$

Con esto, la derivada del beneficio respecto a T será:

$$\frac{\delta CPA_3}{\delta T} = \alpha_1 \theta_{tt} \tau \bar{S} k + \alpha_3 \frac{k}{2cap} \sqrt{\frac{c_a}{\theta_{wt} \lambda T}} \quad (4.12)$$

La expresión anterior es positiva para cualquier valor de T , lo que implica que, de acuerdo al modelo 3, un aumento en el flujo total del corredor hace más atractiva la implementación de los servicios expresos.

Con respecto al costo de operación del servicio regular, su efecto neto también se debe calcular considerando que aparece en más de un lugar de la fórmula. Evaluamos la derivada del beneficio en función de estos costos:

$$\frac{\delta CPA_3}{\delta c_a} = \alpha_1 \frac{\delta PTTs}{\delta c_a} + \alpha_2 \frac{\delta CV}{\delta c_a} + \alpha_3 \frac{\delta CI}{\delta c_a} + \alpha_4 \frac{\delta c_a}{\delta c_a} + \alpha_5 \frac{\delta \tau}{\delta c_a} \quad (4.13)$$

Desarrollando los términos se obtiene que:

$$\frac{\delta CPA_3}{\delta c_a} = \alpha_3 \frac{P_{a^*}}{2cap \sqrt{c_a \theta_{wt} \lambda \sum_w T_w}} + \alpha_4 \quad (4.14)$$

Dado que α_4 es negativo, el valor de esta derivada dependerá del valor que tomen el resto de las variables. Evaluando esta expresión en los 972 escenarios, se puede

observar que, a pesar de que en la mayoría de los casos esta derivada es negativa (es decir, un aumento en los costos de operación va en desmedro del desempeño de los expresos), existen seis escenarios donde esta derivada se vuelve positiva. Esto implica que el efecto de los costos de operación sobre los beneficios de los servicios expresos dependerá del escenario analizado.

Observando la fórmula (2.3) se puede observar que los costos de operación del servicio regular c_a dependen de los costos de operación por distancia y por tiempo, y del largo y tiempo de ciclo de dicho servicio. Esto implica que estas variables tendrán también el mismo efecto que c_a en los beneficios de los servicios expresos, que deben ser estudiados para cada caso aplicando la fórmula (4.14).

Los resultados acá expuestos sirven para tener una idea del potencial que poseen los servicios expresos como medida de gestión bajo diferentes escenarios. Una forma de predecir este desempeño consistiría directamente en aplicar la fórmula (4.7). Sin embargo, para que las predicciones hechas por la fórmula sean válidas, los valores de los parámetros deben encontrarse razonablemente cerca de los valores utilizados en el experimento, de otra forma la extrapolación podría perder validez. De todas formas, y a modo de resumen, el experimento realizado en este capítulo permitió concluir que los servicios expresos resultan más atractivos cuando:

- El tiempo de detención en paradero de los buses es mayor.
- El número de paraderos por viaje (o el largo promedio de los viajes) es mayor.
- El total de viajes en el sistema es mayor.
- El valor del tiempo de viaje es mayor.
- Los viajes se encuentran más concentrados en ciertos pares.

- La carga del arco crítico, medida en proporción al total de viajes, es mayor.
- La capacidad de los vehículos es menor.
- El valor del tiempo de espera es menor.

Resultaría interesante, como futura línea de investigación, intentar respaldar estos resultados de forma teórica, a partir de expresiones analíticas para los costos sociales óptimos expresados en función de los parámetros recién listados. Una primera aproximación podría ser trabajar con un corredor simplificado, donde los usuarios se comporten de forma sencilla, y con pocos servicios de formato que favorezca a la optimización, como los presentados en el Capítulo 3.

5. DISEÑO DE SERVICIOS EXPRESOS SOBRE UNA RED

En este capítulo se presenta un algoritmo para el diseño de servicios expresos en una red. Éste es obtenido adaptando la heurística para corredores presentada al final del Capítulo 3. A continuación se discute cómo se hace esta adaptación, y en la sección siguiente se muestra un ejemplo de implementación del método de diseño sugerido.

5.1 Algoritmo de diseño de servicios expresos para una red

En un sentido general, el problema de diseño de servicios expresos para una red es similar al problema de diseño en un corredor, si consideramos que la topología de la red es capturada por el modelo a través de una matriz de distancias que es sencilla de construir en ambos casos. Sin embargo, existen algunas particularidades del problema en red que lo hacen más complejo. Si separamos el problema en las dos etapas propuestas en el Capítulo 3, es decir, diseño de servicios y optimización de frecuencias, podemos diferenciar los desafíos que implica el problema para una red en cada caso.

Con respecto al problema de optimización de frecuencias, el problema puede crecer en complejidad debido a diversos factores. El primero de estos factores consiste en que trabajar con una red por lo general obliga a tomar en cuenta un número mayor de nodos. Esto implica un aumento en el número de variables del problema de optimización, que resulta crítico particularmente en el caso de las variables para modelar los trasbordos. Un mayor número de paraderos también puede implicar la necesidad de trabajar con un mayor número de servicios (sin importar, por ahora, de dónde éstos son obtenidos) lo que también aumenta el número de variables de

decisión y de óptimos locales existentes para el problema. Esto último se explica tomando en cuenta que el problema de optimización de frecuencias debe decidir a cuáles servicios entregar frecuencias positivas, y a cuáles de éstos resulta conveniente entregar frecuencia nula. Entre más servicios se consideran, probablemente existirá también un mayor número de subconjuntos de servicios que entreguen una solución mínima local al problema, lo que hará más difícil su resolución.

Por otra parte, esta misma complejidad del sistema hace que el comportamiento de los usuarios pueda ser más complicado de modelar, y al mismo tiempo más relevante para la optimización. En un corredor aislado, por ejemplo, proveer un único servicio regular asegura que exista una conexión directa entre cualquier origen y destino, lo que hace menos crítica la necesidad de incorporar los trasbordos en la modelación. En una red, sin embargo, es esperable que en la mayor parte de los casos existan viajes que se vean forzados a realizar trasbordos. Más aún, en un problema sobre un corredor es posible que un mismo viaje pueda tener distintos paraderos como candidatos a punto de trasbordo, pero podemos asegurar que el viaje siempre se realizará por sobre el único corredor en consideración. En el caso de red, en cambio, existen viajes que se pueden realizar por diferentes partes de la red. Para poder modelar adecuadamente este tipo de escenarios un comportamiento de usuarios a rutas mínimas podría a veces ser insuficiente, y podría resultar de utilidad considerar alguna de las variantes para el comportamiento de usuarios propuestas en el Capítulo 3. Sin embargo, es importante reconocer que las heurísticas de dicho capítulo asumen que la demanda T_w por el corredor es conocida. En el caso de una red, no es posible

conocer la demanda por corredor, porque para los viajes que no están contenidos íntegramente en un corredor ésta dependerá de la oferta de servicios y la elección de ruta por parte de los usuarios.

A pesar de las dificultades acá expuestas, el problema de optimización de frecuencias para una red, conocido un conjunto de servicios definidos previamente por el modelador, puede ser abordado con un enfoque prácticamente idéntico al utilizado en el caso para un corredor aislado, cuidando que la complejidad del problema no se vuelva excesiva. Esto se puede lograr implementando las técnicas descritas en el Capítulo 3 para lidiar con trasbordos, específicamente la limitación del número puntos de trasbordo habilitados, la codificación de servicios no cíclicos (es decir, la utilización de restricciones de continuidad de frecuencias) y la separación de las etapas de asignación y de optimización de frecuencias para ser resueltas en forma secuencial e iterativa.

Con respecto al diseño de servicios, el problema se vuelve considerablemente más complejo. De partida, el problema de diseño sin utilizar servicios expresos deja de ser trivial, ya que los corredores pueden ser definidos de múltiples formas en una red, lo que lleva a diferentes soluciones, y el número de corredores puede ser muy alto como para simplemente alimentar al modelo de todos ellos. Incluso limitando los posibles nodos extremos para la definición de los corredores las combinaciones de posibles terminales del corredor se dispara con el número de nodos, y además cada par de terminales puede ser conectado por varias rutas. Si suponemos que se conoce una configuración razonable para los corredores a utilizar, el problema de optimizar las frecuencias de los servicios regulares que operan sobre ellos tampoco deja de ser

complejo: las frecuencias óptimas para dichos servicios no se pueden obtener directamente a partir de una fórmula sencilla como la que se entrega en (3.10), por causa de la probable existencia de traslapes entre servicios, rutas alternativas para distintos viajeros, y trasbordos obligados.

Resulta natural intentar abordar este problema aplicando alguna variante de la heurística de diseño de servicios para un corredor. Para hacer esto posible, es necesario tener alguna definición previa sobre cuáles serán los corredores a considerar para la modelación, de forma de aplicar la heurística en cada uno de ellos por separado. Sin embargo, este enfoque posee ciertas limitaciones. Por una parte, dado que los servicios a generar se encontrarán contenidos sólo dentro de los corredores a considerar, el modelador se ve enfrentado a descartar de entrada la posibilidad de ofrecer una conexión directa entre ciertos pares, o entregar al modelo un número alto de corredores, lo que se traduce en un aumento en la complejidad del problema, dificultando su solución.

Definido un conjunto de trazados iniciales a partir de los cuales se aislarán los corredores (que equivalen a posibles trazados de servicios regulares en un problema sin expresos), el enfoque que se propone para enfrentar el problema de generación de servicios consiste en, partiendo de una solución inicial para una red, tomar un trazado de la red y aislarlo como corredor, y aplicar sobre éste las heurísticas de generación. En términos más precisos, aislar un trazado implica dos cosas: primero, identificar cuáles viajes de la solución vigente son efectuados en servicios contenidos por el trazado, y, segundo, encontrar las condiciones que deben cumplir las

frecuencias de los servicios (vigentes y por generar) sobre el trazado para no violar la continuidad de frecuencias con el resto de los servicios de la red.

Con esto, el problema de diseño de servicios para una red puede ser resuelto iterando entre la optimización de las frecuencias de la red completa, y la generación de servicios para corredores aislados. Cabe notar que, de forma similar a lo que sucede con el enfoque binivel para la optimización de frecuencias explicado en el Capítulo 3, en la etapa de generación para un corredor aislado los servicios son diseñados asumiendo que la asignación de viajes al corredor no se verá modificada, lo que en realidad al optimizar a nivel de red no será así. Sin embargo, esto no es un problema, ya que la generación de servicios para asignación fija baja los costos totales del sistema, y la posterior liberación de la asignación de los usuarios disminuye estos costos todavía más. Esto asegura que cada paso de la iteración avanza en la reducción de los costos totales, lo que a su vez garantiza la convergencia del algoritmo.

Otra simplificación que se propone en el Capítulo 2 y que puede resultar crucial para reducir la complejidad del problema de red a un nivel tratable, corresponde a la limitación de los puntos de trasbordo del sistema. Para un correcto funcionamiento del algoritmo, esta elección se debe hacer cuidando que estos puntos tengan sentido para los trazados previamente definidos. Como regla general, debiesen ser puntos de trasbordo, como mínimo, todos los nodos donde dos trazados se intersectan, y los puntos donde comienza o termina un traslape entre trazados.

