



Universidad de Valladolid

FACULTAD DE CIENCIAS

TRABAJO FIN DE GRADO

GRADO EN ESTADÍSTICA

Métodos escalables de resolución para el Dial-a-Ride Problem (DARP)

Autor:

Sergio García Prado



Universidad de Valladolid

FACULTAD DE CIENCIAS

TRABAJO FIN DE GRADO

GRADO EN ESTADÍSTICA

Métodos escalables de resolución para el Dial-a-Ride Problem (DARP)

Autor:

Sergio García Prado

Tutor:

Jesús Saez Aguado

Abstract

[TODO]

Resumen

[TODO]

Este trabajo puede ser consultado a través del siguiente enlace:
<https://github.com/garciparedes/tfg-pickup-and-delivery>

Agradecimientos

[TODO]

Prefacio

[TODO]

Índice general

Resumen	1
Agradecimientos	3
Prefacio	5
1. Introducción	9
1.1. Introducción	9
1.2. Objetivos	9
1.3. Metodología	9
1.4. Problema de recogidas y envíos	9
1.5. Aplicaciones	9
1.6. Conclusiones	9
2. Formulación del Problema	11
2.1. Introducción	11
2.2. Contextualización del problema	12
2.3. Notación	23
2.4. Formulación básica	27
2.5. Otras formulaciones	31
2.6. Extensiones	31
2.7. Tiempo Real	32
2.8. Conclusiones	33
3. Métodos de Resolución	35
3.1. Introducción	35
3.2. Métodos de Resolución Exactos	36
3.3. Métodos de Resolución basados en Heurísticas	41
3.4. Métodos de Resolución basados en Metaheurísticas	49
3.5. Conclusiones	50

4. Implementación y Resultados	51
4.1. Introducción	51
4.2. Implementación	51
4.3. Resultados	51
4.4. Conclusiones	52
5. Conclusiones Generales y Próximos pasos	53
5.1. Introducción	53
5.2. Conclusiones	53

Capítulo 1

Introducción

1.1. Introducción

[TODO]

1.2. Objetivos

[TODO]

1.3. Metodología

[TODO]

1.4. Problema de recogidas y envíos

[TODO]

1.5. Aplicaciones

[TODO]

1.6. Conclusiones

[TODO]

Capítulo 2

Formulación del Problema

2.1. Introducción

El problema *Dial-a-Ride* (o *DARP* en modo abreviado) representa una de las modelizaciones más interesantes en el ámbito de los problemas de *optimización combinatoria*. Esto se debe a la gran cantidad de situaciones del mundo real que pueden ser representadas siguiendo dicho esquema. Sin embargo, antes de profundizar en los aspectos más detallados que caracterizan este modelo, es necesario describir el contexto del mismo así como la clase problemas a la cual pertenece. Una vez se haya completado dicha tarea, se estará en condiciones suficientes para poder describir tanto la versión básica como las extensiones más interesantes, tando desde el punto de vista de los aspectos matemáticos, como desde la cantidad de situaciones reales que permiten resolver.

En los últimos años, el número de compañías que basan su modelo de negocio en alguna variante relacionada con el problema *Dial-a-Ride* ha crecido de manera desmesurada. Algunos ejemplos son aquellas basadas en el reparto de comida (u otros productos) desde los establecimientos hasta la casa de los solicitantes, o la modernización del sector del transporte privado en ciudad gracias a la solicitud de viajes desde el dispositivo móvil. Es importante darse cuenta de que dichas compañías llevan a cabo la planificación en tiempo real ya que en contadas ocasiones los clientes solicitan sus servicios con demasiada antelación. Sin embargo, para poder desarrollar soluciones interesantes sobre entornos en tiempo real, lo primero es poder comprender el problema de manera detallada en su versión estática, para después poder llevar a cabo las simplificaciones y extensiones pertinentes que permiten la toma de decisiones de manera eficiente en entornos dinámicos. Por tanto, este capítulo se centra especialmente en la descripción del modelo estático. Sin embargo, en el apartado 2.7 se indican las características especiales del problema cuando este se pretende resolver en tiempo real.

En cuanto a la organización del capítulo, en el apartado 2.2 se describe de manera detallada el problema *Dial-a-Ride* desde el punto de vista de la jerarquía de clases a la cual pertenecen

siguiendo un enfoque descendente (o *top-down*) en los apartados 2.2.1 a 2.2.6. Seguidamente, en el apartado apartado 2.3 se indica la notación que se pretende seguir a lo largo del documento (y que pretende ser utilizada a modo de referencia en capítulos posteriores). El siguiente paso es determinar la *Formulación Básica*, lo cual se lleva a cabo en el apartado 2.4. Posteriormente se describen otras formulaciones equivalentes en el apartado 2.5 para después discutir posibles extensiones del problema en el apartado 2.6. Seguidamente, (tal y como se ha indicado previamente) en el apartado 2.7 se comentan las peculiaridades del modelo cuando se aplica en situaciones en tiempo real. Finalmente, en el apartado 2.8 se describen brevemente los aspectos más importantes del capítulo.

2.2. Contextualización del problema

A continuación se procede a describir la clase de problemas matemáticos a la cual pertenece el problema *Dial-a-Ride*. Dicha descripción se llevará a cabo de fuera hacia dentro, esto es desde la categoría de problemas más amplia hasta la más concreta, pasando por una breve contextualización así como ejemplificación de problemas similares. En la figura 2.1 se muestra un breve resumen gráfico acerca de las clases de problemas que se describen.

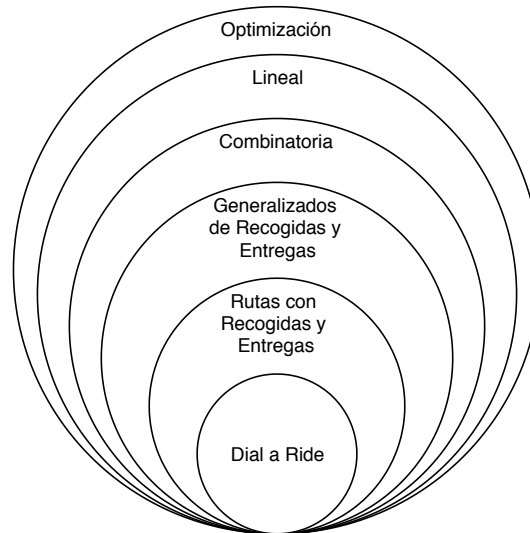


Figura 2.1: Contextualización del *Problema Dial-a-Ride* desde el punto de vista de la jerarquía de problemas a la que pertenece.

2.2.1. Problemas de Optimización

La clase de *problemas de optimización* representa una de las áreas de investigación más interesante en la actualidad, ya que muchas de las innovaciones obtenidas en dicho campo permiten resolver problemas aplicables al mundo real de manera práctica que antes únicamente podían ser resueltos

teóricamente. En concreto, los problemas de optimización son aquellos que se basan en la minimización (o maximización) de una determinada función objetivo (posiblemente vectorial, lo cual define modelos multiobjetivo) de manera que se satisfaga un conjunto de restricciones previamente fijadas sobre un conjunto de variables de decisión que afectan mutuamente a la satisfacibilidad de las restricciones y el valor de la función objetivo.

Dichas variables de decisión pueden ser tanto categóricas como numéricas (discretas o continuas), lo cual genera una gran cantidad de subproblemas diferentes (Nótese que las variables categóricas con k niveles diferentes pueden ser representadas de manera sencilla a partir de $k - 1$ variables binarias). De la misma manera, tanto el valor de función objetivo como las restricciones pueden tener una naturaleza muy diferente: estas pueden estar formadas por funciones lineales de las variables de decisión, como por complicadas funciones no lineales que complican el proceso de obtención del valor óptimo del problema. De manera matemática, la ecuación 2.1 define la formulación de optimización, donde tanto las funciones $f_i(\cdot)$ como $g_k(\cdot)$ son funciones arbitrarias que proyectan el vector de variables de decisión n -dimensional \mathbf{x} en un espacio unidimensional (generando un valor escalar).

$$\begin{aligned} \underset{\mathbf{x} \in \mathbb{R}^n}{\text{Minimizar}} \quad & f_i(\mathbf{x}), \quad \forall i \in \{1, \dots, I\} \\ \text{sujeto a} \quad & g_k(\mathbf{x}) \leq 0, \quad \forall k \in \{1, \dots, K\} \end{aligned} \tag{2.1}$$

Ecuación 2.1: Formulación del modelo de Optimización General

Muchos de los problemas que resolvemos a diario en nuestra vida cotidiana son en cierta medida *problemas de optimización*, desde qué elementos decidimos añadir a nuestra mochila cada día (basados en restricciones de capacidad, funciones objetivo de utilidad y variables de decisión binarias) hasta la detección del rostro por nuestros teléfonos móviles para aplicar un filtro de la manera más realista posible en una videollamada (basados en restricciones de forma, funciones objetivo multidimensionales y millones de variables de decisión numéricas).

2.2.2. Problemas de Optimización Lineal

Una de las categorías de problemas de optimización más ampliamente estudiados por su relativa simplicidad (ya se han desarrollado métodos capaces de obtener soluciones óptimas en un número reducido de pasos) y su gran capacidad de modelización ante muchas situaciones del mundo real son los *problemas de optimización lineal*. Dichos problemas se caracterizan por estar compuestos por variables de decisión compuestas por transformaciones lineales respecto de la función objetivo. Esto quiere decir que tanto el valor de la función objetivo como las posibles restricciones escritas en forma de desigualdades están compuestas por sumas de las variables de decisión multiplicadas por determinados pesos.

En la ecuación 2.2 se muestra a modo de ejemplo la formulación de un problema de optimización lineal (del cual se hablará posteriormente). Como se puede apreciar, en este caso las funciones arbitrarias definidas en la formulación general han sido sustituidas por transformaciones lineales respecto de las variables de decisión. Para la resolución de problemas de este tipo se han desarrollado una gran cantidad de métodos, entre los que destaca un algoritmo altamente eficiente el cual se conoce como *Simplex* [KM70]. Este algoritmo se basa en pivotaje entre soluciones de manera que tras cada iteración se llegue a una solución igual o mejor. Una de las mayores ventajas de la formulación de un problema como lineal es que algoritmos como *Simplex* proporcionan garantías de optimalidad al alcanzar el valor óptimo al terminar completamente su ejecución.

$$\begin{aligned}
& \text{Minimizar} && \sum_{i=1}^N \sum_{j=1}^M c_{ij} x_{ij} \\
& \text{sujeto a} && \sum_{i=1}^N x_{ij} = s_i, \quad \forall i \in \{1, \dots, N\} \\
& && \sum_{j=1}^M x_{ij} = d_j, \quad \forall j \in \{1, \dots, M\} \\
& && x_{ij} \geq 0, \quad \forall i \in \{1, \dots, N\}, \forall j \in \{1, \dots, M\}
\end{aligned} \tag{2.2}$$

Ecuación 2.2: Formulación de un modelo de *Optimización Lineal*. En concreto, el *Problema de Transporte*.

A pesar de que el algoritmo *Simplex* siempre proporcione resultados óptimos, en algunas ocasiones no se posee la capacidad suficiente de cálculo para llegar a la mejor solución. Por lo tanto, se han desarrollado una gran cantidad de métodos conocidos como *heurísticos* (de los cuáles se hablará más en detalle en los apartados 3.3 y 3.4) que a pesar de proporcionar unos buenos resultados, no ofrecen ninguna garantía de optimalidad. Por contra, también existen otros métodos exactos que son capaces de llegar al valor óptimo utilizando menor cantidad de recursos (algunos de los cuales se describen en el apartado 3.2).

Antes de describir algunos de los ejemplos y aplicaciones reales más destacadas basadas en *problemas de optimización lineal*, es necesario describir unos de los problemas más populares de esta categoría, el cual se conoce como *problema de transporte* y se caracteriza por permitir representar de manera matemática la tarea sobre cómo distribuir un conjunto de recursos procedentes de N puntos de origen hasta M puntos de destino, donde cada trayecto tiene un coste diferente, y cada origen y destino unas capacidades de oferta y demanda. La formulación sobre dicho problema se corresponde con la utilizada a modo de ejemplo en la ecuación 2.2.

Sin embargo, los problemas de optimización lineal permiten representar una amplia cantidad de situaciones de nuestra vida diaria. Entre ellos se encuentran los *problemas de mezclas* (donde se pretende generar un compuesto con unas ciertas características a partir de la combinación de otros

tratando de reducir los costes) aunque los problemas de optimización lineal también son de gran utilidad en el ámbito de la economía y los estudios de mercado, permitiendo representar de una manera relativamente sencilla el comportamiento de los clientes ante cambios de precio u otras variables más elaboradas.