Como una última simplificación del problema, que facilita en gran medida el tratamiento de las restricciones de continuidad de frecuencias, se asumirá que

paraderos de ambos sentidos coinciden geográficamente y que pueden ser representados por un único nodo. Sin embargo, al momento de aplicar la heurística de generación de servicios se deberá adaptar el corredor separando los nodos de ida y de vuelta, debido a la forma en que está concebido el algoritmo. Considerando todos los supuestos y precisiones expuestos, el algoritmo para la resolución del problema de diseño de servicios expresos en red se detalla a continuación:

Algoritmo para el diseño de servicios expresos en una red:

1. Definir un conjunto de trazados, \mathcal{T} , que cubran la red completa. Estos trazados pueden poseer traslapes, y pueden ser, por ejemplo, inspirados en los que operan en la red actual en la práctica. Construir el conjunto de líneas \mathcal{L} que contiene a todos los servicios regulares para cada trazado, separados por sentido.
2. Utilizar el modelo de optimización de frecuencias sobre los servicios ya generados para la red completa, sin generación de nuevos servicios, y sin considerar restricción de capacidad de los buses. Para esto se puede aplicar directamente el modelo de Leiva, cuidando de que los tiempos de viaje entre paraderos por servicio capturen adecuadamente la topología de la red. En caso de que el problema sea muy grande, es posible aplicar alguna de las simplificaciones propuestas en el Capítulo 2, especialmente la de limitación de los puntos de trasbordo, y el enfoque binivel de solución para separar la asignación de viajes de la optimización de frecuencias.

3. A partir de la solución actual, seleccionar un trazado $t \in \mathcal{T}$. Construir para este corredor una matriz de demanda T_w^t con $w \in \mathcal{W}_t$, donde \mathcal{W}_t es el conjunto de pares contenidos en el trazado t . Esta matriz incluirá a todas las etapas de viaje de la solución actual que son realizados íntegramente en algún servicio que esté contenido por el trazado, es decir:

$$T_w^t = \frac{\sum_{l \in \mathcal{L}} f_l^w \delta_l^t \delta_l^w}{\sum_{l \in \mathcal{L}} f_l^w \delta_l^w} \sum_{s \in \mathcal{W}} V_w^s, \quad \forall w \in \mathcal{W}_t \quad (5.1)$$

En la expresión anterior, el parámetro δ_l^t indica si el servicio l pertenece al trazado t . Un servicio pertenecerá a un trazado sólo si éste se encuentra contenido por completo en éste, aunque un servicio puede pertenecer a más de un trazado. En la expresión anterior se utilizan las frecuencias auxiliares f_l^w que, como ya se ha explicado en el Capítulo 2, equivalen a la frecuencia nominal f_l del servicio l cuando éste es atractivo para viajar en la sección w , y cero cuando no lo es.

4. Calcular las frecuencias exógenas de cada nodo del trazado, $i \in \mathcal{N}_t$, definidas como:

$$F_i^t = \sum_{l \in \mathcal{L}} \vartheta_{li}^+ f_l (1 - \delta_l^t) - \sum_{l \in \mathcal{L}} \vartheta_{li}^- f_l (1 - \delta_l^t) \quad (5.2)$$

El término ϑ_{li}^+ toma valor unitario cuando el servicio l ingresa al trazado t o comienza su recorrido en el nodo i , mientras que ϑ_{li}^- lo hace cuando el servicio l sale del trazado t o termina su recorrido en el nodo i .

5. Aplicar la heurística de generación de servicios en corredores al corredor que coincide con el trazado t , considerando la matriz de demanda T_w^t , adaptada

para el corredor codificado con nodos de ida y de vuelta, y considerando capacidad ilimitada. Para la heurística de generación se debe considerar el problema separado por sentidos. En la formulación de este problema se debe reemplazar la restricción de continuidad de frecuencias (2.42) por:

$$\sum_{l \in \mathcal{L}} \vartheta_{li}^+ f_l - \sum_{l \in \mathcal{L}} \vartheta_{li}^- f_l + F_i^c = 0, \quad \forall i \in \mathcal{N}_1^c \quad (5.3)$$

Cabe recalcar que en esta última expresión se trabaja con nodos separados por ida y vuelta, de acuerdo a la notación con la que funcionan los algoritmos de generación del Capítulo 3. El conjunto \mathcal{N}_1^t representa a los paraderos en un sentido del corredor.

6. Comparar el resultado obtenido en el paso anterior con el costo total de asignar estos flujos sobre los servicios actualmente ofrecidos para el corredor. Si los servicios generados ofrecen una mejor solución que los servicios actuales, actualizar el conjunto \mathcal{L} , cambiando los servicios actuales del trazado t por los servicios generados en el paso 5.
7. Optimizar las frecuencias para la red con los nuevos servicios, sin considerar restricción de capacidad. Volver al paso 3 seleccionando un nuevo trazado, hasta que no sea posible encontrar una mejora a la solución actual para ninguno de los trazados en \mathcal{T} , o hasta cumplir con algún otro criterio de convergencia.
8. Ajustar las frecuencias de forma de satisfacer la restricción de capacidad, aplicando la heurística HC_2 .

Como se explicaba anteriormente, dado que la función objetivo nunca puede subir en una iteración, el algoritmo converge a una solución de equilibrio. Sin embargo, para

que el algoritmo funcione correctamente, es necesario verificar en cada subproblema que la solución obtenida no sea peor que la con la que se cuenta actualmente. Esto se debe a que, a pesar de que en teoría la solución óptima de cada nuevo problema no puede ser peor que la anterior, puesto que siempre la solución actual está contenida en el nuevo dominio, el problema cuenta con múltiples óptimos locales y es posible que algunas instancias los métodos de solución utilizados para resolver los subproblemas arrojen soluciones subóptimas.

Debido a que el modelo de optimización de frecuencias es invocado en repetidas ocasiones durante la ejecución del algoritmo, se recomienda utilizar sus variantes más eficientes en términos de tiempo de ejecución. En particular, se recomienda limitar el número de trasbordos a los puntos estratégicos de la red, y separar las etapas de asignación y optimización. Como se discute en el Capítulo 3, en la etapa de asignación es posible modelar el comportamiento de los usuarios como una minimización de rutas o de estrategias, lo cual queda a criterio del modelador, aunque en general rutas mínimas debiera ser más rápido de calcular y representa una aproximación conservadora para los costos de los usuarios.

Para la generación de servicios, y de acuerdo a lo que se concluye de los experimentos del Capítulo 3, para un problema de estas características resulta conveniente utilizar las heurísticas de generación que reportan beneficios sin necesidad de realizar iteraciones, es decir, HG_{1a} (generación de líneas iniciales, indicador simple), HG_2 (generación de servicios cortos) y HG_3 (generación de servicios zonales).

Con el fin de mantener la complejidad del problema de optimización de frecuencias a un nivel tratable, resulta conveniente limitar el número máximo de servicios a considerar en cada iteración, e imponer una frecuencia mínima a los servicios a considerar. Una solución que respete estas condiciones se puede obtener simplemente eliminando iterativamente los servicios de menor frecuencia, volviendo a optimizar las frecuencias en cada paso. Alternativamente, podría incluirse en esta etapa una verificación del efecto de aumentar drásticamente la frecuencia del servicio antes de eliminarlo, ya que a veces los servicios expresos sólo muestran su potencial cuando son suficientemente frecuentes como para que los usuarios los esperen para utilizarlos. Este efecto se puede apreciar en el ejemplo de multiplicidad de óptimos locales que se entrega en el Capítulo 2.

Como se explica en el paso 5 del algoritmo, es necesario hacer una conversión de notación entre los pasos de optimización de frecuencias de la red y de generación de servicios. Esta conversión consiste en pasar de la notación donde los paraderos de ambos sentidos son representados por nodos separados que utilizada en la etapa de generación de servicios para corredores, a la notación donde ambos sentidos son representados por un único nodo, utilizada para la optimización de frecuencias sobre la red. Esta conversión puede representar una pequeña molestia, pero puede ser fácilmente automatizada. Además, permite eliminar del análisis todos los nodos externos al corredor en la etapa de generación. Por último, esta conversión también se hace necesaria debido a que la notación binaria para los servicios en corredores no es suficiente para representar los servicios en red, donde además se debe indicar orden. En resumidas cuentas, esta conversión de notación es necesaria debido a que

estas diferentes etapas trabajan una con una red y la otra con un corredor, y para cada caso resulta conveniente utilizar una notación distinta.

El problema con capacidad se aborda en esta heurística de la misma forma que en el modelo base de optimización de frecuencias: se obtiene primero una buena solución para el problema sin capacidad, y a partir de esta solución se busca una nueva que satisfaga la restricción de capacidad. De acuerdo a las pruebas realizadas en el Capítulo 3, la heurística HC_2 representa un buen equilibrio entre velocidad de ejecución y calidad de soluciones obtenidas y resulta adecuada para la solución de este problema. Este enfoque posee las mismas limitaciones ya mencionadas en el Capítulo 2, es decir, puede ser costoso al implicar varias soluciones del problema de optimización de frecuencias, y que la solución que se obtiene finalmente es una aproximación al óptimo.

No se recomienda ajustar la capacidad en los pasos intermedios del algoritmo debido a que, como el paso 3 no considera que los pasajeros se pueden reasignar en el resto de la red, el resultado de este paso puede ser infactible en términos de capacidad en caso de atraer viajes de otros corredores. Además, al aplicar la heurística de capacidad en pasos intermedios del algoritmo puede generar subidas en los costos entre iteraciones seguidas, lo que dificulta la convergencia del algoritmo.

A continuación se expone un ejemplo donde se muestra esta heurística en acción sobre una red de quince paraderos, consistente en tres corredores que confluyen en un punto central de la red.

5.2 Aplicación del algoritmo

En esta sección se presenta un ejemplo de implementación del algoritmo propuesto para el problema en una red de transporte público. La red a estudiar se presenta en la Figura 5-1:

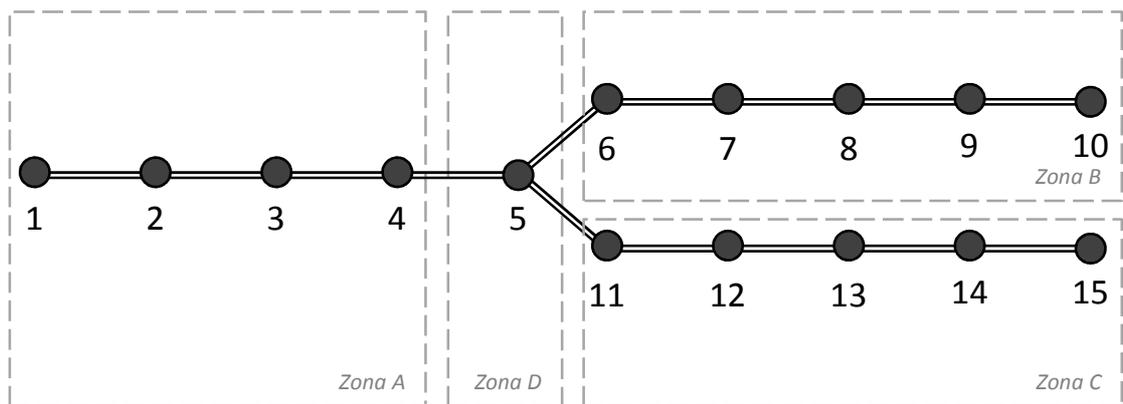


Figura 5-1: Esquema de la red a estudiar

Como se puede apreciar en la figura, esta red consta de 15 paraderos, organizados en tres tramos (zonas A, B y C) que confluyen en un nodo central (nodo 5, o zona D). La matriz de viajes a optimizar sobre esta red fue construida agregando para los diferentes corredores la matriz de demanda del corredor Pajaritos utilizada en los experimentos del Capítulo 3, con algunas modificaciones adicionales. Los principales movimientos de flujo que ocurren sobre la red se muestran en forma agregada en el esquema de la Figura 5-2 y en la Tabla 5-1.