2.2.3. Problemas de Optimización Combinatoria

Uno de los factores más relevantes cuando se trata de *problemas de optimización lineal* se refiere a la naturaleza de las variables de decisión, en que (como se ha comentado anteriormente) estas pueden ser de una naturaleza continua o discreta. En este sentido, los problemas pueden clasificarse en 4 categorías: problemas de optimización continua puros (donde existen métodos muy eficientes que permiten ser resueltos en un tiempo razonable), problemas de *optimización entera*, problemas de optimización binaria (o de *optimización combinatoria*) y problema de optimización mixtos (donde mezclan ambos tipos de variables). Para profundizar más en el tema de la *optimización entera* y *optimización combinatoria* se recomienda [WN14]

A pesar de que la exigencia de que las variables de decisión sean discretas en un primer momento puede parecer muy sencilla en un primer momento (por ser muy intuitiva a nivel conceptual), esta complica mucho la labor de optimización del problema. En modo abstracto, esto se debe a que si el espacio de soluciones de un problema de optimización lineal era visto como un *prima* en un espacio n -dimensional (donde n es el número de variables de decisión) y los puntos de interés se refieren a los vértices, que presentarán el máximo/mínimo valor óptimo, en el caso de las variables discretas el *prima* se transforma en una estructura “pixelada” lo cual incrementa de manera exponencial el número de vértices (y por tanto de puntos de evaluación) del problema en la búsqueda del valor óptimo.

Las dificultades descritas en el párrafo anterior provocan que la clase de problemas de optimización de variables discretas, entre los que se encuentra la de los problemas de optimización combinatoria (a la cuál a su vez pertenecen los de rutas, de los que se hablará posteriormente), hacen que el espacio de soluciones se extienda de tal manera que para casos relativamente pequeños, este sea inabordable. A pesar de existir métodos exactos (de los cuales se hablará en el apartado 3.2), la mayor parte de la literatura sobre este tema se ha dedicado a la investigación de métodos que a pesar de no ofrecer garantías de optimalidad, proporcionan un buen acuerdo entre eficiencia computacional y calidad de las soluciones obtenidas (los cuales se detallan en los apartados 3.3 y 3.4).

Los problemas de *optimización combinatoria* se caracterizan porque las variables de decisión del modelo son de carácter binario. Dicha razón permite representar situaciones muy interesantes y diversas del mundo real, lo cual ha hecho que los campos de aplicaciones de este tipo de problemas sean desde la generación de rutas de vehículos con garantías de conectividad, hasta problemas de

$$\begin{aligned}
& \text{Maximizar} && \sum_{i=1}^N u_i x_i \\
& \text{sujeto a} && \sum_{i=1}^N w_i x_i \leq W, \quad \forall i \in \{1, \dots, N\} \\
& && x_i \in \{0, 1\}, \quad \forall i \in \{1, \dots, N\}
\end{aligned} \tag{2.3}$$

Ecuación 2.3: Formulación de un modelo de *Optimización Combinatoria*. En concreto, el *Problema de la Mochila*.

decisión relacionados con la toma o no de una determinada acción. Dentro de los problemas de decisión, existe uno que destaca sobre el resto por su relativa simplicidad teórica, pero a la vez extremada practicidad, el cual se conoce como *problema de la mochila*. A pesar de tener muchas vertientes y extensiones, la idea básica de este es la de maximizar el valor en el proceso de selección de un subconjunto de elementos, dadas unas limitaciones de capacidad y un valor determinado para cada producto. La “dificultad” de este modelo reside en que los elementos no pueden ser escogidos parcialmente, por lo que surge un problema de combinatoria entre los elementos que añadir o no al subconjunto que forma la solución. A modo de ejemplo, en la ecuación 2.3 se muestra la versión más básica del *problema de la mochila*, donde tal y como se puede apreciar, las variables de decisión tan solo pueden tomar valores $\{0, 1\}$ lo cual determina la presencia o no del elemento i -ésimo en el subconjunto solución.

[TODO: describir problemas NP]

Tal y como se ha remarcado a lo largo de todo el apartado, es especialmente importante indicar que la mayor dificultad surgida al añadir las restricciones en el soporte de las variables de decisión genera un mayor espacio de puntos de evaluación, lo cual complica extremadamente los métodos de generación de soluciones. Por contra, esta dificultad amplía en gran medida la capacidad de representación de los modelos, lo cual es algo beneficioso ya que permite aprovechar en mayor medida recursos que de otra forma serían imposibles de aprovechar de la misma manera.

2.2.4. Problemas Generalizados de Recogidas y Entregas

Los problemas de optimización descritos hasta ahora permiten abarcar una gran cantidad de situaciones basadas en tratar de aprovechar en mayor medida uno o más recursos disponibles sobre una colección de restricciones. En el caso de los problemas de *optimización combinatoria*, dicho proceso de aprovechamiento de recursos puede ser visto como un conjunto de problemas de decisión codependientes entre si sobre la utilización (o no) de cada uno de ellos. Sin embargo, en dichos problemas la relación de codependencia no queda fijada de manera explícita, sino que esta es marcada por cada problema concreto. Entonces, tiene sentido pensar que distintos problemas puedan presentar similitudes desde el punto de vista de las relaciones de dependencia entre sus

variables de decisión. Para el resto de subclases de problemas la descripción se apoya en cierta medida en la jerarquía descrita en [PDH08] por su sencillez conceptual pero a la vez potencia desde el punto de vista de los problemas que engloba.

Los *problemas generalizados de recogidas y entregas* se estudiaron inicialmente a partir del *problema del viajante*, del cual hay registros de su existencia desde el siglo *XIX*. A pesar de ello, uno de los primeros papers que describen de manera detallada su formulación [MTZ60] no se publicaría hasta los años 60. Con el paso de los años se ha desarrollado un amplio marco de conocimiento alrededor de este tipo de problemas, entre los que destaca especialmente [TV02]. Estos problemas surgen precisamente como una subcategoría englobada dentro del conjunto de los problemas de *optimización combinatoria*. La restricción que más caracteriza este tipo de problemas se basa en el mantenimiento tanto de orden en el cumplimiento de tareas como del mantenimiento de un concepto de estado a lo largo dicho proceso. Dicho estado en algunos subproblemas es muy débil, mientras que en otros representa una componente muy importante durante el proceso de resolución del mismo.

En muchos casos, los *problemas generalizados de recogidas y entregas* (o *Vehicle Routing Problems*) pueden resumirse de la siguiente manera: Dada una colección de tareas a realizar en un determinado lugar del espacio (con posibles restricciones de orden, duración, capacidad, etc.) así como una colección de vehículos cuya función es moverse de un punto a otro en el espacio generado por las tareas (con posibles restricciones de duración, capacidad, etc.) generando un determinado incremento de coste, se trata de encontrar la *ruta* (o *tour*) para cada vehículo que satisface más y/o con menor coste las tareas disponibles. Entonces, dicho marco engloba una gran cantidad de problemas de optimización caracterizados por dicha restricción, donde están contenidos problemas relativamente dispares, desde el *problema del viajante* (que se describirá brevemente a continuación) como el *problema de recogidas y entregas* (que se describirá de manera más detallada en próximos apartados).

Una de las características más importantes desde el punto de vista de la resolución en este tipo de problemas es la estructura secuencial que estos presentan, así como el elevado número de variables de decisión necesario para su formulación formal. Dichas características conllevan interesantes peculiaridades que convierten en prácticamente inabordable la búsqueda de soluciones factibles utilizando técnicas de resolución genéricas. Por contra, proporcionan ciertas ventajas y simplificaciones sobre el espacio de soluciones las cuales permiten tomar una gran ventaja a métodos de resolución específicos. Entre ellas, la idea que más destaca es la una reducción de orden logarítmico del tamaño del espacio de soluciones conforme las tareas van siendo seleccionadas.

Para la formulación de *problemas generalizados de recogidas y entregas*, se utiliza una serie de restricciones comunes para poder modelar situaciones coherentes. Para ello en la ecuación 2.3 se

incluye la formulación de un problema de rutas que será utilizado a modo de ejemplo. Dicha formulación se corresponde con la versión más simple del *problema del viajante*. Es importante remarcar que la modelización de este problema asume que únicamente existe un vehículo. Antes de proceder con la descripción de las restricciones es necesario describir mínimamente la notación que se utiliza en este caso: Suponemos que el problema está compuesto por N tareas, que el punto de inicio se indexa con la posición 0 (y que también deberá ser la posición final), que las variables de decisión x_{ij} representan la acción de que el vehículo se desplace desde la tarea i -ésima hasta la tarea j -ésima (por lo que X será una matriz de tamaño $(N + 1) \times (N + 1)$) y los movimientos conllevan un coste determinado por c_{ij} . Además, se añade un vector auxiliar de variables de decisión U de longitud N el cual se explicará posteriormente. A continuación se describen dichas restricciones:

- **Unicidad:** Dado que este tipo de problemas se basan en la construcción de rutas cuyo objetivo principal es visitar una colección de tareas que representan una determinada posición en un hiperplano espacial, es necesario restringir el número de veces que estas se visitan a únicamente una. La forma de llevar a cabo esto es restringiendo el número de *enlaces* entre cualquier otra tarea y la tarea en cuestión (así como la tarea en cuestión y cualquier otra tarea) a uno. Esto impone tanto la restricción citada, como que todas las tareas tengan que ser visitadas (existen técnicas para relajar dicha restricción en problemas no factibles totalmente mediante la redefinición como un *problemas multiobjetivo*). Entonces, en la formulación de la ecuación 2.4 las dos primeras restricciones generan la propiedad de unicidad.
- **Secuencialidad:** Además de construir rutas que permitan visitar todas las tareas del problema determinadas en el problema, también es necesario que dichas rutas cumplan las restricciones de secuencialidad entre un “enlace” y el siguiente. Típicamente este tipo de restricciones se conocen como de eliminación de subtours y es un tema ampliamente estudiado en la literatura entre las que destacan las de *Miller-Tucker-Zemlin* descritas en [MTZ60]. Sin embargo, también se han desarrollado soluciones específicas para problemas generalizados de recogidas y entregas tal y como se describe en [DL91]. En su versión más simple, estas restricciones se basan en el apoyo de un vector de variables de decisión (en nuestro caso U) que anotan el orden (a partir del índice o del momento temporal dependiendo de la formulación utilizada) de las tareas a partir de la cual se apoyan las restricciones para que los “enlaces” tengan una relación secuencial entre si. En el caso de la formulación descrita en la ecuación 2.4 esto se consigue a partir de la tercera restricción (que “asigna” y obliga a que exista la relación de orden en los índices anotados en el vector U).

Tal y como se puede apreciar a partir del ejemplo anterior, modelar *problemas generalizados de recogidas y entregas* es una tarea relativamente sencilla. Sin embargo, es importante remarcar que en la formulación de la ecuación 2.4 (*problema del viajante*) únicamente se requiere visitar todas las tareas y la solución óptima se consigue encontrando aquella que menor coste proporcione. Sin

$$\begin{aligned}
& \text{Minimizar} && \sum_{i=0}^N \sum_{j=1, j \neq i}^N c_{ij} x_{ij} \\
& \text{sujeto a} && \sum_{i=0, i \neq j}^N x_{ij} = 1, \quad \forall i \in \{1, \dots, N\} \\
& && \sum_{j=0, j \neq i}^N x_{ij} = 1, \quad \forall j \in \{1, \dots, N\} \\
& && u_i - u_j + N x_{ij} \leq N - 1, \quad 1 \leq i \neq j \leq N \\
& && x_{ij} \in \{0, 1\}, \quad \forall i \in \{1, \dots, N\}, \forall j \in \{1, \dots, N\}
\end{aligned} \tag{2.4}$$

Ecuación 2.4: Formulación de un modelo *Generalizado de Recogidas y Entregas*. En concreto, el *Problema del viajante*.

embargo, el marco de *problemas generalizados de recogidas y entregas* permite modelar situaciones mucho más complicadas. Esto se consigue ampliando las restricciones descritas inicialmente. Algunas de las más comunes se describen a continuación:

- *Capacidad:* Muchos de los problemas generalizados de recogidas y entregas se basan en cargar y/o recoger una determinada mercancía para después llevar al almacén (u otro destino), por lo tanto es necesario tener en cuenta la capacidad del vehículo así como tener constancia de la ocupación del mismo en cada momento de la ruta. Dicha descripción se amplía para el caso de los *problemas de recogidas y entregas* en el apartado 2.4.
- *Ventanas Temporales:* Las tareas no siempre perduran a lo largo del tiempo, sino que tienen unas determinadas franjas temporales en que pueden ser servidas, por lo que dicha información debe ser incluida en la formulación del modelo para que este sea lo más próximo posible a la realidad. Dicha descripción se amplía para el caso de los *problemas de recogidas y entregas* en el apartado 2.4.
- *Duración:* Existen ocasiones en que es necesario imponer restricciones de duración a lo largo de la ruta. Dichas restricciones pueden ser de diferentes tipos, desde el punto de vista de la duración total de la ruta, así como del tiempo que pasa desde el principio de la ruta hasta que se visita una determinada tarea, desde que se visita una determinada tarea hasta el final de la ruta, etc. Dicha descripción se amplía para el caso de los *problemas de recogidas y entregas* en el apartado 2.4.