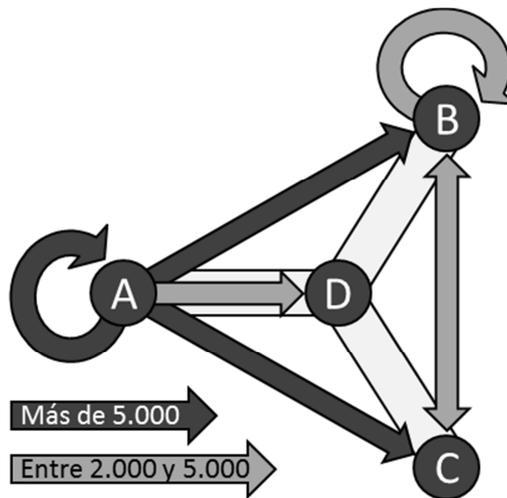


Figura 5-2: Principales movimientos de la red a estudiar

Tabla 5-1: Matriz de viajes agregada por zona de la red a estudiar

	A	B	C	D
A	6.341	5.045	6.614	2.132
B	1.309	2.645	3.058	917
C	669	2.891	1.696	226
D	69	1.375	1.343	0

Como se desprende de la Figura 5-2, existe una gran cantidad de viajes que se generan en la zona A y que se distribuyen por toda la red, como podría ser el caso de los viajes en punta tarde de una ciudad donde la actividad comercial se desarrolla en un sector específico (zona A) y donde los habitantes están distribuidos por toda la ciudad.

Las próximas figuras ilustran el perfil de carga para las zonas A, B y C de la red.

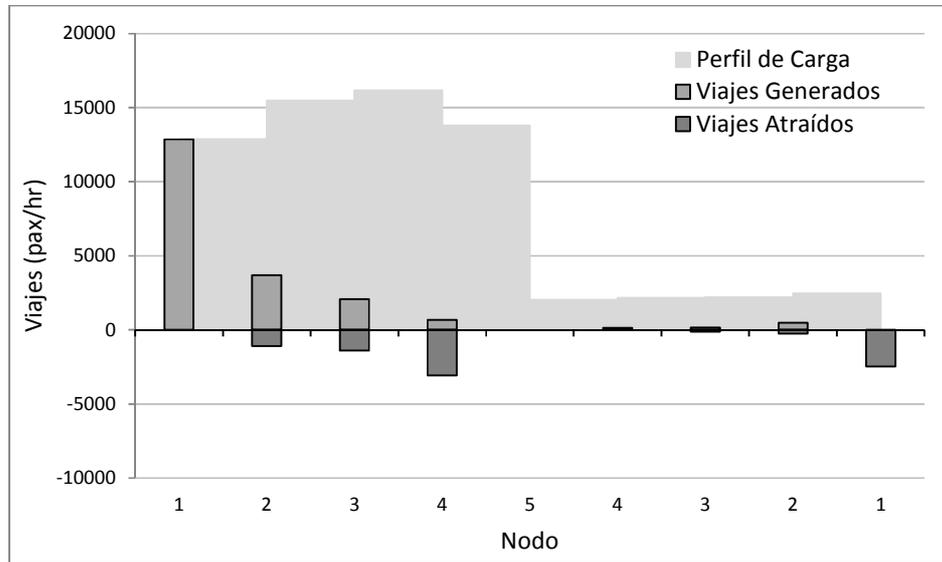


Figura 5-3: Perfil de carga zona A

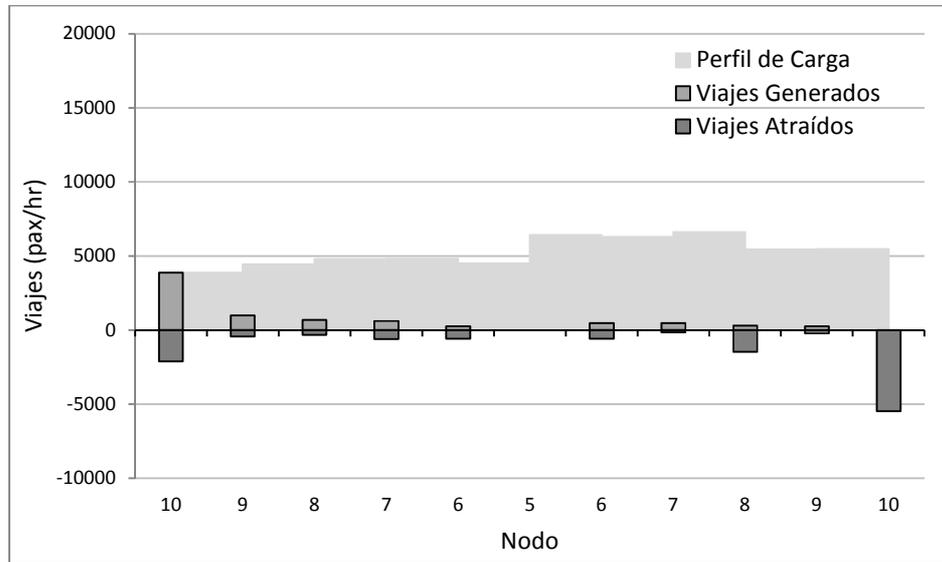


Figura 5-4: Perfil de carga zona B

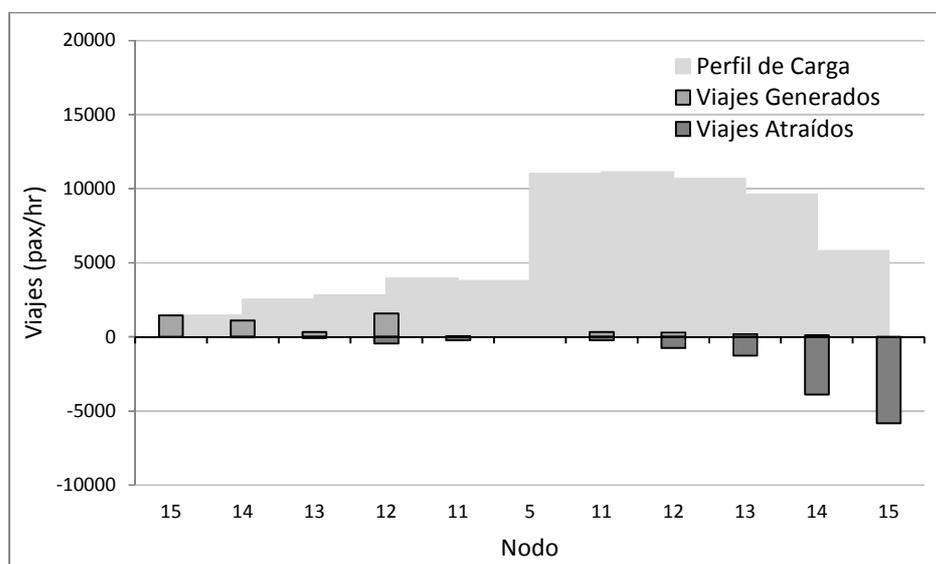


Figura 5-5: Perfil de carga zona C

Estos gráficos muestran la generación y atracción de viajes junto con el perfil de carga dentro cada zona de la red, separadas por sentido: hacia el centro (lado izquierdo de los gráficos) y desde el centro (lado derecho). Estos perfiles se condicen con lo ilustrado en la Figura 5-2 y entregan una visión más profunda de la estructura de los viajes en la red. Por ejemplo, se puede constatar que la zona A genera muchos viajes y captura pocos a partir de la asimetría de sus perfiles por sentido, y se puede apreciar además que la generación de viajes está principalmente concentrada en el nodo 1 en particular. Los perfiles también permiten observar que el arco más cargado de la red corresponde al que va del nodo 4 al nodo 5 en la zona A, y que los tramos más cargados de la red ocurren en la zona A sentido centro, y en la zona C con sentido hacia la periferia.

La Tabla 5-2 a continuación muestra otros parámetros que describen a la red, su demanda, y condiciones de operación.

Tabla 5-2: Parámetros del modelo

Parámetro	Valor	Unidades
λ	1	
θ_{tr}	75	\$
θ_{wt}	30	\$/min
θ_{tt}	15	\$/min
Velocidad	25	km/hr
Distancia entre paradas	1	Km
τ	2	min
c_L	250	\$/bus-Km
c_T	2.500	\$/bus-hr
Capacidad	100	pax/bus

Los valores utilizados en los parámetros del modelo corresponden a aproximaciones a los valores típicamente considerados en instancias de este tipo para la realidad de la ciudad de Santiago, Chile. La única salvedad corresponde a τ , el tiempo de detención por parada, que está ligeramente exagerado en este experimento con el fin de hacer la instancia de estudio más propensa a la utilización de los servicios expresos, de acuerdo a los resultados del Capítulo 4. De todas formas, un valor alto del tiempo de detención también se puede interpretar como una situación donde la red a optimizar representa una situación más agregada del problema, donde los nodos representan grupos de paradas agrupadas en zonas, instancia en la que un valor de dos minutos para τ no representaría en ningún caso una exageración.

En este experimento se asume que los tiempos de viaje y los costos de operación se pueden calcular como se propone en las fórmulas (2.1) a (2.5), y se asume que se puede despreciar el efecto de maniobras de retorno de los buses en tiempos y costos.

Para la aplicación de la heurística consideraremos tres trazados sobre la red, formados por los nodos del 1 al 10 (trazado 1), del 1 al 15 (trazado 2), y del 10 al 15 (trazado 3). Además, se limitarán los trasbordos de la red al paradero 5 exclusivamente.

En lo que sigue se muestra la aplicación del algoritmo paso por paso sobre la red en estudio.

- **Iteración 1: Optimización de frecuencias red completa**

En esta primera iteración se aplica el modelo de optimización de frecuencias sobre la red completa, considerando como servicios del conjunto \mathcal{L} a los seis servicios regulares que operan de extremo a extremo de los tres trazados definidos. Las líneas, frecuencias y otros indicadores obtenidos al final de esta primera iteración se presentan en las tablas a continuación:

Tabla 5-3: Resultados iteración 1

l	Paraderos											Frec. (bus/hr)	Carga Máx. (pax/bus)
1	1	2	3	4	5	6	7	8	9	10	46,8	154,3	
2	10	9	8	7	6	5	4	3	2	1	51,4	38,7	
3	1	2	3	4	5	11	12	13	14	15	49,8	180,8	
4	15	14	13	12	11	5	4	3	2	1	45,2	20,3	
5	10	9	8	7	6	5	11	12	13	14	15	37,7	97,2
6	15	14	13	12	11	5	6	7	8	9	10	42,3	84,9

Tabla 5-4: Indicadores iteración 1

Indicador	Valor (\$/hr)
<i>OC</i>	1.100.337
<i>TTC</i>	13.018.830
<i>WTC</i>	1.100.487
<i>TRC</i>	0
<i>UC</i>	14.119.317
<i>SC</i>	15.219.655
<i>FTTC</i>	8.191.080
<i>SC'</i>	7.028.575

Esta última tabla presenta un resumen de los costos de la solución obtenida. Cada fila corresponde a: costos de operación, costos por tiempo de viaje, espera y trasbordo, costos de los usuarios, costos sociales, la parte fija de los costos por tiempo de viaje y los costos sociales corregidos (es decir, que dejan fuera la parte fija recién mencionada). De este resultado se puede observar, en primer lugar, que el óptimo para la solución con sólo servicios regulares es una solución sin trasbordos y que aprovecha la flexibilidad que entrega la opción de trabajar con continuidad de frecuencias en vez de servicios cíclicos. En este ejemplo, si se trabajara con servicios cíclicos (que equivale a imponer $f_1 = f_2$, $f_3 = f_4$ y $f_5 = f_6$) se obtendría un costo total corregido de 7.030.047, levemente mayor al obtenido por la vía escogida. Cabe notar que para las siguientes iteraciones el uso de servicios cíclicos complicaría considerablemente la implementación del algoritmo al obligar a generar servicios completos para ida y vuelta, considerando sus respectivas combinaciones, y obligaría a replantear las heurísticas de generación de servicios.

Se observa además que la solución cumple la igualdad entre los costos del operador y los tiempos de espera. La pequeña diferencia entre estos valores se explica por la tolerancia del solver utilizado en la resolución del problema.

- **Iteración 2: Generación de servicios para el trazado 1**

Para poder generar los servicios del trazado 1 es necesario primero determinar la matriz de viajes de este trazado, la que se obtiene aplicando la fórmula (5.1). A modo de ejemplo, obtendremos $T_{1,2}^1$, es decir, cuántos viajes desde 1 a 2 utilizan el trazado 1 en la solución actual. En primer lugar se observa que dado que no existen trasbordos en este caso, se tiene que $V_s^w = T_w$ cuando $w = s$, y $V_s^w = 0$ en los casos restantes. Aplicando fórmula (5.1) se obtiene que:

$$T_{1,2}^1 = \frac{f_1}{f_1 + f_3} \sum_{s \in \mathcal{W}} V_{1,2}^s = T_{1,2} \frac{f_1}{f_1 + f_3} = \frac{1.088 \cdot 46,8}{46,8 + 49,8} = 527,2^{pax}/hr \quad (5.4)$$

Por simplicidad en los cálculos anteriores se utilizan las frecuencias en lugar de las frecuencias auxiliares debido a que todos los servicios involucrados son atractivos para el par. También resulta necesario calcular las frecuencias exógenas al trazado para asegurar el balance de frecuencias en la red completa. Para el nodo 1 la única frecuencia entrante de servicios de otros trazados corresponde a f_3 , y la única frecuencia saliente de otro trazado es f_4 , con lo que $F_1 = f_3 - f_4 = 4,57buses/hr$. En el nodo 10 un cálculo similar arroja que $F_{10} = -4,57buses/hr$, mientras que en resto de los nodos estos valores son nulos ya que no existen líneas comenzando o terminando en ellos.

Resolviendo el problema de generación de servicios utilizando la heurística consolidada del Capítulo 3, utilizando las partes de la heurística recomendadas antes en este capítulo, se obtienen los resultados que se muestran en las siguientes tablas:

Tabla 5-5: Resultados iteración 2

<i>l</i>	Paraderos											Frec. (bus/hr)	Carga Máx. (pax/bus)
1	1	2	3	4	5	6	7	8	9	10	28,3	98,1	
2	10	9	8	7	6	5	4	3	2	1	39,9	16,3	
3	1	2	3	4	5	11	12	13	14	15	49,8	179,7	
4	15	14	13	12	11	5	4	3	2	1	45,2	20,7	
5	10	9	8	7	6	5	11	12	13	14	15	37,7	97,2
6	15	14	13	12	11	5	6	7	8	9	10	42,3	84,1
7	10	9	8	5	1	45,7	35,0						
8	1	10	29,7	90,8									
9	1	2	4	5	6	8	9	10	23,0	76,0			

Tabla 5-6: Indicadores iteración 2

Indicador	Valor (\$/hr)
<i>OC</i>	1.324.254
<i>TTC</i>	12.104.130
<i>WTC</i>	1.194.934
<i>TRC</i>	0
<i>UC</i>	13.299.064
<i>SC</i>	14.623.318
<i>FTTC</i>	8.191.080
<i>SC'</i>	6.432.238
Ahorro Iteración	8,5%
Ahorro Acumulado	8,5%

En esta iteración el algoritmo agrega tres nuevos servicios para el trazado 1. Las frecuencias de los servicios regulares del trazado 1 también se ven afectadas, mientras que las de los otros corredores permanecen inalteradas. Se observa además que la optimización del trazado 1 tampoco introduce nuevos trasbordos. Esta primera iteración conlleva una disminución en los costos ajustados totales de un 8,5%. Es importante notar que para el cálculo de estos costos se reasignan los viajes de la red en lugar de forzarlos a realizar las mismas etapas de viaje consideradas para la optimización. Esto no afecta en nada el desarrollo del algoritmo, puesto que la próxima etapa considera como entrada la matriz de los T_w sin asignar y los servicios existentes, sin considerar ni la asignación ni las frecuencias actuales.

- **Iteración 3: Optimización de frecuencias red completa**

Dado que la etapa anterior se enfoca sólo en la optimización de un único corredor e ignora por completo el resto de la red, es probable que los nuevos ahorros de tiempo ofrecidos por el trazado 1 atraigan algunos viajes hacia él, disminuyendo los costos totales del sistema, ya que dado que los costos del operador son fijos en una asignación, y que la capacidad no está siendo tomada en cuenta por el momento, cualquier disminución de los tiempos de viaje de un individuo se traduce en una disminución en los costos de usuario del sistema. Los resultados de esta iteración son los que se presentan en las siguientes tablas:

Tabla 5-7: Resultados iteración 3

<i>l</i>	Paraderos											Frec. (bus/hr)	Carga Máx. (pax/bus)		
1	1	2	3	4	5	6	7	8	9	10		16,2	122,4		
2	10	9	8	7	6	5	4	3	2	1		14,2	15,8		
3	1	2	3	4	5	11	12	13	14	15		50,0	184,7		
4	15	14	13	12	11	5	4	3	2	1		44,7	14,5		
5	10	9	8	7	6	5	11	12	13	14	15		38,8	94,8	
6	15	14	13	12	11	5	6	7	8	9	10		44,1	84,4	
7	10	9	8	5	1								62,9	28,0	
8	1	10											29,7	90,8	
9	1	2	4	5	6	8	9	10						26,0	86,8

Tabla 5-8: Indicadores iteración 3

Indicador	Valor (\$/hr)
<i>OC</i>	1.259.106
<i>TTC</i>	12.003.896
<i>WTC</i>	1.259.148
<i>TRC</i>	67.050
<i>UC</i>	13.330.094
<i>SC</i>	14.589.200
<i>FTTC</i>	8.191.080
<i>SC'</i>	6.398.120
Ahorro Iteración	0,5%
Ahorro Acumulado	9,0%

Se puede observar como la nueva optimización de frecuencias entrega una solución levemente mejor que la anterior, y que en este caso sí aparecen algunos trasbordos en el sistema. Cabe notar que existe la posibilidad de que al cabo de esta asignación algún viaje contenido por el trazado 2 ó 3 perciba algún beneficio transbordando para utilizar los servicios expresos del trazado 1 en parte de su recorrido. Sin embargo, y a

pesar de que la generación de servicio trabaja a nivel de etapas de viaje, cada etapa de optimización de frecuencias trabaja sobre la matriz T_w , generando la posibilidad de que un viaje que transborda en una iteración pueda ser efectuado en una sola etapa en alguna iteración posterior si los cambios sobre la red así lo ameritan.

- **Iteración 4: Generación de servicios para el trazado 2**

De manera similar a la iteración 2, se debe calcular T_w^2 y F_i para poder optimizar los servicios del trazado. En este caso, dada la existencia de trasbordos, el cálculo de T_w^c es levemente más complejo que en el caso anterior. El par $w = (5,1)$, por ejemplo, resulta atractivo para los viajes originados en los nodos 6, 7, 11, 12, 13, 14 y 15 que tienen como destino el nodo 1, que obtienen un ahorro en el tiempo esperado de viaje realizando un trasbordo en el nodo 5. El total de estos viajes que trasbordan es de 730, que sumados a los 61 viajes existentes en el par da un valor para $\sum_{s \in W} V_{5,1}^s$ de 791 viajes a la hora. Estos viajes se reparten en proporción a las frecuencias de los servicios atractivos que atienden el par. Dado que el servicio 7 omite 6 paraderos entre 5 y 1, generando un ahorro de 6 minutos con respecto a los otros servicios disponibles, y considerando que su tiempo esperado de espera es inferior al minuto, este servicio resulta atractivo por sí solo. Como el servicio 7 corresponde al trazado 1, podemos concluir que los 791 viajes están siendo efectuados a través del trazado 1 en la solución actual y que por consiguiente $T_{5,1}^2 = 0$ en este caso.

Los valores de las frecuencias exógenas en esta iteración son de $F_1 = -5,30pax/hr$ y $F_{15} = 5,30pax/hr$, y adoptan valores nulos en el resto de los nodos. Los resultados de la iteración 4 se muestran a continuación:

Tabla 5-9: Resultados iteración 4

<i>l</i>	Paraderos											Frec. (bus/hr)	Carga Máx. (pax/bus)	
1	1	2	3	4	5	6	7	8	9	10		16,2	129,5	
2	10	9	8	7	6	5	4	3	2	1		14,2	14,6	
3	1	2	3	4	5	11	12	13	14	15		21,0	88,2	
4	15	14	13	12	11	5	4	3	2	1		33,5	11,7	
5	10	9	8	7	6	5	11	12	13	14	15		38,8	98,9
6	15	14	13	12	11	5	6	7	8	9	10		44,1	84,4
7	10	9	8	5	1							62,9	28,0	
8	1	10										29,7	90,8	
9	1	2	4	5	6	8	9	10				26,0	92,9	
10	1	2	3	14	15							41,1	136,2	
11	15	14	12	5	3	2	1					26,4	16,1	
12	1	2	3	4	5	13	14	15				19,4	84,9	
13	15	14	5	2	1							16,4	12,4	

Tabla 5-10: Indicadores iteración 4

Indicador	Valor (\$/hr)
<i>OC</i>	1.470.931
<i>TTC</i>	11.095.401
<i>WTC</i>	1.312.135
<i>TRC</i>	39.450
<i>UC</i>	12.446.985
<i>SC</i>	13.917.916
<i>FTTC</i>	8.191.080
<i>SC'</i>	5.726.836
Ahorro Iteración	10,5%
Ahorro Acumulado	18,5%

Como se puede apreciar, esta iteración incorpora cinco nuevos servicios y genera nuevos ahorros en los costos totales de la red, llegando a un ahorro acumulado de un

18,5%. Se puede observar también una reducción en el número de trasbordos, que se explica con la incorporación de alternativas más rápidas para los viajes largos del trazado 2 que en algunos casos ya no justifican realizar un trasbordo.