Hasta ahora, se han descrito los *problemas generalizados de recogidas y entregas* en los que únicamente se considera un vehículo. Sin embargo, es relativamente sencillo ampliar la formulación a casos en los cuales el modelo permite la utilización de varios vehículos. Nótese que esto no es equivalente a la construcción de un modelo para cada vehículo ya que el objetivo es que estos se repartan tareas entre si. Para ello existen distintas formulaciones, entre las que destacan la utilización de modelos de 3 índices (añadiendo una nueva dimensión a X que determina el vehículo)

o de 2 índices (añadiendo nuevas restricciones que restringen el solapamiento de rutas). Cada una de estas alternativas presentan ventajas y desventajas que se describen más en detalle para el caso del *problema de recogidas y entregas* en el apartado 2.4.

2.2.4.1. Relación con problemas de Secuenciación de Tareas

Tras la descripción de los *problemas generalizados de recogidas y entregas* es fácil darse cuenta de la gran similitud que existe con los *problemas de secuenciación de tareas* ya que, al igual que estos, se basan en la optimización de un conjunto de tareas a servir. Antes de proceder con la relación y diferencias entre ambas categorías es necesario detallar que por *problemas de secuenciación de tareas* nos referimos a problemas *Job-Shop Scheduling*, en los cuales el objetivo es la planificación de una serie de tareas codependientes entre si de tal manera que estas se consigan resolver optimizando la duración total.

A pesar de que el enfoque conceptual es muy similar y tal (y como se expone en [BPS03] existen diferencias muy sutiles desde el punto de vista de la formulación), tanto el objetivo como el contexto de ambas categorías de problemas es muy diferente. En [BPS03] se exponen 5 razones por las cuales deben ser considerados como dos problemas diferentes, las cuales se proceden a enumerar a continuación:

- Recursos disponibles: Generalmente los *problemas generalizados de recogidas y entregas* planifican un número elevado de vehículos mientras que los *problemas de secuenciación* suelen poseer un número reducido de servidores.
- Dependencias temporales: Los *problemas generalizados de recogidas y entregas* presentan tareas independientes entre si (excepto en *problemas de rutas con recogidas y entregas* como se describe en próximos apartados) mientras que los *problemas de secuenciación* presentan una elevada componente de dependencias temporales entre las tareas.
- Estructura de tiempo: En los *problemas generalizados de recogidas y entregas* el mayor peso temporal se corresponde en el proceso de enlace entre una tarea y la siguiente, mientras que en el *problema de secuenciación* el peso temporal reside principalmente en la duración de la propia tarea.
- Función Objetivo: En la mayoría de casos los *problemas generalizados de recogidas y entregas* tratan de reducir los costes para generar los trayectos óptimos, mientras que en los *problemas de optimización* el objetivo es minimizar la duración total del proceso (desde el inicio de la primera tarea hasta la finalización de la última).
- Ventanas Temporales: Por lo general, la mayoría de *problemas generalizados de recogidas y entregas* se caracterizan por estar compuestos por “amplias” ventanas temporales, mientras que los *problemas de secuenciación* suelen tener intervalos mucho más ajustados.

A pesar de que dichas diferencias se refieren más a la estructura de la instancia específica del problema a resolver que a la propia modelización del mismo, estas también afectan en cierta manera a la forma de definir la formulación. Dichas diferencias son apreciables sobre todo desde el punto de vista de la definición de la función objetivo. Sin embargo, en el aspecto en que más se diferencian entre si es en la estrategia de resolución a seguir en cada caso desde el punto de vista de la utilización de heurísticas de búsqueda. Es fácil darse cuenta de que las estrategias que funcionen de manera adecuada con los *problemas generalizados de recogidas y entregas* probablemente no se comporten como deberían en *problemas de secuenciación*, y viceversa. Por todas estas razones, deben ser estudiados como problemas diferentes a pesar de tener muchas similitudes entre si.

Tal y como se ha descrito a lo largo del apartado, los problemas generalizados de recogidas y entregas representan un marco de modelización muy interesante que permite representar una gran cantidad de situaciones reales, desde el caso más sencillo relacionado con el *problema del viajante* hasta modelizaciones muy complejas como las de empresas de reparto a gran escala que requieren de la construcción de rutas de reparto que aprovechen al máximo los recursos de la misma. El ejemplo más notable en este campo es el sistema de distribución de productos de empresas como *Amazon* dado que sin técnicas de optimización sería imposible alcanzar un nivel de servicio similar sin caer en costes de servicio inmanejables.

2.2.5. Problemas de Rutas con Recogidas y Entregas

Hasta ahora se ha definido la categoría de *problemas generalizados de recogidas y entregas* de manera general, indicando que estos son aquellos problemas basados en la planificación de trayectos que visitan una serie de puntos definidos como tareas. Sin embargo, este es un marco muy general que a su vez permite modelar una gran cantidad de problemas. Este apartado se centra en el caso de *problemas de rutas con recogidas y entregas*. Dicha modelización se caracteriza porque las tareas no solo se basan en la entrega (o recogida) de una determinada mercancía desde (o hasta) el almacén de origen (o destino), si no que el concepto de tarea queda dividido en dos. Una de ellas representa la recogida de mercancía en un determinado punto, mientras que otra representa la entrega de mercancía en otro punto diferente. Nótese que este modelo se contrapone al enfoque de los *problemas de rutas* donde las mercancías son recogidas (o entregas) hasta (o desde) un almacén común.

Entonces, la característica de recogida y entrega en posiciones arbitrarias del hiperplano espacial sobre el cual se ha definido el problema ofrece una gran potencia desde el punto de vista de los problemas de modelización que permite representar. Sin embargo, dicha potencia conlleva una mayor dificultad desde el punto de vista de los métodos de resolución donde el espacio de soluciones es inabordable. Sin embargo, tal y como se comenta al principio del apartado 2.2.4 la estructura de codependencia entre variables de decisión permite tomar ventaja reduciendo notablemente el espacio de búsqueda.

Dentro de esta clase de problemas, es necesario hacer una importante distinción entre las dos grandes alternativas en que se pueden producir las recogidas y entregas. Estas pueden ser o no pareadas por lo tanto, a continuación se describen las diferencias más notables entre ambas clases:

1. **Pareadas:** Cuando las tareas son pareadas es necesario que exista una recogida para que haya una entrega y ambas deben cumplir la misma capacidad. Además, estas no pueden ser servidas en ordenes opuestos (primero se debe producir la recogida y después la entrega).
2. **No Pareadas:** En el caso de problemas con tareas no pareados las limitaciones anteriormente descritas son eliminadas. Es decir, no es necesario que exista una recogida para su correspondiente entrega (y viceversa). Es de especial importancia darse cuenta de que para que esto sea posible se debe asumir que la mercancía es uniforme entre si (al menos en la formulación más básica) de tal manera que pueda ser intercambiable entre todas las tareas.

2.2.6. Problemas Dial-a-Ride

El *problema Dial-a-Ride* se engloba dentro del marco de los *problemas de rutas con recogidas y entregas* donde las tareas son pareadas entre si. Es decir, donde existe una tarea que representa la recogida con una capacidad igual a la de su correspondiente tarea de entrega, y además estas han de ser realizadas siguiendo un orden estricto. A este tipo de problemas se les conoce como *problemas de recogidas y entregas* o *Pickup and Delivery Problem (PDP)*. Es en esta categoría de problemas donde surge el concepto de *viaje*, que se define como el trayecto dentro de una ruta que realiza cada una de las mercancías desde que estas se recogen hasta que se entregan. Nótese entonces que cada viaje esta formado por dos tareas.

A pesar de que el *problema Dial-a-Ride* pertenece a la categoría de *problemas de recogidas y entregas*, este se caracteriza porque la mercancía a transportar está formada por personas (siendo el ejemplo más significativo el problema de planificación de una compañía de taxis), lo cual implica la creación de restricciones adicionales. Típicamente se añaden restricciones que limitan la duración máxima de cada uno de los viajes a un tiempo máximo, algo que no es tan importante en el caso de mercancías pero si en el de personas.

[TODO: Hablar sobre las descripciones definidas en [CL07]]

2.2.6.1. Aplicaciones en Situaciones Reales

[TODO]

[TODO]

2.3. Notación

Una vez contextualizado de una manera detallada el *problema Dial-a-Ride* es fácil intuir cuáles son los factores que más complican la implementación y el uso de métodos que permitan resolver instancias reales que tan frecuentemente se producen en situaciones del mundo real. En este caso, el *problema Dial-a-Ride* es relativamente sencillo desde el punto de vista conceptual (tal y como se ha indicado anteriormente, se trata de cumplir un conjunto de tareas de transporte desde un punto de recogida hasta un punto de entrega a partir del uso de un conjunto de vehículos). Sin embargo, a pesar de su sencillez conceptual, en muchas ocasiones es complicado representar el problema en lenguaje matemático. Por dicha razón, es de vital importancia definir una nomenclatura común a lo largo de todo el trabajo que permita facilitar la representación matemática de las ideas que se pretende transmitir. Para ello, se incluye este apartado, el cual (como se ha comentado previamente) pretende servir como una guía de referencia que clarifique el sentido de ciertas definiciones futuras.

A pesar de lo expuesto en el párrafo anterior, la tarea de definir una nomenclatura común y aplicable a las posteriores explicaciones no es una tarea sencilla. La dificultad surge principalmente porque los términos no siempre se utilizan de la misma manera y en ocasiones son muy dependientes del contexto. Sin embargo, esto se tratará de hacer en la medida de lo posible. Una vez introducido el apartado, en el apartado 2.3.1 se procede a enumerar todos los términos necesarios para comprender adecuadamente el *problema Dial-a-Ride* para posteriormente definir las variables que se utilizarán en los procesos de formulación, así como en los posteriores métodos de resolución en subsiguientes capítulos. Seguidamente, en el apartado 2.3.2 se definen las Constantes que se utilizan en las formulaciones matemáticas y, finalmente, en el apartado 2.3.3 se describen las variables de decisión necesarias para llevar a cabo la formulación básica del problema.

2.3.1. Palabras clave

A pesar de que para la formulación del problema no sea estrictamente necesario detallar las palabras clave que se utilizan para tratar el problema (y sus correspondientes métodos de resolución), se cree que estos pueden facilitar el entendimiento posterior de las variables utilizadas. Además, estos conceptos serán de vital importancia para definir los métodos de resolución basados en heurísticas y metaheurísticas en capítulos posteriores.

- **Posición:** Vector de coordenadas que representa el lugar en que se ha de realizar una determinada acción (recoger o entregar) para resolver la instancia del problema.
- **Superficie:** El subespacio formado por el conjunto de todas las posiciones que forman la instancia del problema. Este es quien define la métrica de distancia (tanto espacial como temporal) entre las distintas posiciones del problemas.
- **Servicio:** Está compuesto por una posición, un tiempo para ser efectuado, y una ventana temporal.