- **Iteración 5: Optimización de frecuencias red completa**

En esta iteración se vuelven a optimizar las frecuencias de la red completa, obteniéndose los siguientes resultados:

Tabla 5-11: Resultados iteración 5

<i>l</i>	Paraderos											Frec. (bus/hr)	Carga Máx. (pax/bus)
1	1	2	3	4	5	6	7	8	9	10	16,0	114,1	
2	10	9	8	7	6	5	4	3	2	1	8,3	14,1	
3	1	2	3	4	5	11	12	13	14	15	0,0	0,0	
4	15	14	13	12	11	5	4	3	2	1	18,3	17,0	
5	10	9	8	7	6	5	11	12	13	14	15	47,1	90,3
6	15	14	13	12	11	5	6	7	8	9	10	29,4	84,9
7	10	9	8	5	1	57,3	30,7						
8	1	10	29,5	91,6									
9	1	2	4	5	6	8	9	10	37,9	98,0			
10	1	2	3	14	15	43,9	127,4						
11	15	14	12	5	3	2	1	12,1	32,2				
12	1	2	3	4	5	13	14	15	22,5	103,8			
13	15	14	5	2	1	53,8	31,6						

Tabla 5-12: Indicadores iteración 5

Indicador	Valor (\$/hr)
<i>OC</i>	1.402.602
<i>TTC</i>	10.838.711
<i>WTC</i>	1.402.541
<i>TRC</i>	166.725
<i>UC</i>	12.407.977
<i>SC</i>	13.810.579
<i>FTTC</i>	8.191.080
<i>SC'</i>	5.619.499
Ahorro Iteración	1,9%
Ahorro Acumulado	20,0%

Nuevamente el efecto de esta etapa es relativamente menor en comparación con las etapas de generación. Sin embargo, los cambios en la configuración del sistema no son menores. Llama la atención en particular la aparición de una frecuencia nula en un servicio regular, que genera trasbordos forzosos en algunos pares (de 1 a 7, por ejemplo). Esto habla del poder de la herramienta implementada, que es capaz de identificar este tipo de oportunidades y explotarlo para la reducción de los costos del sistema. Cabe notar que por construcción se optó por que el algoritmo mantuviera siempre los servicios regulares dentro del conjunto \mathcal{L} , incluso cuando sus frecuencias se anulan, nada más que por un tema de orden.

- **Iteración 6: Generación de servicios para el trazado 3**

Los resultados de esta iteración se presentan en las siguientes tablas:

Tabla 5-13: Resultados iteración 6

<i>l</i>	Paraderos											Frec. (bus/hr)	Carga Máx. (pax/bus)		
1	1	2	3	4	5	6	7	8	9	10		16,0	105,7		
2	10	9	8	7	6	5	4	3	2	1		8,3	15,4		
3	1	2	3	4	5	11	12	13	14	15		0,0	0,0		
4	15	14	13	12	11	5	4	3	2	1		18,3	13,6		
5	10	9	8	7	6	5	11	12	13	14	15		25,1	46,3	
6	15	14	13	12	11	5	6	7	8	9	10		28,2	30,2	
7	10	9	8	5	1								57,3	32,1	
8	1	10											29,5	91,6	
9	1	2	4	5	6	8	9	10						37,9	98,0
10	1	2	3	14	15								43,9	127,4	
11	15	14	12	5	3	2	1						12,1	18,9	
12	1	2	3	4	5	13	14	15						22,5	109,7
13	15	14	5	2	1								53,8	19,3	
14	15	13	12	5	10								29,0	112,0	
15	10	14	15										28,8	61,9	
16	10	9	7	6	5	12	13	14	15					8,8	55,0
17	10	9	7	6	5	12	14	15						7,5	48,8
18	10	9	7	5	12	14	15						4,8	45,8	

Tabla 5-14: Indicadores iteración 6

Indicador	Valor (\$/hr)
OC	1.605.882
TTC	9.946.836
WTC	1.552.174
TRC	150.000
UC	11.649.010
SC	13.254.892
FTTC	8.191.080
SC'	5.063.812
Ahorro Iteración	9,9%
Ahorro Acumulado	28,0%

Nuevamente, la generación de servicios entrega importantes ahorros en los costos sociales. Esto resulta esperable al ser la primera iteración que se enfoca en una parte de la red que hasta el momento estaba siendo atendido exclusivamente por servicios regulares.

- **Iteraciones 7 a 13**

A contar de este punto, el algoritmo prosigue su curso alternando los trazados 1, 2 y 3, con sus respectivas etapas de optimización de frecuencias de red intermedias. Las tablas a continuación muestran la solución obtenida en la iteración 13, después de completar otra vuelta completa de los tres trazados.

Tabla 5-15: Resultados iteración 13

<i>l</i>	Paraderos											Frec. (bus/hr)	Carga Máx. (pax/bus)	
1	1	2	3	4	5	6	7	8	9	10		24,4	54,6	
2	10	9	8	7	6	5	4	3	2	1		11,1	19,2	
3	1	2	3	4	5	11	12	13	14	15		0,0	0,0	
4	15	14	13	12	11	5	4	3	2	1		41,9	19,5	
5	10	9	8	7	6	5	11	12	13	14	15		0,0	0,0
6	15	14	13	12	11	5	6	7	8	9	10		0,0	0,0
7	10	9	8	5	1							44,0	35,8	
8	1	10										29,9	90,3	
9	1	4	5	6	8	9	10					43,4	113,1	
10	10	9	5	1								20,5	21,2	
11	15	14	5	2	1							51,0	22,4	
12	1	2	3	4	5	13	14	15					25,2	68,6
13	1	2	14	15								18,0	99,3	
14	1	2	3	14	15							27,4	139,4	
15	15	12	5	10								45,9	75,5	
16	10	14	15									28,2	63,1	
17	10	7	6	5	12	14	15					18,5	51,8	
18	10	9	7	6	5	11	12	13	14	15		21,4	49,5	

Tabla 5-16: Indicadores iteración 13

Indicador	Valor (\$/hr)
<i>OC</i>	1.648.084
<i>TTC</i>	9.588.485
<i>WTC</i>	1.648.064
<i>TRC</i>	209.325
<i>UC</i>	11.445.875
<i>SC</i>	13.093.959
<i>FTTC</i>	8.191.080
<i>SC'</i>	4.902.879
Ahorro Acumulado	30,2%

Como se puede apreciar, el efecto de estas últimas iteraciones sobre el ahorro total es mucho menor que el de las primeras seis iteraciones. En este punto se optó por detener el algoritmo debido a que el efecto sobre el ahorro total de cada iteración ya resultaba muy bajo. De hecho, las últimas seis iteraciones en conjunto aportaron con una disminución menor al 2%.

La Figura 5-6 ilustra la evolución de los costos para las diferentes iteraciones realizadas.

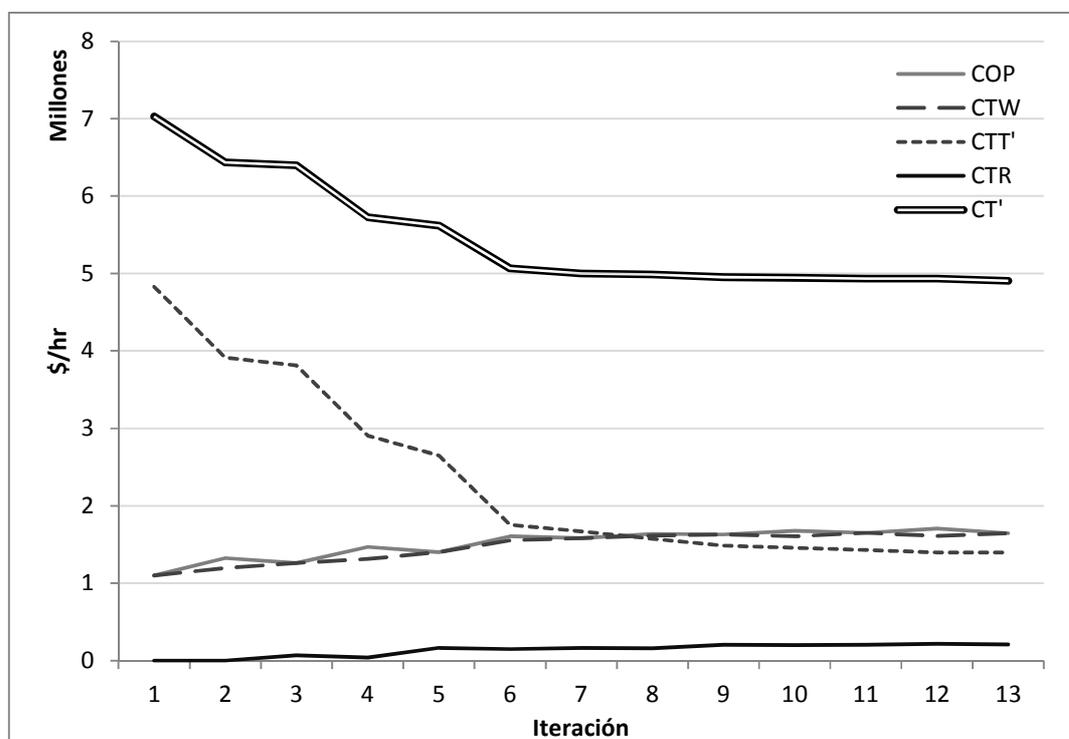


Figura 5-6: Evolución de los indicadores

En el gráfico se puede apreciar con claridad cómo en las primeras seis iteraciones se alcanza casi la totalidad de los beneficios entregados por el algoritmo. En este caso en particular se puede observar, además, que la mayor parte de los ahorros provienen

de una disminución en los tiempos de viaje de los usuarios, que viene de la mano de un leve aumento en el resto de los costos involucrados.

Otro resultado interesante que se puede apreciar en el gráfico es que las curvas de los costos de operación y de espera están prácticamente superpuestas, coincidiendo en las iteraciones impares, que son las iteraciones donde se optimizan las frecuencias de la red completa. Esto confirma el resultado discutido en el Capítulo 2, que dice que cuando las frecuencias son óptimas para un sistema irrestricto, los costos del operador se igualan con los de espera de los usuarios, incluso cuando los trasbordos son permitidos.

- **Ajuste de capacidad y simplificación de la solución**

Es posible que en la práctica sea deseable limitar el número de servicios expresos a operar en cada dirección de un corredor para hacerlo más fácil de utilizar por los usuarios. En particular, no parece tener mucho sentido entregar sobre un trazado servicios muy parecidos entre sí, como lo son los servicios 13 y 14, y 17 y 18. Un criterio para simplificar la solución que es posible aplicar en este caso es, para cada par de servicios, tomar el de menor frecuencia y forzarla a cero. La solución resultante de aplicar este procedimiento se entrega a continuación:

Tabla 5-17: Resultados solución simplificada

<i>l</i>	Paraderos										Frec. (bus/hr)	Carga Máx. (pax/bus)
1	1	2	3	4	5	6	7	8	9	10	23,8	55,9
2	10	9	8	7	6	5	4	3	2	1	13,2	18,3
3	15	14	13	12	11	5	4	3	2	1	39,6	20,5
4	10	9	8	5	1						64,6	29,3
5	1	10									29,6	91,3
6	1	4	5	6	8	9	10				43,0	114,2
7	15	14	5	2	1						47,6	23,9
8	1	2	3	4	5	13	14	15			25,4	64,7
9	1	2	3	14	15						43,2	129,5
10	15	12	5	10							44,3	78,1
11	10	14	15								29,1	61,1
12	10	9	7	6	5	11	12	13	14	15	33,8	68,5

Tabla 5-18: Indicadores solución simplificada

Indicador	Valor (\$/hr)
<i>OC</i>	1.604.398
<i>TTC</i>	9.742.974
<i>WTC</i>	1.604.057
<i>TRC</i>	191.400
<i>UC</i>	11.538.431
<i>SC</i>	13.142.829
<i>FTTC</i>	8.191.080
<i>SC'</i>	4.951.749
Ahorro Acumulado	29,5%

Como se puede ver, la nueva solución (para la que se omiten los servicios regulares de frecuencia nula) es más simple que la anterior, presentando costos sociales muy cercanos. Este resultado se condice con lo observado en Leiva et al. (2010) donde se

prueba que limitar el número de servicios tiene un efecto muy bajo sobre la calidad de la solución al trabajar en un corredor aislado.

La solución acá expuesta viola la capacidad máxima de los buses en dos de sus servicios. Para corregir esta situación aplicaremos sobre este resultado la heurística HC_2 descrita en el Capítulo 3. Los resultados obtenidos son los que se presentan en las tablas a continuación:

Tabla 5-19: Resultados solución con capacidad

<i>l</i>	Paraderos										Frec. (bus/hr)	Carga Máx. (pax/bus)
1	1	2	3	4	5	6	7	8	9	10	23,0	57,9
2	1	4	5	6	8	9	10				49,1	100,0
3	1	10									29,1	92,8
4	10	9	8	7	6	5	4	3	2	1	15,1	17,0
5	10	9	8	5	1						67,3	28,2
6	1	2	3	4	5	13	14	15			23,7	68,7
7	1	2	3	14	15						56,0	100,0
8	15	14	13	12	11	5	4	3	2	1	43,4	19,1
9	15	14	5	2	1						55,2	20,8
10	10	9	7	6	5	11	12	13	14	15	34,7	67,4
11	10	14	15								28,9	61,7
12	15	12	5	10							44,7	77,6

Tabla 5-20: Indicadores solución con capacidad

Indicador	Valor (\$/hr)
<i>OC</i>	1.722.174
<i>TTC</i>	9.743.839
<i>WTC</i>	1.507.914
<i>TRC</i>	191.400
<i>UC</i>	11.443.154
<i>SC</i>	13.165.328
<i>FTTC</i>	8.191.080
<i>SC'</i>	4.974.248

Naturalmente esta solución tiene un costo más alto que la del caso irrestricto, pero la diferencia no es tan alta debido a que la solución irrestricta sólo posee dos líneas que violan la restricción de capacidad, y no por demasiado. Como punto de referencia para evaluar la calidad de esta solución se aplicó también la heurística de capacidad sobre la solución inicial para un servicio regular de la primera iteración. Los resultados de dicho caso se muestran a continuación:

Tabla 5-21: Resultados solución sin expresos con capacidad

<i>l</i>	Paraderos											Frec. (bus/hr)	Carga Máx. (pax/bus)			
1	1	2	3	4	5	6	7	8	9	10					68,5	100,2
2	10	9	8	7	6	5	4	3	2	1					72,4	29,1
3	1	2	3	4	5	11	12	13	14	15					92,9	100,3
4	15	14	13	12	11	5	4	3	2	1					88,9	11,1
5	10	9	8	7	6	5	11	12	13	14	15				35,9	95,7
6	15	14	13	12	11	5	6	7	8	9	10				39,8	85,3

Tabla 5-22: Indicadores solución sin expresos con capacidad

Indicador	Valor (\$/hr)	
	Sin Expresos	Con Expresos
<i>OC</i>	1.586.650	1.722.174
<i>TTC</i>	13.018.830	9.743.839
<i>WTC</i>	813.905	1.507.914
<i>TRC</i>	0	191.400
<i>UC</i>	13.832.735	11.443.154
<i>SC</i>	15.419.385	13.165.328
<i>FTTC</i>	8.191.080	8.191.080
<i>SC'</i>	7.228.305	4.974.248

Se puede apreciar que la reducción porcentual de los costos ajustados totales del sistema son de un 31,2%, ligeramente mayor que en el caso sin congestión, como era de esperar de acuerdo a los resultados obtenidos en el Capítulo 4, donde se muestra que los ahorros que entregan los servicios expresos crecen en presencia de congestión.

La Figura 5-7 presenta una síntesis de la configuración de servicios obtenida para la red en estudio.

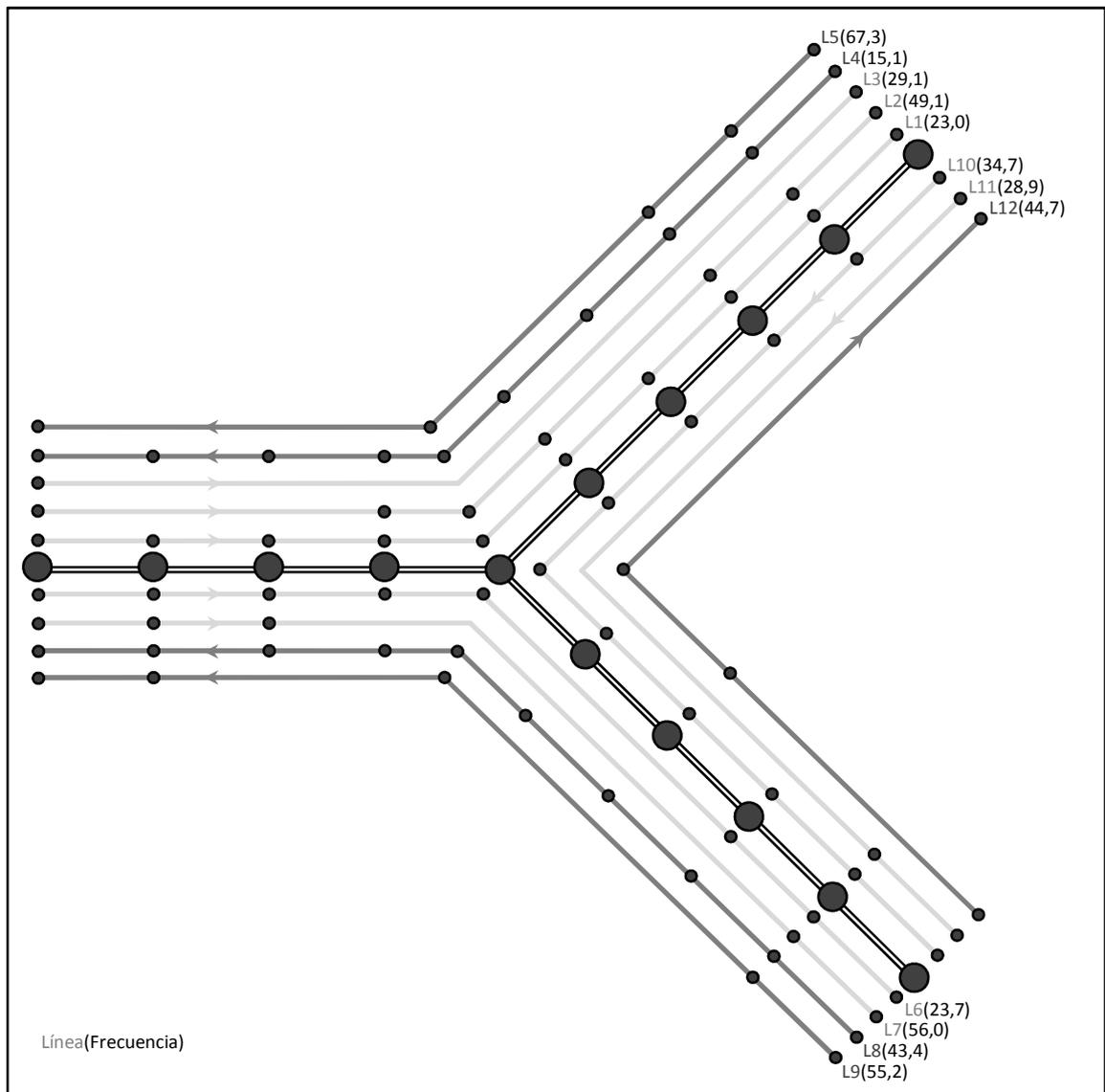


Figura 5-7: Diagrama solución

Los resultados y el diseño obtenido prueban que el modelo es capaz de entregar para una red un conjunto de servicios expresos y sus frecuencias que generen ahorros para el sistema. Esta solución respeta la restricción de capacidad máxima de los vehículos, sin impedir que los usuarios minimicen sus itinerarios de viaje, escogiendo los

servicios que resulten atractivos en cada tramo de su viaje, y realizando trasbordos en caso de ser conveniente.

El modelo acá presentado ofrece un gran potencial para ser aplicado en escenarios con datos extraídos de la práctica y con redes de mayor tamaño. Esto porque, por una parte, los subproblemas de generación de servicios no crecen en complejidad con el crecimiento de la red, sino sólo con la longitud de los trazados, los que se pueden mantener dentro de un tamaño conveniente. Por otra parte, a pesar de que el problema de optimización de frecuencias puede volverse cada vez más complejo, las simplificaciones propuestas lo hacen mucho más abordable que en su formulación original. Además, al postergar el ajuste de capacidad para el final del algoritmo se reduce en un número importante las veces que este problema es invocado.