- **Viaje:** Está compuesto por los servicios de recogida y entrega de una mercancía. Además contiene información necesaria para poder proceder a su realización, es decir, la capacidad (o el tamaño de la carga a desplazar), el tiempo máximo admisible desde la recogida hasta la entrega el tiempo de servicio en las posiciones de recogida y entrega, etc. En la literatura es común diferenciar entre dos tipos de viajes a partir de las ventanas temporales. Cuando estas se fijan sobre la posición de recogida se denominan *outbound* mientras que cuando se hacen sobre la posición de entrega son denominadas *inbound*. Sin embargo, esto es simplemente terminología del problema ya que desde un punto de vista de la formulación del modelo ambos conceptos podrían ser aplicados a la vez sobre un viaje si se fijan de manera pertinente las ventanas temporales.
- **Trabajo:** Representa el conjunto de todos los viajes a realizar durante una determinada instancia del problema.
- **Vehículo:** Representa el actor encargado de llevar a cabo un subconjunto de viajes que permiten resolver una instancia del problema de manera parcial (o completa). Se define por una posición inicial, así como otra final, así como por una determinada capacidad máxima y un tiempo máximo de servicio. Además, tiene asociada una determinada ventana temporal.
- **Flota:** Está compuesta por el conjunto de todos los vehículos disponibles para resolver la instancia del problema. En ciertas ocasiones el tamaño de la flota de vehículos es fija, mientras que en otras ocasiones se permite que la instancia del problema sea planificada con tantos vehículos como sea posible. Dichas ideas se discutirán posteriormente.
- **Parada:** Representa la acción de visitar una determinada posición por un vehículo. Esta está compuesta por el instante de tiempo en que se visita así como la razón, la cual puede ser una recogida o entrega referida a un determinado viaje.
- **Viaje Planificado:** Representa el enlace entre un viaje y sus dos correspondientes paradas (la de recogida y la de entrega), que (como se ha dicho previamente) mantienen los instantes de tiempo en que sucede.
- **Ruta:** Representa una serie de viajes planificados por un determinado vehículo.
- **Planificación:** Está formada por el conjunto de rutas que permiten resolver la instancia del problema.
- **Resultado:** Está compuesta por la planificación que resuelve parcial (o totalmente) la instancia del problema, junto con la correspondiente flota de vehículos y el trabajo formado por un conjunto de viajes a realizar.

Una vez, descritas las palabras clave más importantes es fácil darse cuenta de que los conceptos más importantes para la resolución del *problema Dial-a-Ride* son el de *viaje* y el de *ruta*, ya que son quienes contienen la información tanto del qué resolver (*viae*), como del cómo resolver (*ruta*).

2.3.2. Constantes

El siguiente paso antes de proceder a la descripción del *problema Dial-a-Ride* de manera matemática a través de su formulación es definir el significado de las constantes que se utilizan en el mismo. A continuación se incluye el listado junto con su correspondiente definición.

- P : Conjunto de nodos generado por las posiciones referidas a recogidas. Esto es el conjunto formado por la secuencia $\{1, 2, \dots, n\}$.
- D : Conjunto de nodos generado por las posiciones referidas a entregas. Esto es el conjunto formado por la secuencia $\{n + 1, n + 2, \dots, 2n\}$.
- N : Conjunto de nodos generado por todas las posiciones de la superficie. Esto es el conjunto formado por $\{0, 2n + 1\} \cup P \cup D$.
- $A_{(i,j)}$: El arista que se encarga de relacionar el nodo definidos por la posición i -ésima con el nodo definido por la posición j -ésima. Nótese por tanto que inicialmente la cardinalidad es de $\text{card}(N)^2$. Sin embargo, asumiendo que no se pueda dar la situación de visitar un nodo desde si mismo (para respetar las restricciones de unicidad) este número se reduce a $\text{card}(N)(\text{card}(N) - 1)$. Sin embargo, es fácil darse cuenta de que muchos de los aristas serán marcados como “prohibidos” por las restricciones del modelo (No se puede visitar una posición de entrega antes que su correspondiente posición de recogida).
- $G = (N, A)$: El grafo generado por el conjunto de nodos sobre el conjunto de aristas, sobre el cual se apoya toda la planificación.
- K_k : Conjunto de índices generados por los vehículos disponibles para resolver el problema.
- c_{ij} : Coste por moverse del nodo i -ésimo al nodo j -ésimo.
- t_{ij} : Tiempo para moverse del nodo i -ésimo al nodo j -ésimo.
- e_i : Cota inferior de la ventana temporal sobre el nodo i -ésimo.
- l_i : Cota superior de la ventana temporal sobre el nodo i -ésimo.
- d_i : Tiempo de servicio sobre el nodo i -ésimo.
- q_i : Capacidad del servicio sobre el nodo i -ésimo. Nótese que para los nodos 0 y $2n + 1$ la capacidad a servir deberá ser 0 mientras que se deberá cumplir que $q_i \geq 0$ si $i \in P$ y $q_i = q_{i-n}$ si $i \in D$.

- T_k : Duración máxima de ruta por el vehículo k -ésimo.
- Q_k : Capacidad máxima del vehículo k -ésimo.
- E_k : Cota inferior de la ventana temporal sobre el vehículo k -ésimo.
- L_k : Cota superior de la ventana temporal sobre el vehículo k -ésimo.

Nótese la equivalencia conceptual entre las definiciones descritas en este apartado con respecto al apartado anterior (de *Palabras clave*). Se ha decidido describir los dos enfoques para permitir una mayor versatilidad (y comprensión) a lo largo del documento.

2.3.3. Variables de decisión

Por último, se procede a definir el significado de las variables que se utilizan en el mismo. A continuación se incluye el listado junto con su correspondiente definición. Antes de nada es importante remarcar que existen dos formulaciones populares: basadas en 3 índices (la cual se seguirá en la formulación básica) y basadas en 2 índices (la cual se expone en futuros apartados). En este caso se describe la de 3 índices por ser más general. A pesar de ello es trivial entender la simplificación cuando se trate de la de 2 índices.

- x_{ij}^k : Variable de naturaleza binaria que toma el valor 1 cuando existe un enlace entre los nodos i -ésimo y j -ésimo llevado a cabo por el vehículo k -ésimo
- u_i^k : Variable de decisión continua que representa el instante de tiempo en que comienza la ruta generada por el vehículo k -ésimo llega al nodo i -ésimo.
- w_i^k : Variable de decisión continua que representa la capacidad ocupada en la ruta generada por el vehículo k -ésimo en el nodo i -ésimo.

A pesar de que dichas variables de decisión serán las utilizadas principalmente durante el proceso de formulación de los problemas, puede darse el caso de que en algunas situaciones concretas (para tratar de corregir la linealidad de restricciones, simplificarlas o algún paso similar, se utilicen variables diferentes).

Para concluir el apartado se quiere remarcar lo siguiente: Tal y como se ha repetido en varias ocasiones a lo largo del mismo, es de vital importancia definir de manera adecuada la notación utilizada para así simplificar en la medida de lo posible la aclaración de los conceptos que se construyen posteriormente a partir de la misma. En este caso se pretende seguir una notación común que después sea aplicable a la definición de los distintos métodos de resolución. Por lo tanto se ha preferido llevar a cabo de la manera más general posible.

2.4. Formulación básica

Tras contextualizar de una manera detallada el *problema Dial-a-Ride*, tanto desde el punto de vista de la situación que pretende resolver, partiendo desde una base muy general (*Problemas de Optimización* en el apartado 2.2.1) hasta las características concretas del mismo (*Problemas Dial-a-Ride* en el apartado 2.2.6), así como la descripción de la notación que se utiliza a lo largo del documento, tanto desde una perspectiva de vocabulario como matemática, ya se está en condiciones suficientes como para presentar la formulación específica para dicho problema. En este caso, se describe la formulación básica (u original) formada por 3 índices en el tensor tridimensional que contine las variables de decisión binarias x_{ij}^k . En la literatura existen distintas representaciones de la misma, entre las que se encuentran [Cor06; RCL07; CL07] entre otros. Sin embargo, en este caso se ha preferido mostrar una versión más cercana a la formulación como modelo lineal, simplificando previamente las restricciones no lineales, las variables redundantes etc.

$$\begin{aligned}
& \text{Minimizar} && \sum_{k \in K} \sum_{j \in V} \sum_{i \in V} c_{ij}^k x_{ij}^k \\
& \text{sujeto a} && \sum_{k \in K} \sum_{j \in V} x_{ij}^k = 1, && \forall i \in P \\
& && \sum_{i \in V} x_{0i}^k = 1, && \forall k \in K \\
& && \sum_{i \in V} x_{i,2n+1}^k = 1, && \forall k \in K \\
& && \sum_{j \in V} x_{ij}^k - \sum_{j \in V} x_{n+i,j}^k = 0, && \forall i \in P, \forall k \in K \\
& && \sum_{j \in V} x_{ji}^k - \sum_{j \in V} x_{ij}^k = 0, && \forall i \in P \cup D, \forall k \in K \\
& && u_i^k + d_i + t_{ij} - M_{ij}^k(1 - x_{ij}^k) \leq u_j^k, && \forall i, j \in V, \forall k \in K \\
& && w_i^k + q_j - W_{ij}^k(1 - x_{ij}^k) \leq w_j^k, && \forall i, j \in V, \forall k \in K \\
& && u_{2n+1}^k - u_0^k \leq T_k, && \forall k \in K \\
& && e_i \leq u_i^k, && \forall i \in V, \forall k \in K \\
& && u_i^k \leq l_i, && \forall i \in V, \forall k \in K \\
& && t_{i,n+i}^k \leq u_{n+i}^k - (u_i^k + d_i), && \forall i \in P, \forall k \in K \\
& && u_{n+i}^k - (u_i^k + d_i) \leq L_k, && \forall i \in P, \forall k \in K \\
& && \max\{0, q_i\} \leq w_i^k, && \forall i \in V, \forall k \in K \\
& && w_i^k \leq Q_k + \min\{0, q_i\}, && \forall i \in V, \forall k \in K \\
& && x_{ij}^k \in \{0, 1\}, && \forall i, j \in V, \forall k \in K
\end{aligned} \tag{2.5}$$

Ecuación 2.5: Formulación del modelo basado en 3 índices para el *Problema Dial-a-Ride*.

En la ecuación 2.5 se muestran tanto las restricciones como la función objetivo del problema. En este último caso, es sencillo darse cuenta de que el valor a optimizar en el modelo básico es el coste total de la planificación, que viene dado por la acumulación de costes para cada enlace entre posiciones generadas por los servicios del problema. Esto se puede ver de forma sencilla en la ecuación (2.6).

$$\sum_{k \in K} \sum_{j \in V} \sum_{i \in V} c_{ij}^k x_{ij}^k \quad (2.6)$$

En cuanto a las restricciones, estas pueden ser agrupadas en tres grupos principales: restricciones de unicidad (apartado 2.4.1), de conectividad (apartado 2.4.2), de acumulación (apartado 2.4.3), y de tiempo y capacidad (apartado 2.4.4). A continuación se describen un poco más en detalle los distintos grupos de restricciones.

2.4.1. Restricciones de Unicidad

Tal y como se ha indicado previamente en el apartado 2.2.4, las restricciones de unicidad en los problemas de rutado son las encargadas de asegurarse de que cada servicio sea realizada una única vez, así como que sea realizada obligatoriamente. En este caso, se pretende además de que distintos vehículos no realicen el mismo servicio. Dicha restricción se muestra en la ecuación (2.7). Es interesante darse cuenta de que esta restricción tan solo se aplica a los servicios correspondientes a recogidas. La razón de ello, es que por transitividad procedente de otro conjunto de restricciones sería redundante.

Dicho conjunto additional de restricciones se refiere a la necesidad de que si una ruta visita la posición correspondiente a un servicio de recogida, entonces será necesario que también visite la posición del correspondiente servicio de entrega. Esto es necesario en el *problema Dial-a-Ride* ya que tal y como se comentó previamente, en este caso las mercancías no son intercambiables entre distintos pares de recogida/entrega. La restricción correspondiente se describe en la ecuación (2.8).

$$\sum_{k \in K} \sum_{j \in V} x_{ij}^k = 1, \quad \forall i \in P \quad (2.7)$$

$$\sum_{j \in V} x_{ij}^k - \sum_{j \in V} x_{n+i,j}^k = 0, \quad \forall i \in P, \forall k \in K \quad (2.8)$$

Una vez añadidas dichas restricciones, ya se puede asegurar que para llegar a una solución factible del problema, todas las posiciones en que se localizan sus correspondientes servicios hayan sido visitadas, así como que las posiciones de inicio y fin de ruta se cumplan.

2.4.2. Restricciones de Secuenciación

Las restricciones de secuenciación (las cuales se comentaron brevemente en el apartado 2.2.4) se refieren a la necesidad de que tras cada parada en un servicio, únicamente se pueda continuar hacia otro servicio. Es decir, que no surjan problemas relacionados con la secuencialidad de la ruta. Esto se lleva a cabo a partir de la restricción descrita en la ecuación (2.9). Nótese que esto tan solo se aplica a las posiciones relacionadas con servicios, ya que las posiciones referidas al inicio y fin de ruta, el comportamiento ha de ser diferente.