En problemas de mayor tamaño se presentan nuevas oportunidades a las que la naturaleza del algoritmo podría permitir sacar partido. La codificación de los trazados por separado, por ejemplo, permite incluir en el modelo rutas directas con los ahorros de tiempo provenientes de la factibilidad de usar rutas alternativas en las vías. Por ejemplo, en el problema resuelto en este capítulo, hubiera sido posible (si la matriz de demanda así lo requiriera) definir un trazado como los paraderos 9, 10, 15, y 14, por ejemplo, y que sacara provecho de la cercanía geográfica entre 10 y 15 (asumiendo por un momento que la Figura 5-1: Esquema de la red a estudiar representa espacialmente la ubicación de los nodos) para ahorrar en tiempos de viaje. La codificación trazados con atajos acarrea nuevas posibilidad que podrían ayudar en gran medida a la solución del problema en casos más complejos. Consideremos, por ejemplo, el caso de un corredor de gran tamaño. A pesar de que en principio no

parece natural definir múltiples trazados sobre él, estos “subtrazados” permiten optimizar el problema enfocándose en sus partes. Estos subtrazados, además, no tienen por qué representar tramos de paraderos seguidos, sino que pueden considerar cualquier combinación de paraderos que parezca razonable probar, aprovechando además la posible existencia de atajos, que en corredores más largos tienen mayor probabilidad de existir.

Una debilidad del modelo aquí propuesto podría ser el hecho de depender del criterio del modelador para la definición de los trazados, en forma análoga a como el modelo de Leiva debía ser alimentado de servicios. Sin embargo, la configuración de los trazados no necesariamente es tan compleja como la de los servicios a ofrecer, y en la práctica se puede extraer de los servicios que ya operan sobre la red. Concretamente, se propone como criterio para la generación de los trazados iniciales simplemente utilizar los trazados de los servicios regulares existentes. En el caso de no existir una solución actual para el problema, la literatura ofrece un gran número de modelos que generan los trazados para una red (se mencionan algunos en el Capítulo 1) que se podrían utilizar como punto de partida para el problema.

El modelo para el diseño de servicios expresos en red admite algunas variantes que podrían resultar de interés en ciertos escenarios en particular. Estas variantes corresponden a las mismas que se proponen en el Capítulo 2: más allá de la relajación de la condición de que los servicios sean cíclicos y de la limitación de los puntos de trasbordo, que son simplificaciones claves para permitir que el problema sea abordable, podría ser interesante agregar además la separación por niveles de la asignación y la optimización de frecuencias en las etapas de optimización de

frecuencias del sistema completo. Esta simplificación, a su vez, abre la puerta a experimentar con otros tipos de comportamiento de usuarios, tales como minimización de estrategias, o admitir clases de usuario.

Naturalmente, una próxima etapa del desarrollo de esta línea de investigación consistiría en diseñar una herramienta computacional para la resolución del problema de diseño acá discutido. Esta herramienta idealmente debería ser capaz de incluir la etapa de generación de trazados, además de proveer como opciones las diversas técnicas de solución y simplificaciones propuestas, además de las variantes del modelo ya discutidas.

6. CONCLUSIONES

Como se menciona en un comienzo de la presente tesis, el objetivo principal de este trabajo consistía en determinar cómo se pueden diseñar servicios expresos para buses en una red de transporte público urbano. En el Capítulo 5 de este documento se presenta una heurística que permite diseñar servicios expresos para una red de transporte público, y que además cumple con las condiciones impuestas en un comienzo: considerar capacidad en los vehículos, representar un comportamiento racional de los usuarios, generar nuevos servicios, y, como se indicó recién, ser aplicable sobre una red.

Para construir esta heurística se diseñó, en primer lugar, un algoritmo para la solución del problema de diseño para corredores aislados. Este problema, a su vez, fue separado en dos niveles: la optimización de frecuencias para un conjunto fijo de servicios, abordada en el Capítulo 2, y la generación de servicios expresos, abordada en el Capítulo 3.

Con respecto al problema de optimización de frecuencias, el trabajo toma como punto de partida el modelo de Leiva y propone ciertas mejoras y simplificaciones necesarias para que pudiera ser implementado en los posteriores algoritmos. En primer lugar, se plantea una nueva forma de modelar los servicios, relajando la imposición de que tuviesen que ser cíclicos, lo que hace más eficiente la codificación de los servicios al permitir explorar un mayor dominio usando menos servicios. En segundo lugar, se propone una simplificación consistente en limitar los puntos de trasbordo a ciertos nodos definidos por el modelador. Finalmente, se realiza una mejora a la heurística de ajuste de capacidad de las soluciones, reduciendo su tiempo de ejecución en forma drástica.

Además de las mejoras y simplificaciones mencionadas, se sugieren y plantean otras variantes para el modelo que, a pesar de no ser necesarias para la solución del problema en red, representan interesantes líneas de investigación futura. La primera de estas variantes consiste en plantear el problema de asignación como un problema en un nivel separado del problema de asignación. La segunda, que puede ser combinada con la medida anterior, consiste en la consideración de diferentes formas de comportamiento de usuarios, tales como estrategias mínimas, o clases de usuarios. Finalmente, se discute también otras formas de capacidad que en otros escenarios pueden resultar de interés para la solución del problema. En particular, se plantea el problema donde los paraderos imponen una condición de frecuencia máxima proveniente del espacio limitado para la operación de los vehículos. Este tipo de restricción pareciera ser de gran importancia en sistemas altamente congestionados, donde la solución sin expresos resulta infactible como consecuencia de dicha restricción.

Finalmente, en el Capítulo 2 se presentan también dos resultados analíticos que pueden resultar útiles para el desarrollo de heurísticas de generación o para analizar los resultados del modelo. El primero de estos resultados consiste en la demostración de que en el problema de diseño de servicios expresos en ausencia de congestión los costos del operador en el óptimo son iguales a los costos de espera de los usuarios. Un segundo resultado que se entrega es que es posible conocer un umbral para el número de paraderos que un servicio expreso debe omitir de forma de poder asegurar que a los usuarios les convendrá esperar por este servicio específicamente, y que permite hacer simplificaciones al comportamiento para algunos análisis posteriores.

En el Capítulo 3, como ya se mencionaba antes, se aborda el problema de la generación de servicios para el problema de diseño en un corredor aislado. Para esto se proponen dos tipos de heurísticas: las que funcionan sin tomar en cuenta la capacidad de los vehículos, y las que asumen que los buses operan a capacidad.

Dentro de las del primer tipo existen dos heurísticas que trabajan sobre un formato impuesto de servicio, ya sea servicio corto o zonal, y determina las condiciones para los diferentes servicios de cada tipo que permiten determinar cuáles de éstos pueden representar un ahorro al ser alimentados al problema de optimización de frecuencias. Las otras dos heurísticas de esta categoría estudian el efecto de agregar o eliminar paraderos de un servicio, tomando en cuenta el efecto sobre los pasajeros involucrados y sobre el operador. Uno de estos dos servicios considera además los tiempos de viaje existentes sobre la red a partir de una solución inicial, de forma de que puede ser utilizado para agregar nuevos servicios a una solución conocida.

Las heurísticas de generación trabajan con los mismos servicios de formato conocido mencionados (o la combinación de estos), sacando provecho de la observación de que los servicios que atraviesan el arco crítico deben repartirse la carga crítica entre ellos, lo que permite en algunos casos determinar las frecuencias óptimas de los servicios involucrados, o simplificar su cálculo. Este tipo de enfoque abre una línea de trabajo que puede ser interesante, ya que muestra que, para ciertos casos específicos, el problema de generación se puede resolver analíticamente y a optimalidad. Sin embargo, este enfoque posee la limitación de que trabaja con un número acotado de servicios, lo que en muchos casos significa que sus soluciones son superadas por las soluciones encontradas por heurísticas que no acotan el número de servicios a optimizar.

La heurística para el diseño de servicios expresos en corredores propuesto en el Capítulo 3 fue implementada en el Capítulo 4, donde se experimenta con cerca de mil escenarios para responder la segunda pregunta que se plantea dentro de los objetivos de esta tesis, consistente en determinar las características de un corredor que lo vuelven conveniente para utilizar servicios expresos.

En dicho capítulo se calibra un modelo lineal que predice los beneficios otorgados por los servicios expresos (medidos como el ahorro porcentual de los costos sociales corregidos) a partir de cinco variables explicativas: el ahorro potencial en tiempos de viaje (es decir, cuánto disminuyen los costos por tiempo de viaje si todos los viajes omiten todos los paraderos posibles de omitir), la concentración de los viajes de la matriz origen destino (aproximada como el coeficiente de variación de esta matriz), el nivel de congestión del sistema (entendida como la razón entre la frecuencia mínima del servicio regular y su frecuencia óptima para el caso irrestricto), los costos unitarios de operación del servicio regular, y el tiempo de duración de una parada. El parámetro asociado a estas variables es positivo en todos los casos, salvo en el del costo de operación.

Se concluye en el Capítulo 4 que son factores favorables para el desempeño de los servicios expresos los que influyen positivamente en el ahorro potencial de tiempos de viaje, es decir, el valor del tiempo de viaje, el largo promedio de los viajes, los tiempos de detención (cuyo efecto es además mayor que los mencionados al aparecer como variable aparte en el modelo), y el nivel de demanda agregada de la matriz, cuyo efecto, debido a que influye en el factor de congestión, también es mixto pero siempre positivo. También son factores favorables el nivel de concentración de los viajes, la razón entre el flujo crítico

versus el total de viajes, y un menor valor del tiempo de espera, ya que hace disminuir el nivel de congestión.

Con respecto al efecto de los costos de operación, los resultados muestran que pueden ser tanto favorables como desfavorables para el desempeño de los servicios expresos. Esto se debe a que los costos de operación presentan dos efectos en el modelo: uno directo, acompañado de un factor negativo, que se compensa con uno indirecto por la influencia de este valor en la frecuencia óptima irrestricta y en el factor de congestión, y que el efecto dominante dependerá de las características propias de cada escenario.

Resulta de interés, como futura línea de investigación, buscar un respaldo teórico de las conclusiones obtenidas en este capítulo, a partir de las propiedades analíticas del problema.

Una posible estrategia para abordar este problema podría consistir en trabajar con casos simplificados como los planteados en el Capítulo 3, que permiten llegar a expresiones analíticas para los ahorros en costos sociales de los servicios expresos.

La heurística para corredores presentada en el Capítulo 3 es extendida en el Capítulo 5 para su aplicación en una red. La estrategia adoptada consiste en dividir la red en trazados (permitiendo traslapes), que son optimizados en forma secuencial. Más específicamente, el algoritmo consiste en la alternación de dos fases. En una primera fase, las frecuencias de la red completa son optimizadas, para los servicios existentes, aplicando el modelo de Leiva mejorado propuesto en el Capítulo 2. En una segunda fase, se selecciona un trazado y se fija su matriz de demanda actual, para posteriormente generar nuevos servicios y optimizarlos aplicando la heurística del Capítulo 3, cuidando de respetar los desbalances de frecuencias que pueden existir al interactuar con servicios de otros trazados.

El algoritmo propuesto mostró poseer un gran potencial para disminuir los costos del sistema, obteniéndose una reducción del orden del 30% en los costos sociales corregidos en el escenario donde fue puesto a prueba. El algoritmo es capaz de resolver el problema para una red como la del ejemplo sin problemas, lo que hace pensar que puede ser aplicado exitosamente sobre redes de mayor tamaño tomadas de casos reales. Con este fin se proponen, además de las simplificaciones mencionadas en el Capítulo 2, algunas mejoras al algoritmo que sacan provecho de la estructura de corredores con la que éste trabaja.