Dicho comportamiento requiere que las posiciones de inicio y fin de ruta sean visitadas por todas las rutas (por lo que no se aplica ni la unicidad ni la secuencialidad). Sin embargo, es necesario ser capaz de controlar las posiciones de origen y destino de la planificación (que típicamente coinciden aunque podría no ser así). Asumiendo que todas las rutas comienzan y terminan en las posiciones 0 y $2n + 1$ respectivamente, es necesario añadir las restricciones descritas en las ecuaciones (2.10) y (2.11).

$$\sum_{j \in V} x_{ji}^k - \sum_{j \in V} x_{ij}^k = 0, \quad \forall i \in P \cup D, \forall k \in K \quad (2.9)$$

$$\sum_{i \in V} x_{0i}^k = 1, \quad \forall k \in K \quad (2.10)$$

$$\sum_{i \in V} x_{i,2n+1}^k = 1, \quad \forall k \in K \quad (2.11)$$

Una vez incluidas las restricciones de secuencialidad, ya es posible conseguir rutas válidas desde el punto de vista de las posiciones que se visitan y el orden entre estas. Sin embargo, estas no tienen en cuenta ni tiempos ni capacidades por lo que es necesario añadir un conjunto de restricciones adicionales para poder modelar de manera completa el *problema Dial-a-Ride*.

2.4.3. Restricciones de Acumulación

En problemas basados en el mantenimiento de un cierto nivel de estado para la planificación, donde este se basa en incrementos y decrementos sobre determinadas variables, es necesario incluir distintas restricciones que “asignen” a las variables relacionadas sus correspondientes valores. En el caso del *problema Dial-a-Ride* es necesario mantener la información tanto del instante de tiempo como de la capacidad con que se llega a la posición de un determinado servicio. Dichas restricciones pueden ser expresadas de distintas formas, siendo algunas de ellas muy intuitivas pero no lineales. En este caso, se incluye la versión lineal tanto para tiempos como para capacidades en las ecuaciones (2.12) y (2.13) respectivamente.

$$u_i^k + d_i + t_{ij} - M_{ij}^k(1 - x_{ij}^k) \leq u_j^k, \quad \forall i, j \in V, \forall k \in K \quad (2.12)$$

$$w_i^k + q_j - W_{ij}^k(1 - x_{ij}^k) \leq w_j^k, \quad \forall i, j \in V, \forall k \in K \quad (2.13)$$

Las restricciones inicialmente no lineales pueden ser vistas como implicaciones en el sentido siguiente: “Si se utiliza el enlace entre las posiciones i -ésima y j -ésima, entonces el incremento será de y ”. Para linearizar este tipo de restricciones sobre variables de decisión de carácter binario, se puede utilizar una cota superior. En [Cor06] se indica que los valores descritos en las ecuaciones (2.14) y (2.15) son suficientes para tiempos y capacidad respectivamente.

$$M_{ij}^k = \max\{0, l_i + d_i + t_{ij} - e_j\} \quad (2.14)$$

$$W_{ij}^k = Q_k + \min\{0, q_i\} \quad (2.15)$$

Las restricciones definidas en este apartado permiten mantener el estado a nivel de tiempo y capacidad para cada ruta en cada parada de la misma. Sin embargo, estas no aportan ninguna restricción sobre la lógica del problema. A pesar de ello, es necesario incluirlas en la formulación para poder apoyarnos en ellas durante la definición de las restricciones propiamente relacionadas con la lógica del problema.

2.4.4. Restricciones de Tiempo y Capacidad

Una vez descritas todas aquellas restricciones que permiten mantener de manera coherente tanto la unicidad y secuenciación de paradas, como el estado a lo largo de las rutas, se pueden añadir las restricciones más específicas y que utilizan al resto como base.

$$u_{2n+1}^k - u_0^k \leq T_k, \quad \forall k \in K \quad (2.16)$$

$$e_i \leq u_i^k, \quad \forall i \in V, \forall k \in K \quad (2.17)$$

$$u_i^k \leq l_i, \quad \forall i \in V, \forall k \in K \quad (2.18)$$

$$t_{i,n+i}^k \leq u_{n+i}^k - (u_i^k + d_i), \quad \forall i \in P, \forall k \in K \quad (2.19)$$

$$u_{n+i}^k - (u_i^k + d_i) \leq L_k, \quad \forall i \in P, \forall k \in K \quad (2.20)$$

$$\max\{0, q_i\} \leq w_i^k, \quad \forall i \in V, \forall k \in K \quad (2.21)$$

$$w_i^k \leq Q_k + \min\{0, q_i\}, \quad \forall i \in V, \forall k \in K \quad (2.22)$$

En primer lugar, es necesario fijar un tiempo máximo de ruta, lo cual se puede llevar a cabo apoyándose en las variables de estado de tiempo, mediante la restricción descrita en la ecuación (2.16). De manera similar, para respetar las ventanas temporales de llegada a cada posición para poder

cumplir con un servicio es necesario utilizar las restricciones definidas en las ecuaciones (2.17) y (2.18) que controlan los tiempo más tempranos y tardíos de llegada respectivamente.

La restricción que diferencia el *problema de recogidas y entregas* del *problema Dial-a-Ride* consiste en la fijación de un tiempo máximo de trayecto por cada viaje. Esto es el tiempo que pasa desde el servicio de recogida hasta el de entrega para un viaje concreto. Las restricciones correspondientes se incluye en las ecuaciones (2.19) y (2.20).

Finalmente, es necesario restringir de manera apropiada la capacidad máxima que el vehículo puede transportar durante cada etapa de la ruta. Para ello, se han de actualizar de manera coherente tanto las cargas como las descargas, lo cual queda resuelto con las restricciones de acumulación descritas anteriormente. Por lo tanto, en este caso, tan solo es necesario asegurar que la capacidad resultante tras cada servicio siempre se encuentre dentro del intervalo válido $([0, Q_k])$. Esto se puede llevar a cabo de manera sencilla mediante el uso de las restricciones definidas en las ecuaciones (2.21) y (2.22).

[TODO]

2.5. Otras formulaciones

[TODO]

2.5.1. Formulación de 2 índices (1)

[TODO]

Formulación basada en [RCL07]

2.5.2. Formulación de 2 índices (2)

[TODO]

Formulación basada en [RCL07]

2.6. Extensiones

[TODO]

2.6.1. Extensiones basadas en Restricciones

[TODO]

1. **Heterogeneidad:** [TODO]
2. **Propiedades de las Rutas:** [TODO]
3. **Calidad del Viaje:** [TODO]

[TODO]

2.6.2. Extensiones basadas en la Función Objetivo

[TODO]

2.6.2.1. Funciones Objetivo Simples

[TODO]

1. **Dial a Ride:** [TODO]
2. **Pickup and Delivery:** [TODO]
3. **Taxi Sharing:** [TODO]

2.6.2.2. Funciones Multiobjetivo

[TODO]

[TODO]

2.6.3. Modelos para vehículos eléctricos

[TODO]

Basado en [EM12].

2.6.4. Modelos para vehículos autónomos

Basado en [Dia95; BKG19].

2.7. Tiempo Real

[TODO]

2.8. Conclusiones

[TODO]

Capítulo 3

Métodos de Resolución

3.1. Introducción

Los problemas de *optimización combinatoria* constituyen una de las categorías más interesantes dentro del área de la optimización de variables. Esto se debe a la gran aplicabilidad de los mismos en situaciones de la vida real, así como la gran reducción de costes que se puede alcanzar cuando estos son aplicados en los puntos estratégicos de cualquier proceso de producción.

Sin embargo, esta categoría de problemas de optimización presenta un mayor grado de dificultad en su resolución, tal y como se ha indicado a lo largo del capítulo 2. Entre otros, esto se debe a la imposibilidad de utilizar técnicas basadas en gradientes (que en problemas con variables de decisión continuas proporcionan simplificar mucho su resolución). Puesto que estas técnicas no son aplicables, la estrategia alternativa se basa en la enumeración de posibles configuraciones de variables de manera inteligente hasta alcanzar aquella que genere el resultado óptimo para la instancia del problema que se pretenda resolver.

Por tanto, a pesar de no ser posible el apoyo en técnicas basadas en gradientes, si que existe la alternativa basada en enumeración de configuraciones como enfoque para dar con el óptimo. A pesar de ello, es fácil darse cuenta de que esta estrategia no es lo suficientemente potente como para permitir resolver problemas reales (o incluso de pequeño tamaño). La dificultad radica en la explosión combinatoria generada por la enumeración de soluciones. Por ejemplo, en un problema compuesto por 20 variables de decisión de naturaleza binaria, el número de configuraciones a evaluar alcanzaría el valor de $2^{20} = 1048576$, lo cual no es escalable a problemas de gran tamaño, donde hay cientos o miles de variables de decisión con las que trabajar.

A pesar de esta dificultad, es posible enfrentarse a problemas de *optimización combinatoria* de tamaños relativamente grande teniendo en cuenta distintas estrategias que permiten ignorar un gran número de configuraciones no factibles dadas las restricciones descritas por el problema en cuestión. Es por ello que muchos de los métodos de resolución se apoyan en estas garantías para

enfocar la búsqueda sobre un subespacio de configuraciones factibles. Otra de las ideas más utilizadas por los métodos de resolución es apoyarse en los resultados de evaluación de configuraciones anteriores para que el coste computacional de evaluación de configuraciones actuales sea mucho menor. Dicha estrategia algorítmica se conoce como *Programación Dinámica*, la cual se describe en mayor detalle en [Bel54].

A lo largo del capítulo se describen distintas situaciones en que es posible aplicar las técnicas anteriormente descritas sobre el *problema Dial-a-Ride* para tratar de reducir en la medida de lo posible la complejidad computacional que este presenta. Entonces, el resto del capítulo se organiza de la siguiente manera: En el apartado 3.2 se describen las distintas técnicas de resolución exactas que la literatura especializada en el problema a probado para tratar de resolver el problema. Seguidamente, en el apartado 3.3 se describen estrategia basadas en heurísticas (tanto de construcción como mejora de soluciones) que a pesar de no ofrecer ninguna garantía de optimalidad, ofrecen un buen equilibrio entre tiempo de cómputo y calidad de las soluciones. Después, en el apartado 3.4 se describen un conjunto de metaheurísticas basadas en la utilización de las heurísticas descritas en el apartado anterior de manera inteligente de tal manera que las configuraciones generadas proporcionan soluciones de mayor calidad. Por último, en el apartado 3.5 se incluyen unas conclusiones generales acerca de los métodos de resolución aplicados al *problema Dial-a-Ride*.

3.2. Métodos de Resolución Exactos

Los métodos de resolución exactos representan una de las categorías de resolución de problemas de optimización más interesantes debido a distintos factores, entre los que se encuentran aquellas situaciones en que es estrictamente necesario obtener el valor óptimo para una instancia dada. Sin embargo, esta categoría no solo es interesante por dicha razón, sino que el estudio de estos problemas además proporciona en muchas ocasiones razonamientos y justificaciones teóricas sobre el problema que pretenden resolver que ayudan a comprender mejor los factores del mismo. Lo interesante de este punto es que esto sirve de utilidad para el diseño de diferentes métodos de resolución así como su justificación.

Tal y como se ha descrito anteriormente, el método de resolución exacta más básico consiste en la enumeración de todas las configuraciones posibles dadas la variables de decisión del problema para después determinar como configuración óptima aquella que maximiza/minimiza la función objetivo. Esto es teóricamente posible por la naturaleza combinatoria del problema pero a la vez prácticamente imposible por la misma razón.

Puesto que los problemas de *optimización combinatoria* se pueden modelar como problemas de *optimización lineal* mediante un conjunto de restricciones de naturaleza lineal, la siguiente alternativa a la enumeración de configuraciones es la resolución como un problema de esta naturaleza.

Para ello, una de las opciones es la utilización del algoritmo *Simplex* [KM70] tal y como se indica en el apartado 2.2.2. Como se ha descrito anteriormente, este algoritmo se basa en la búsqueda de la mejor solución contigua a la actual de manera iterativa hasta alcanzar aquella que no tenga una solución contigua mejor. Es posible demostrar que este algoritmo alcanza el valor óptimo cuando se aplica sobre problema de *optimización lineal*. Tal y como se puede apreciar en la figura 3.1, gráficamente puede ser visto como la evaluación de vértices en un grafo de soluciones. La complejidad que esta estrategia presenta en problemas de *optimización combinatoria* reside en el gran número de vértices que surgen del cambio de valor de cada una de las variables de decisión del problema, generando un grafo con gran número de vértices y aristas lo cual hace inalcanzable computacionalmente navegar por el mismo hasta alcanzar la solución óptima.

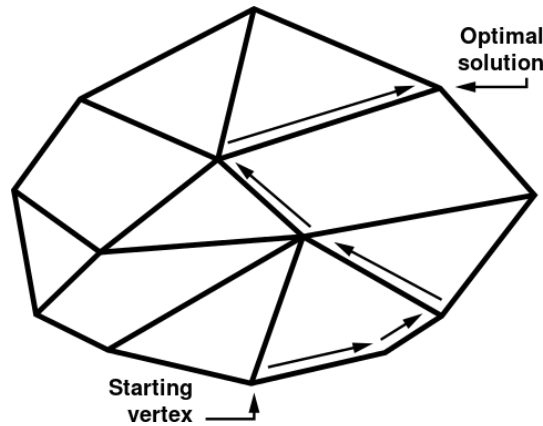


Figura 3.1: Representación gráfica del modo de ejecución del algoritmo *Simplex*.

Hasta ahora, en este apartado se ha descrito únicamente el proceso de resolución mediante algunos métodos exactos. Sin embargo, no se ha realizado ninguna particularización sobre el problema *Dial-a-Ride*. Por tanto, es momento de definir el punto de enlace entre la definición del método y el problema en cuestión. Tal y como se ha descrito a lo largo del capítulo 2 el problema *Dial-a-Ride* se enmarca dentro de los problemas de *optimización lineal*. En la ecuación 2.5 se describe la formulación de 3 índices para este problema, incluyendo tanto las variables de decisión como las restricciones que lo componen. En cuanto a la estrategia de enumeración, es fácil darse cuenta de que esta puede llevarse a cabo tomando todas las combinaciones posibles para las variables binarias del problema (lo cual tal y como se ha comentado anteriormente es algo inabordable en problemas no demostrativos). En cuanto a la estrategia de resolución basada en el algoritmo *Simplex*, tal y como se ilustra en [KM70] es posible construir la matriz simplex a partir de la matriz de restricciones generada por el problema en forma lineal.

A lo largo de la literatura relacionada con métodos de resolución exactos para el problema *Dial-a-Ride*, existen un gran número de referencias basadas en el estudio de técnicas basadas en *Branch and Bound* que tratan de tomar ventaja de alguna manera de la estructura del problema para así

simplificar el camino hasta la solución óptima. En el siguiente apartado se describen de manera más detallada los conceptos en que se apoya este método.

3.2.1. Branch and Bound

Se pueden englobar en la categoría de algoritmos *Branch and Bound* todos aquellos que basan su estrategia de resolución en comenzar por una versión relajada del problema base para posteriormente realizar un proceso jerárquico de cambios sobre la estructura del problema hasta alcanzar el valor óptimo del mismo. Esta estrategia es muy usada sobre problemas de *optimización entera* ya que facilita en gran medida la problemática generada por las variables de decisión de carácter discreto del problema.

La estrategia habitual consiste en partir de una versión del problema sobre la cual las restricciones sobre variables discretas se han relajado a valores sobre intervalos. Es decir, supongamos que la variable de decisión x únicamente puede tomar valores enteros entre 1 y 5, lo que es equivalente a $x \in \{1, 2, 3, 4, 5\}$. Entonces, una relajación lineal podría venir dada por la transformación en $1 \leq x \leq 5$ la cual es más sencilla de resolver por el algoritmo *Simplex*. Posteriormente, el resto de pasos de ramificación consisten en añadir restricciones (mediante la generación de filas o columnas) que permitan satisfacer las restricciones del problema original.

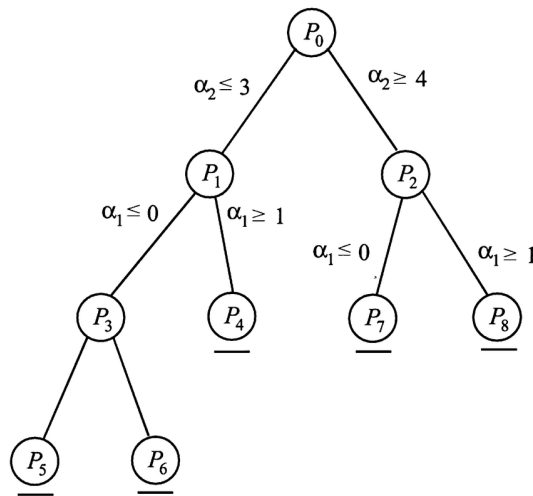


Figura 3.2: Ejemplo de ramificación del problema mediante el algoritmo *Branch and Bound*.

Es importante tener en cuenta que las relajaciones aplicadas al problema inicial para aplicar el algoritmo *Branch and Bound* no tienen por qué ser de naturaleza lineal, si no que se pueden aplicar sobre otros tipos de restricciones. Un ejemplo de ello se podría dar eliminando por ejemplo las restricciones de duración máximo de viaje en el problema *Dial-a-Ride*, o la eliminación de restricciones de ventana temporal en el problema *Pickup and Delivery with Time Windows*. Una vez

generada una solución factible sobre estos casos, el proceso de ramificación consistirá en “modificar” dicha solución hasta alcanzar aquella que cumpla todas las restricciones (de tiempo máximo de viaje, de ventanas temporales, etc.).

Dependiendo de la estrategia de ramificación del algoritmo, este se denomina de distintas maneras. Cuando la ramificación se basa en la generación de filas el algoritmo se conoce como *Branch and Cut*, mientras que si la estrategia se basa en la generación de columnas este se denomina *Branch and Price*. Por último, cuando la estrategia se basa en la combinación de ambos entonces el algoritmo es denominado *Branch and Cut and Price*.

3.2.1.1. Branch and Cut

Tal y como se ha comentado anteriormente, los algoritmos *Branch and Cut* consisten en una caracterización del algoritmo general de ramificación donde el proceso de “estrechamiento” de la versión relajada para cumplir con todas las restricciones impuestas por el problema viene dado por la adición de planos de corte a partir de nuevas restricciones. Es por ello que este método se apoya en estrategias de generación de filas.

Para el caso del problema *Dial-a-Ride*, en [Cor06] se propone una implementación completa basada en el algoritmo *Branch and Cut* la cual se apoya en la generación inicial de una solución del problema relajado (para el cual se eliminan las restricciones que imponen variables de decisión binarias, permitiendo que estas tomen cualquier valor en el intervalo $[0, 1]$) para posteriormente, a partir de un proceso de ramificación añadir restricciones adicionales que fuerzan a las variables de naturaleza binaria con valores no binarios a tomar un valor válido (ahí surge el proceso de ramificación). Para detectar sobre qué variable concreta generar el corte en cada caso, la estrategia seguida se apoya en la definición de unas heurísticas que generan una lista ordenada sobre la cual seleccionar dicha variable. En la línea de este enfoque, también es interesante remarcar [RCL07] donde además se propone un modelo de programación lineal basado en 2 índices, que a pesar de no representar el mismo caso que el problema *Dial-a-Ride*, sus diferencias son mínimas (número indeterminado de vehículos, todos los vehículos con misma capacidad, etc.).

3.2.1.2. Branch and Price

Del mismo modo que el método *Branch and Cut* es una estrategia de generación de filas enmarcada dentro de la familia de algoritmos *Branch and Bound*, el método *Branch and Price* es la versión correspondiente cuya estrategia se basa en la generación de columnas sobre la matriz de restricciones del problema de programación lineal. Es decir, se basa en la inclusión de nuevas variables sobre una versión relajada del problema inicial. Puesto que puede dar lugar a confusión, es necesario detallar cómo se puede llevar esto a cabo para reflejar de manera adecuada la diferencia respecto del método de generación de filas. Mientras que en la estrategia de generación de filas,

la relajación se produce permitiendo que las variables de decisión tomen valores que no satisfacen las restricciones originales, en este caso el enfoque consiste en resolver una versión más simple del problema original para después, mediante un proceso iterativo, ampliar el alcance del problema simplificado hasta alcanzar la definición del mismo que satisfaga todas las restricciones de la versión original. El proceso de ramificación surge cuando no es posible satisfacer las condiciones impuestas tras la inclusión de nuevas variables, donde esta es forzada con la esperanza de que sea satisfecha en la siguiente iteración, creando una ramificación por cada posible alternativa.

Aplicado al problema *Dial-a-Ride*, una posible estrategia se apoya en limitar los aristas disponibles del grafo en una primera iteración, para posteriormente aplicar un proceso inclusión/eliminación de aristas del grafo para así controlar la complejidad del problema a la vez que se exploran distintas soluciones. El proceso de ramificación de soluciones surge cuando son eliminados aristas que en alguna de las soluciones hayan sido utilizados. Para llevar a cabo dicho proceso de inclusión/eliminación en la literatura se han propuesto estrategias apoyadas en *Programación dinámica*.

3.2.1.3. Branch and Cut and Price

Una vez descritos los métodos basados en generación de filas (*Branch and Cut*) así como de columnas (*Branch and Price*), es fácil entender el funcionamiento del método *Branch and Cut and Price*, que se apoya en la combinación de ambas estrategias para reducir así la complejidad de resolución del problema *Dial-a-Ride*. Como es natural, estos cuentan con la ventaja de resolver instancias más fáciles de resolver del problema por tener un menor número de restricciones (eliminación de filas) así como de hacerlo sobre versiones más pequeñas del problema (eliminación de columnas). Tal y como se ha podido comprobar en distintas publicaciones que comparan la utilización de los métodos de generación de filas o columnas de manera separada, frente a las correspondientes versiones combinadas, estas últimas obtienen un nivel de resultados equivalente requiriendo de un menor nivel de cómputo.

Tal y como se ha podido comprobar a lo largo del apartado, los métodos de resolución apoyados por técnicas basadas en ramificación son capaces de ofrecer ventajas relacionadas con la simplificación del problema completo a resolver a costa de requerir la resolución de múltiples versiones relajadas de distintas formas de dicho problema. Cuando el tamaño de la instancia del problema es tan grande que esta no puede ser resuelta “de una sola vez”, estos métodos ofrecen una alternativa interesante por seguir ofreciendo garantías de optimalidad, lo cual en ciertos casos es un requisito necesario.

3.3. Métodos de Resolución basados en Heurísticas

Tal y como se ha explicado anteriormente, los problemas de optimización combinatoria se corresponden en la mayoría de ocasiones con problemas cuya complejidad desde el punto de vista de la resolución no es sencilla. Tal es así que muchos de estos se engloban dentro de la categoría de problemas *NP*, lo cual se puede resumir como aquellos para los cuales no existe una estrategia de resolución exacta que alcance el resultado óptimo en tiempo polinómico. Esto implica que para instancias no demasiado grandes no sea factible alcanzar dicho resultado en un tiempo aceptable.

Para paliar esta problemática, surgen métodos de resolución denominados *estrategias heurísticas*. Estas se caracterizan por ser capaces de proporcionar resultados razonablemente cercanos a solución óptima empleando para ello una cantidad de recursos (generalmente tiempo de cómputo) significativamente menor a la que requerirían otras estrategias con garantías de optimalidad. Este es un compromiso aceptable cuando no es posible llegar a soluciones de otro modo. Por lo tanto, uno de los grandes objetivos de la investigación y el desarrollo de heurísticas consiste en implementar algoritmos que sean capaces de alcanzar soluciones lo más cercanas a la solución óptima, teniendo en cuenta las restricciones de tiempo para que estas puedan ser alcanzadas.