Como próximo paso en esta línea de investigación queda propuesta la implementación de la heurística acá descrita como un software computacional. Esta herramienta debiese ser capaz de incorporar las distintas variantes discutidas en el Capítulo 2, entregar la opción de incluir las diferentes heurísticas de generación de servicios, y generar automáticamente los corredores para el problema. En el Capítulo 5 se entrega una breve discusión sobre posibles formas de enfrentar el problema de la generación de trazados para la heurística propuesta en este trabajo.

BIBLIOGRAFIA

Baaj, M.H. y Mahmassani, H.S. (1995). A Hybrid Route Generation Heuristic Algorithm for the Design of Transit Networks. *Transportation Research Part C*, 3(1), 31–50.

Bradley, M.J. (2010). Ultra-Efficient Bus Rapid Transit Timetabling. Tesis de doctorado, Curtin University of Technology.

Ceder, A. y Stern, H.I. (1981). Deficit Function Bus Scheduling with Deadheading Trip Insertion for Fleet Size Reduction. *Transportation Science*, 15(4), 338–363.

Ceder, A. y Wilson, N. (1986). Bus Network Design. *Transportation Research Part B*, 20(4), 331–344.

Ceder, A. (2001). Efficient Timetabling and Vehicle Scheduling for Public Transport. In: Voß, S. y Daduna J.R. (Eds.), *Computer-Aided Scheduling of Public Transport, Lecture Notes in Economics and Mathematical Systems*, vol. 505. Springer, 37-52.

Ceder, A. (2003). Designing Public Transport Network and Routes. In: Lam, W.H.K., Bell, M.G.H. (Eds.), *Advanced Modeling for Transit Operations and Service Planning*. Pergamon, 59–92.

Cepeda, M., Cominetti, R. y Florian, M. (2006). A Frequency-Based Assignment Model for Congested Transit Networks with Strict Capacity Constraints: Characterization and Computation of Equilibria. *Transportation Research Part B*, 40(6), 437–459.

Chang, S.K. y Schonfeld, P. M. (1991). Multiple Period Optimization of Bus Transit Systems. *Transportation Research Part B*, 25(6), 453–478.

Chen, X., Hellinga, B., Chang, C. y Fu, L. (2012). Optimization of Headways for Bus Rapid Transit System with Stop-Skipping Control. Transportation Research Board Annual Meeting 2012 Paper #12-1999.

Chiraphadhanakul, V. y Barnhart C. (2013). Incremental Bus Service Design: Combining Limited-Stop and Local Bus Services. Enviado a Public Transport.

Chriqui, C. y Robillard, P. (1975). Common Bus Lines. *Transportation Science*, 9(1), 115–121.

Clarence, G.C., y Hurdle, V.F. (1975). An Operating Strategy for a Commuter Bus System. *Transportation Science*, 9(1), 1-20.

Cominetti, R. y Correa, J. (2001). Common-Lines and Passenger Assignment in Congested Transit Networks. *Transportation Science*, 35(3), 250–267.

Cortés, C., Jara-Díaz, S. y Tirachini, A. (2011). Integrating Short Turning and Deadheading in the Optimization of Transit Services. *Transportation Research A*, 45(5), 419-434.

Daganzo, C. (2010). Structure of Competitive Transit Networks. *Transportation Research Part B*, 44(4), 434–446.

de Cea, J. y Fernández, E. (1993). Transit Assignment for Congested Public Transport Systems: An Equilibrium Model. *Transportation Science*, 27(2), 133–147.

Desaulniers, G. y Hickman, M. (2007). Public Transit. In: Barnhart, C., Laporte, G. (Eds.), *Handbooks in Operations Research and Management Science*, vol. 14. North-Holland, Amsterdam, 69–128.

Dial, R.B. (1967). Transit Pathfinder Algorithm. *Highway Research Record*, 205, 67–85.

El-Geneidy, A.M. y Surprenant-Legault, J. (2010). Limited-Stop Bus Service: An Evaluation of an Implementation Strategy. *Public Transport*, 2(4), 291-306.

Ercolano, J.M. (1984). Limited-Stop Bus Operations: An Evaluation. *Transportation Research Record*, 994, 24–29.

Estrada, M., Roca-Riu, M., Badia, H., Robusté, F., Daganzo, C.F. (2011). Design and Implementation of Efficient Transit Networks: Procedure, Case Study and Validity Test. *Transportation Research Part A*, 45(9), 935-950.

Fan, W. y Machemehl, R. (2006). Using a Simulated Annealing Algorithm to Solve the Transit Route Network Design Problem. *Journal of Transportation Engineering*, 132(2), 122–132.

Fernández, J.E., De Cea, J. y Malbran, H. (2008). Demand responsive urban public transport system design: methodology and application. *Transportation Research Part A*, 42(7), 951–972.

Freyss, M., Giesen, R. y Muñoz, J.C. (2013). Continuous Approximation for Skip-Stop Operation in Rail Transit. *Transportation Research part C*, en vías de publicación.

Furth, P.G. y Day, F.B. (1985). Transit Routing and Scheduling Strategies for Heavy Demand Corridors. *Transportation Research Record*, 1011, 23–26.

Furth, P.G. (1986). Zonal Route Design for Transit Corridors. *Transportation Science*, 20(1), 1–12.

Furth, P.G. (1987). Short Turning on Transit Routes. *Transportation Research Record*, 1108, 42–52.

Guihaire, V. y Hao, J.K. (2008). Transit Network Design and Scheduling: A Global Review. *Transportation Research Part A*, 42(10), 1251–1273.

Jansson, J.O. (1980). A Simple Bus Line Model for Optimization of Service Frequency and Bus Size. *Journal of Transport Economics and Policy*, 14(1), 53–80.

Jara-Díaz, S.R. y Gschwender, A. (2003). From the Single Line Model to the Spatial Structure of Transit Services. Corridors or Direct? *Journal of Transport Economics and Policy*, 37(2), 261–277.

Jordan, W.C. y Turnquist, M.A. (1979). Zone Scheduling of Bus Routes to Improve Service Reliability. *Transportation Science*, 13(3), 242–268.

Kochenberger, G. A. (Ed.). (2003). *Handbook in Metaheuristics*. Springer.

Kocur, G. y Hendrickson, C. (1982). Design of Local Bus Service with Demand Equilibration. *Transportation Science*, 16(2), 149–70.

Larrain, H. y Muñoz, J.C. (2008). Public Transit Corridor Assignment Assuming Congestion Due to Passenger Boarding and Alighting. *Networks and Spatial Economics*, 8, 241–586.

Larrain, H., Giesen, R. y Muñoz, J.C. (2010a). Choosing the Right Express Services for a Bus Corridor with Capacity Restrictions. *Transportation Research Record*, 2197, 63-70.

Larrain, H., Giesen, R. y Muñoz, J.C. (2010b). The Effect of OD Trip Dispersion versus Concentration in Express Service Design. In: General Proceedings of the 12th World Conference on Transport Research Society, Paper ID 3406 R.

Leblanc, L. (1988). Transit System Network Design. *Transportation Research Part B*, 22(5), 383–390.

Leiserson, C. E., Rivest, R. L., y Stein, C. (2001). *Introduction to algorithms*. T. H. Cormen (Ed.). The MIT press.

Leiva, C., Muñoz, J.C., Giesen, R. y Larrain, H. (2010). Design of Limited-Stop Services for an Urban Bus Corridor with Capacity Constraints. *Transportation Research Part B*, 44(10), 1186-1201.

Levinson, H.S., Zimmerman, S., Clinger, J. y Rutherford, C.S. (2002). Bus Rapid Transit: An Overview. *Journal of Public Transportation*, 5(2), 1-30.

Mauttone, A. y Urquhart, M.E. (2009). A Route Set Construction Algorithm for the Transit Network Design Problem. *Computers & Operations Research*, 36(8), 2440–2449.

Medina-Tapia, M., Giesen, R., Muñoz, J.C. (2013). A Model for the Optimal Location of Bus Stops and Its Application to a Public Transport Corridor in Santiago. Aceptado en *Transportation Research Record*.

Mohring, H. (1972). Optimization and Scale Economies in Urban Bus Transportation. *American Economic Review*, 62(4), 591–604.

Nguyen, S. y Pallottino, S. (1988). Equilibrium Traffic Assignment for Large Scale Transit Networks. *European Journal of Operational Research*, 37, 176–186.

Nuzzolo, A. y Crisalli, U. (2009). The Schedule-Based Modelling of Transportation Systems: Recent Developments. In: Wilson, N., Nuzzolo, A. (Eds.), *Schedule-Based Modelling of Transportation Networks Theory and Applications*. Springer. 1-26.

Scortia, H. (2010). Design and Evaluation of BRT and Limited-Stop Services. Tesis de doctorado, MIT.

Schwarcz, S. (2004). Service Design for Heavy Demand Corridors: Limited-Stop Bus Service. Tesis de magister, MIT.

Sheffi, Y. (1985). *Urban Transportation Networks*. Englewood Cliffs, NJ. Prentice-Hall.

Silverman, N.C. (1998). Limited-Stop Bus Service at New York City Transit. *Journal of Transportation Engineering*, 124, 503-509.

Spiess, H. (1983). On Optimal Route Choice Strategies in Transit Networks. Publicación 285, Centre de Recherche sur les Transports, Université de Montréal.

Spiess, H. y Florian, M. (1989). Optimal Strategies: A New Assignment Model for Transit Networks. *Transportation Research Part B*, 23(2), 83–102.

Sun, C., Zhou, W. y Wang, Y. (2008). Scheduling Combination and Headway Optimization of Bus Rapid Transit. *Journal of Transportation Systems Engineering and Information Technology*, 8(5), 61-67.

Tétreault, P. y El-Geneidy, A. (2010). Estimating Bus Run Times for New Limited-Stop Service Using Archived AVL and APC Data. *Transportation Research Part A*, 44(6), 390–402.

Tirachini, A., Cortés, C., Jara-Díaz, S. (2010). Optimal Design and Benefits for a Short Turning Strategy for a Bus Corridor. *Transportation*, 38(1) 169-189.

Turnquist, M.A. (1978). A Model for Investigating the Effects of Service Frequency and Reliability on Bus Passenger Waiting Times. *Transportation Research Record*, 663, 70–73.

Vickrey, W. (1955). Some Implications of Marginal Cost Pricing for Public Utilities. *American Economic Review*, 45, 605–20.

Wright, L. (2003). Sustainable Transport: A Sourcebook for Policy-makers in Developing Cities, Module 3b: Bus Rapid Transit.

Wu, J.H. y Florian, M. (1993). A Simplicial Decomposition Method for the Transit Equilibrium Assignment Problem. *Annals of Operations Research*, 44, 245–260.

Wu, J.H., Florian, M. y Marcotte, P. (1994). Transit Equilibrium Assignment: A Model and Solution Algorithms. *Transportation Science*, 28(3), 193–203.