Desde una perspectiva de alto nivel, los métodos de resolución basados en heurísticas típicamente son clasificados en dos grandes categorías según la función que desempeñan, las cuales se conocen como *heurísticas de construcción* y *heurística de mejora*. A pesar de que su nombre es lo suficientemente explicativo, a continuación se detalla la función que representan los métodos pertenecientes a cada una de estas categorías:

- **Heurísticas de Construcción:** la función que desempeñan consiste en la generación de soluciones sin el apoyo de ningún estado previo. Es decir, recibiendo únicamente los valores de entrada necesarios para la formulación del problema, estas son capaces de generar soluciones del problema que satisfacen todas las restricciones impuestas por la formulación del mismo. En el caso de los problemas de rutas de vehículos, estas estrategias se apoyan comúnmente en algoritmos de inserción inspirados en algoritmos greedy y, en algunas ocasiones con un cierto nivel de aleatoriedad. Por lo general, los resultados a los que es posible llegar utilizando únicamente estrategias de este tipo, no son lo suficientemente cercanas al resultado óptimo. Para resolver esta problemática, existen otras estrategias que tratan de paliar esta situación.
- **Heurísticas de Mejora:** debido al grado de dificultad necesario para tomar las decisiones más acertadas durante el proceso de construcción de soluciones, la categoría de *heurísticas de mejora* trata de mejorar la calidad una solución dada. Para ello, como es esperable, estas estrategias necesitan conocer tanto el conjunto de valores de entrada necesarios para la formulación del problema, como una solución inicial del mismo. A partir de estas dos partes, el modo de funcionamiento consiste en aplicar distintas modificaciones a la solución para

así tratar de maximizar (o minimizar) el valor de la función objetivo, lo cual acerca dicha solución al valor óptimo deseado. A este modo de funcionamiento se le conoce como búsqueda local dentro del espacio de soluciones. En el caso de los problemas de rutas de vehículos, las estrategias en que se basan consisten en la aplicación de operadores tales como el intercambio de tareas (o viajes) entre distintos vehículos, cambio del orden dentro del mismo vehículo, etc. Por tanto, el objetivo se trata de buscar el conjunto de transformaciones de la solución que permitan un mayor acercamiento al resultado esperado al final de la ejecución. Esta última afirmación se explicará más en detalle en el apartado 3.3.3

Una vez se han detallado las diferencias entre las dos categorías principales de estrategias de resolución basados en heurísticas, se está en condiciones suficientes como para explicar en detalle algunos de los algoritmos concretos a través de los cuales es posible alcanzar dichas soluciones. El resto del apartado se estructura de la siguiente forma: en el apartado 3.3.1 se detalla la estrategia basada en *fuerza bruta* (la cual como veremos no es aplicable en la práctica en la gran mayoría de casos). Seguidamente, en el apartado 3.3.2 está dedicado a la estrategia basada en *selección aleatoria*. Después se describirá en detalle la estrategia basada en *algoritmos greedy* (la cual es muy popular por su sencillez conceptual y relativamente buenos resultados). Finalmente se discutirá el funcionamiento de la estrategia de *clarke and wright* y un método basado en *redes neuronales* en los apartados 3.3.4 y 3.3.5

3.3.1. Fuerza Bruta

A pesar de que anteriormente se ha comentado que una de las grandes diferencias entre métodos de resolución exactos y métodos de resolución basados en heurísticas son las garantías de optimalidad, esta distinción es un tanto difusa cuando se tratan de clasificar las estrategias de fuerza bruta. Esto se debe a que, si bien es cierto que esta estrategia es capaz de garantizar resultados óptimos cuando termina, la estrategia de resolución se encuentra más próxima a los algoritmos heurísticos que a los métodos que hemos clasificado como exactos, puesto que estos se apoyan en mayor medida en la teoría de programación lineal, mientras que las estrategias heurísticas tienden a acercarse más a estrategias algorítmicas clásicas. Para una descripción más detallada y profunda del tema se recomienda consultar [Cor+09].

Una vez aclarada la peculiaridad de la estrategia de *fuerza bruta*, procedemos a describir en qué se basa: El modo de funcionamiento consiste en la enumeración de todas las posibles configuraciones, para después proceder a filtrar todas aquellas que sean factibles, y de entre estas, escoger aquella que maximize (o minimize) la función objetivo. La definición de esta estrategia en pseudo código se incluye en el algoritmo 1 (donde U representa el universo de posibles soluciones, *feasible* se trata de una función que recibe una solución y se encarga de evaluar si es o no factible, F se compone de todas las soluciones factibles y por último, s se corresponde con la solución óptima, que en este caso se trata de un problema de minimización).

Result: s

```
1  $F \leftarrow \text{filter}(\text{feasible}, U);$   
2  $s \leftarrow \min\{F\};$ 
```

Algorithm 1: Estrategia de resolución basada en *fuerza bruta*.

Sin embargo, tal y como se puede intuir, esta estrategia de resolución requiere de una cantidad de recursos que crece de manera exponencial respecto del tamaño de entrada. Por lo tanto, no es aplicable en la práctica sobre problemas de gran tamaño. Aún así, existen ciertos problemas para los cuales el grado de complejidad es relativamente bajo o el tamaño de las instancias a resolver es pequeño, por lo que esta estrategia es aplicable. En dichos casos, esta es una alternativa interesante ya que, como se ha comentado anteriormente, los algoritmos basados en *fuerza bruta* proporcionan garantías de optimalidad.

Una de las ventajas de los métodos de resolución basado en fuerza bruta es la capacidad de poder aprovecharse de la estructura del problema concreto que se pretende resolver para así poder reducir el espacio de búsqueda de soluciones en gran medida. Esto se debe a que el espacio de posibles soluciones factibles es un subconjunto del espacio de posibles soluciones. Por tanto, si se consigue encontrar alguna estrategia que permita enfocarse únicamente en las soluciones factibles, los requerimientos computacionales serán mucho menores. Un ejemplo de ello podría aplicarse a una estrategia de generación de soluciones en problemas de rutas a partir de una estrategia de inserción al final de la ruta. Entonces, es fácil darse cuenta de que si durante el proceso de inserción de tareas en la ruta, se llega a un estado de infactibilidad, continuar el proceso de insertar nuevas tareas al final de la ruta, continuará generando soluciones infactibles. Por tanto, en este caso es posible reducir el espacio de búsqueda sin perder las garantías de optimalidad.

3.3.2. Selección Aleatoria

La siguiente estrategia heurística a comentar se presenta en muchos sentidos como una versión opuesta a la de *fuerza bruta*. Esta es la heurística basada en *selección aleatoria*, cuya principal característica es el apoyo en una componente de estocasticidad que pretende, a largo plazo, alcanzar resultados razonables utilizando para ello un reducido coste computacional. Entonces, tal y como se puede intuir esta estrategia es opuesta a la de *fuerza bruta* en dos sentidos: (1) la búsqueda en el espacio de soluciones no es exhaustiva, por lo que no se prueban todos los casos posibles (lo cual elimina las garantías de optimalidad) y (2) el coste computacional que conlleva generar soluciones factibles es mucho menor puesto que la decisión sobre qué camino explorar se delega en selección aleatoria.

Tal y como se puede apreciar en el algoritmo 2, la estrategia se basa en generar soluciones aleatorias del universo de posibles soluciones para después evaluar la factibilidad de la escogida en

cada iteración y almacenarla en caso de que esta maximice (o minimice) el valor de la función objetivo. Esto se repite mientras se cumpla una determinada condición, la cual puede basarse en un determinado número de iteraciones, una cantidad de tiempo o estrategias más elaboradas como un determinado ratio de mejora respecto entre la última mejor solución obtenida y la anterior, etc.

<p>Result: s</p> <pre> 1 $s \leftarrow \emptyset$; 2 while <i>condition</i> do 3 $c \leftarrow \text{random}(U)$; 4 if $\neg \text{feasible}(c)$ then 5 continue; 6 end 7 $s \leftarrow \min(s, c)$; 8 end </pre>
--

Algorithm 2: Estrategia de resolución basada en *selección aleatoria*.

Tal y como se puede intuir, por si sola esta heurística no es todo lo eficiente como se espera para ser aplicable en la práctica puesto que, a pesar de la simplicidad que proporciona durante el proceso de generación y evaluación de soluciones, resulta muy improbable que llegue a alcanzar soluciones razonables sin que la estrategia se apoye en cierto conocimiento del problema. Sin embargo, tal y como se comentará más adelante, esta estrategia sirve como base (o componente) de otras mucho más elaboradas que consiguen alcanzar resultados mucho mejores en el largo plazo mediante la utilización de cierta componente aleatoria en alguna parte del proceso de optimización. Desde el punto de vista conceptual, hacer esto consigue añadir cierto ruido en el camino de búsqueda hacia la solución óptima que puede ser positivo dado que este reduce la complicación de “caer” en máximos (o mínimos) locales que en ciertas estrategias heurísticas complican en gran medida el proceso de optimización.

3.3.3. Greedy

Una de las estrategias más conocidas (por su buena relación entre sencillez y calidad de las soluciones) son las basadas en *algorithms greedy*. Estos se corresponden con una de las principales categorías en las cuales se suelen dividir los algoritmos en la mayoría de libros de relacionados. La principal característica de este tipo de algoritmos es que cada una de las decisiones que toman se convierten en invariantes para los próximos pasos que lleve a cabo la estrategia en la búsqueda de la solución final. De ahí el nombre *greedy* (o voraz en español), que trata de remarcar la idea de que cada una de las decisiones que tomadas terminarán formando parte de la solución final. Nótese que esta condición conlleva dos consecuencias principales (1) la simplicidad tanto a nivel conceptual como conceptual al requerir únicamente del estado actual (obtenido a partir

de las decisiones tomadas anteriormente y las alternativas actualmente posibles), y (2) la restricción de reevaluar decisiones anteriores que permitirían llegar a una solución más acercada con el conocimiento obtenido en el momento actual.

Sin embargo, existen ciertos problemas que poseen ciertas características que permiten que los algoritmos de naturaleza voraz sean capaces de llegar al resultado correcto (u óptimo dependiendo de la naturaleza del problema) de manera eficiente y a la vez sencilla. El ejemplo más notable es el de el *problema del cajero*. Dicho problema consiste en la elección sobre qué monedas utilizar de entre un conjunto dado de posibles valores (los cuales no tienen restricciones de cardinalidad) de tal manera que el número de monedas que sea mínimo. A modo de ejemplo, suponiendo el conjunto de valores posibles $\{1, 5, 10, 25, 100\}$, la asignación que minimiza el número de monedas para devolver el valor 34 es de $[25, 5, 1, 1, 1, 1]$. La estrategia para llegar a esta solución consiste en seleccionar siempre la moneda de mayor valor que no supere el resto necesario para llegar al valor deseado. En el algoritmo 3 se muestra el pseudo código de la estrategia, donde S representa la solución, v el valor deseado, C el conjunto de posibles monedas y c una variable temporal. Tal y como se puede apreciar, siguiendo esta estrategia se cumplen las condiciones requeridas por los *algoritmos de naturaleza voraz* ya que únicamente se tiene en cuenta el estado actual para tomar la siguiente decisión y no se modifican las decisiones tomadas anteriormente. Para más información acerca de las características de esta categoría algorítmica así como ejemplos más interesantes se recomienda consultar [Cor+09].

Result: S

```

1  $S \leftarrow [];$ 
2 while  $v \neq 0$  do
3    $c \leftarrow \min_{v-c>0}\{C\};$ 
4    $v \leftarrow v - c;$ 
5    $S \leftarrow S + [c];$ 
6 end
```

Algorithm 3: Estrategia de resolución del *problema del cajero* basada en *algoritmo greedy*.

Desde el punto de vista de los problemas de optimización combinatoria, los *algoritmos de naturaleza voraz* han sido de gran utilidad para muchas situaciones. Sin embargo, tal y como se puede intuir, estos requieren que el problema sobre el que se van a aplicar cumpla la propiedad de ser concavos (una mejora sobre una solución arbitraria en la dirección de la solución óptima garantiza un menor coste en la función objetivo) para que estos puedan alcanzar la solución óptima. Por contra, esto ocurre en muy pocas ocasiones en la mayoría de problemas de optimización combinatoria. Si bien es cierto, hay casos para los cuales esta condición si se cumple (como el problema del cajero), pero en muchos otros (entre los que se encuentran los *problemas de rutas de vehículos*) esta condición no se satisface.

A pesar de ello, tal y como se ha indicado anteriormente, estos algoritmos son capaces de proporcionar resultados relativamente razonables utilizando una cantidad de recursos muy contenida por lo que cuando es tolerable cierto nivel de error respecto de la solución óptima, como es el caso cuando se recurre a estrategias basadas en heurísticas, estos algoritmos proporcionan resultados muy interesantes. Como es natural, el modo en que estos se aplican es muy dependiente de la naturaleza del problema en cuestión. Sin embargo, en el ámbito de los *problemas de optimización combinatoria* es común que la estructura de los mismos tenga una cierta forma: **Por lo general, los algoritmos greedy se construyen como un bucle condicional donde se controlan cuestiones relacionadas con el alcanzamiento de una solución final, ya sea debido a la falta de recursos u otros motivos. La acción que se repite en el bucle típicamente es la selección de la mejor alternativa de entre un conjunto de posibles opciones que modifican tanto el valor de la función de coste como el estado actual de una determinada manera.**

Entonces, un factor decisivo a la hora de aplicar *algoritmos de naturaleza voraz* se basa en elegir de manera adecuada el criterio de selección que se utilizará para elegir la siguiente alternativa que aplicar tras cada iteración del bucle condicional. En el siguiente apartado se describen algunos de los más utilizados en los problemas de optimización combinatoria.

3.3.3.1. Criterios de Selección

[TODO]

3.3.3.2. Randomized Greedy

A pesar de que la elección de un criterio de selección adecuado para el problema que se pretende resolver, existen ocasiones en que estos no son lo suficientemente “buenos” como se podría esperar. Por la propia naturaleza de los algoritmos de naturaleza voraz, en muchas ocasiones no es suficiente el contexto actual para escoger la decisión más acertada a largo plazo.

Para tratar de suavizar el peso de dicha decisión uno de los enfoques posibles es el de añadir un cierto grado de aleatoriedad en el momento de elegir la mejor opción a realizar. Es decir, en lugar de seguir una estrategia que escoja siempre la opción que maximice (o minimice) el valor del criterio de selección, se escoge aleatoriamente la siguiente opción de entre las mejores posibles. Como se ha comentado, la idea subyacente tras esta estrategia es que se asume que el criterio de selección escogido, a pesar de ser óptimo en el momento de su utilización (resultado parcial), probablemente no lo sea respecto del resultado completo. Entonces, la componente aleatoria trata de reducir dicho efecto generado por las “malas decisiones” tomadas.

Generalmente, para llevar a cabo la selección aleatoria se mantiene una lista ordenada con las k mejores opciones posibles (de acuerdo con el criterio de selección) para después utilizar

algun generador aleatorio de valores de una distribución aleatoria discreta que finalmente eligen la alternativa que finalmente utilizar. Estas distribuciones pueden ser de distinta naturaleza, aunque las más comunes son la la distribución uniforme discreta o alguna distribución armónica decreciente respecto del orden obtenido por el criterio de selección.

Como es esperable, estas estrategias tienden a mejorar los resultados obtenidos por los métodos greedy simples en promedio, pero obtienen peores resultado en el peor caso. Por lo tanto, su uso está muy extendido en metaheurísticas basadas en la generación de muchas soluciones. En concreto, este método se utiliza por las metaheurísticas *GRASP* tal y como se comentará en el apartado 3.4.1.

3.3.4. Clarke and Wright

Una de las estrategias heurísticas más conocidas para los problemas de optimización de rutas (y que es también aplicable al problema *Dial-a-Ride*) es la de ahorros (*Savings*) definida por Clarke y Wright en [CW64]. La estrategia esta definida asumiendo que no existen limitaciones en cuanto al número de vehículos disponibles para resolver el problema, por lo que esto es algo a tener en cuenta para su utilización. Sin embargo, es posible realizar pequeñas modificaciones que permitan su utilización con restricciones del número de vehículos. Una manera trivial de hacer esto es la de eliminar tantas rutas como sea necesario hasta que se cumplan las restricciones del número de vehículos sobre la solución generada por la heurísticas, aunque también se pueden aplicar otros enfoques como una penalización de costes elevada que limite el uso de más vehículos.

En cuanto al modo de funcionamiento de esta heurística, en un paso inicial genera una ruta desde almacen a cada tarea (en el caso del problema *Dial-a-Ride* esto consistirá en las dos paradas que requiere el viaje, la de recogida y la de entrega). Esto hace que sean en una situación inicial sean necesarios tantos vehículos como tareas a satisfacer. Posteriormente, se lleva a cabo un proceso iterativo de fusión de rutas mientras los costes de estas sean reducidos. Es decir, se calculan el coste total que genera cada par de rutas de manera independiente para posteriormente compararlo con el coste obtenido de combinar ambas rutas. En cada ciclo del algoritmo se escoge la opción que minimiza los costes y esta exploración se repite hasta que no haya opciones de mejora posibles. Es interesante remarcar que las estrategias de combinación de rutas no están limitados a la concatenación, sino que la estrategia se basa en la valoración de todos los posibles puntos en que una ruta se puede incluir en la otra. Esto hace que sea una estrategia de muy exhaustiva en el sentido de ser capaz de probar muchas situaciones.

Sin embargo, el motivo de la popularidad que ha recibido esta heurística durante tanto tiempo no solo se debe al método en si, sino a que además en [CW64] se describe la manera en que poder calcular el nivel de ahorro (de ahí el nombre de la estrategia) de la combinación de dos rutas sin tener que calcularlo los costes completos de ambas rutas y después hacer la diferencia respecto de

la versión combinada. La estrategia en que se apoyan para llevar esto a cabo es la diferencia entre los costes de los tramos de inicio y fin de la ruta que se va a insertar frente a los costes de los tramos de unión entre ambas rutas. Nótese que esta estrategia es altamente eficiente cuando las tareas tienen restricciones muy flexibles tanto desde el punto de vista de las ventanas temporales como de la capacidad del vehículo. En problemas más complejos como el *Dial-a-Ride* estas mejoras han de ser tomadas con prudencia ya que es necesario verificar la factibilidad de la ruta resultante.

3.3.5. Redes Neuronales

En los últimos años, el aumento de la capacidad computacional (sobre todo aplicable a operaciones de naturaleza matricial apoyado en gran parte por la evolución de las *unidades de computación gráficas (GPUs)*) a propiciado un incremento exponencial la cantidad de aplicaciones reales basadas en técnicas basadas en redes neuronales (cuyos requerimientos computacionales encajan perfectamente en dicho marco). Además de las aplicaciones prácticas, también ha surgido un gran incremento desde el punto de vista de la cantidad de artículos científicos relacionados con la materia. Entre otros temas, uno de los contextos en que se ha tratado de estudiar la factibilidad de este tipo de técnicas han sido los problemas de optimización combinatoria. Algunos de los más destacados son [Pot93; GO03; LJX04; MC09; Bel+16]. Tal y como se puede apreciar, estos trabajos tienen en común la utilización de enfoques basados en *redes neuronales recurrentes* así como *mapas auto-organizados*. A continuación se describen brevemente este tipo de arquitecturas:

1. **Redes Neuronales Recurrentes:** Consisten en una clase de redes neuronales con la característica de utilizar los pesos de ciertas neuronas de las capas internas como valores de entrada, creando una relación de recurrencia con la que se espera obtener la capacidad de tener un estado interno en la red. Es por ello que su utilización en el campo del análisis de lenguaje natural es muy interesante.
2. **Mapas Auto-Organizados:** Se trata de una arquitectura de red neuronal que se caracteriza por tener en cuenta la localización de las neuronas para el entrenamiento de la misma (que se organizan en forma de rejilla). Esto implica que los pesos de estas estén basados en la posición de las mismas. Es por ello que su utilización en procesos de visualización de datos utilizados en tareas de clustering es muy interesante.

Tras haber descrito brevemente estas dos arquitecturas de redes neuronales es fácil darse cuenta de las ideas en las que se apoyan los artículos anteriormente citados. En el caso de las técnicas basadas en redes neuronales recurrentes se espera “descubrir” un patrón de resolución común inherente a un problema de rutas concreto y no visible de manera natural, que las capacidades recurrentes de la red sean capaz de captar. En el caso de las técnicas basadas en mapas auto-organizados, la idea es que (por la estructura en forma de mapa de la red) se descubran patrones de cercanía entre las distintas localizaciones a visitar durante la ruta, con las que después se puedan reconstruir caminos de coste reducido.

Dichas estrategias presentan un gran potencial sobre aquellos problemas que presenten las características comentadas anteriormente. Sin embargo, hasta la fecha dichas ideas se encuentran todavía en fases muy tempranas de investigación por lo que no son aplicables en situaciones reales (en [Bel+16] fueron necesarias 7 horas de computación para resolver una instancia del problema del viajero de 50 localizaciones). Sin embargo, si que forman parte de un interesante nicho de investigación que podría ser muy prometedor en el futuro.

En este apartado se han descrito las principales estrategias heurísticas utilizadas en problemas de optimización de rutas, las cuales sirven de base para resolver las características concretas que se presentan en el problema *Dial-a-Ride*. Tal y como se comentó al principio del mismo, estas estrategias destacan por su relativa sencillez, tanto conceptual como computacional. Sin embargo, en muchas ocasiones no proporcionan resultados suficientemente buenos (desde el punto de vista de su distancia a la solución óptima) como se requiere. Para contrarestrar dicha situación surge una nueva clase de estrategias de optimización compuesta por otras estrategias más sencillas la cuál se discute en el próximo apartado.

3.4. Métodos de Resolución basados en Metaheurísticas

[TODO]

3.4.1. GRASP

[TODO]

3.4.2. Simulated Anneling

[TODO]

3.4.3. Tabu Search

[TODO]

3.4.4. Ant Colony

[TODO]

3.4.5. Variable Neighborhood Search

[TODO]

3.4.6. Large Neighborhood Search

[TODO]

[TODO]

3.5. Conclusiones

[TODO]

Capítulo 4

Implementación y Resultados

4.1. Introducción

[TODO]

4.2. Implementación

4.2.1. Objetivos

[TODO]

4.2.2. Visión a alto nivel

[TODO]

4.2.3. Componentes

[TODO]

4.3. Resultados

[TODO]

4.3.1. Definición del experimento

[TODO]

4.3.2. Tablas de resultados

[TODO]

4.3.3. Análisis de resultados

[TODO]

4.4. Conclusiones

[TODO]

Capítulo 5

Conclusiones Generales y Próximos pasos

5.1. Introducción

[TODO]

[TODO: Definir Secciones.]

5.2. Conclusiones

[TODO]

Bibliografía

- [Bel+16] Irwan Bello y col. «Neural combinatorial optimization with reinforcement learning». En: *arXiv preprint arXiv:1611.09940* (2016).
- [Bel54] Richard Bellman. *The theory of dynamic programming*. Inf. téc. Rand corp santa monica ca, 1954.
- [BKG19] Claudia Bongiovanni, Mor Kaspi y Nikolas Geroliminis. «The electric autonomous dial-a-ride problem». En: *Transportation Research Part B: Methodological* 122 (2019), págs. 436-456.
- [BPS03] J Christopher Beck, Patrick Prosser y Evgeny Selensky. «Vehicle routing and job shop scheduling: What's the difference?» En: *ICAPS*. 2003, págs. 267-276.
- [CL07] Jean-François Cordeau y Gilbert Laporte. «The dial-a-ride problem: models and algorithms». En: *Annals of operations research* 153.1 (2007), págs. 29-46.
- [Cor+09] Thomas H Cormen y col. *Introduction to algorithms*. MIT press, 2009.
- [Cor06] Jean-François Cordeau. «A branch-and-cut algorithm for the dial-a-ride problem». En: *Operations Research* 54.3 (2006), págs. 573-586.
- [CW64] Geoff Clarke y John W Wright. «Scheduling of vehicles from a central depot to a number of delivery points». En: *Operations research* 12.4 (1964), págs. 568-581.
- [Dia95] Robert B Dial. «Autonomous dial-a-ride transit introductory overview». En: *Transportation Research Part C: Emerging Technologies* 3.5 (1995), págs. 261-275.
- [DL91] Martin Desrochers y Gilbert Laporte. «Improvements and extensions to the Miller-Tucker-Zemlin subtour elimination constraints». En: *Operations Research Letters* 10.1 (1991), págs. 27-36.
- [EM12] Sevgi Erdoğan y Elise Miller-Hooks. «A green vehicle routing problem». En: *Transportation Research Part E: Logistics and Transportation Review* 48.1 (2012), págs. 100-114.
- [GO03] Hassan Ghaziri e Ibrahim H Osman. «A neural network algorithm for the traveling salesman problem with backhauls». En: *Computers & Industrial Engineering* 44.2 (2003), págs. 267-281.
- [KM70] Victor Klee y George J Minty. *How good is the simplex algorithm*. Inf. téc. WASHINGTON UNIV SEATTLE DEPT OF MATHEMATICS, 1970.

- [LJX04] Kwong-Sak Leung, Hui-Dong Jin y Zong-Ben Xu. «An expanding self-organizing neural network for the traveling salesman problem». En: *Neurocomputing* 62 (2004), págs. 267-292.
- [MC09] Thiago AS Masutti y Leandro N de Castro. «A self-organizing neural network using ideas from the immune system to solve the traveling salesman problem». En: *Information Sciences* 179.10 (2009), págs. 1454-1468.
- [MTZ60] Clair E Miller, Albert W Tucker y Richard A Zemlin. «Integer programming formulation of traveling salesman problems». En: *Journal of the ACM (JACM)* 7.4 (1960), págs. 326-329.
- [PDH08] Sophie N Parragh, Karl F Doerner y Richard F Hartl. «A survey on pickup and delivery problems». En: *Journal für Betriebswirtschaft* 58.1 (2008), págs. 21-51.
- [Pot93] Jean-Yves Potvin. «State-of-the-art survey—the traveling salesman problem: A neural network perspective». En: *ORSA Journal on Computing* 5.4 (1993), págs. 328-348.
- [RCL07] Stefan Ropke, Jean-François Cordeau y Gilbert Laporte. «Models and branch-and-cut algorithms for pickup and delivery problems with time windows». En: *Networks: An International Journal* 49.4 (2007), págs. 258-272.
- [TV02] Paolo Toth y Daniele Vigo. *The vehicle routing problem*. SIAM, 2002.
- [WN14] Laurence A Wolsey y George L Nemhauser. *Integer and combinatorial optimization*. John Wiley & Sons, 2014.

