

01184



UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO

FACULTAD DE INGENIERIA

UNA ESTRATEGIA PARA RESOLVER EL
PROBLEMA DE INVENTARIO - DISTRIBUCION

T E S I S

QUE PARA OBTENER EL GRADO DE:

DOCTORA EN INGENIERIA

(INVESTIGACION DE OPERACIONES)

P R E S E N T A :

MAYRA ELIZONDO CORTES

DIRECTOR DE TESIS: DR. RICARDO ACEVES GARCIA



2005

m345096



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

∞ Dedicado con todo mi cariño:

A mis papás Ronald y Elvira, la mejor parte de mi vida.

A mis hermanos Francisco y Nadia, parte de mí misma.

A Erubiel, mi mejor amigo y el hombre que amo.

*A toda mi familia,
quien me sigue dando,
muchos de los mejores momentos de mi vida.*

Agradecimientos:

Al Dr. Aceves por todo lo que compartimos durante estos años.

A la Dra. Idalia por su apoyo, cariño e invaluable amistad.

Al Dr. Luis Chías Becerril, al Dr. Manuel Ordorica Mellado, al Dr. Servio Tulio Guillén, al Dr. Juan Pablo Antún y al Dr. Sergio Fuentes Maya, por el apoyo y la disposición que me brindaron y por sus valiosos comentarios en la mejora del presente trabajo.

A la Dirección General de Estudios de Posgrado (DGEP) y al Consejo Nacional de Ciencia y Tecnología (CONACYT), por el apoyo otorgado para la realización de mis estudios de Doctorado.

A la Universidad Nacional Autónoma de México con el orgullo que siento de formar parte de sus filas,

POR MI RAZA HABLARÁ EL EPÍRITU.

Resumen

En el presente trabajo, se desarrolla una estrategia de solución eficiente para instancias no triviales del Problema de Inventario-Distribución o IRP (Inventory-Routing Problem) bajo dominio de decisión de tiempo. La estrategia consta de tres fases que minimizando los costos de inventario y transportación, y asegurando el abasto de producto, actúa como sigue: en la Fase I se define a qué clientes atender en el horizonte de planeación; la Fase II da como resultado la asignación de vehículos a clientes y la determinación de la cantidad de producto que se entregará a cada cliente cuando sea visitado; finalmente la Fase III indica en qué secuencia deben ser visitados. La parte esencial del trabajo se ubica en la segunda fase de la estrategia, ya que al utilizar la técnica de Descomposición Cruzada Separable, se logra solucionar al IRP que es esencialmente NP-Duro, a través de la iteración sucesiva entre problemas de transporte, obteniendo así, una estrategia de solución eficiente en cuanto a complejidad computacional. Al final se evalúa la eficiencia computacional de la estrategia.

Índice

Introducción.	3
Capítulo 1. El Problema de Inventario-Distribución (IRP).	7
1.1 Marco contextual.	7
1.2 Implicaciones de la solución del problema de Inventario-Distribución.	12
1.3 Formulación del modelo de Inventario-Distribución.	15
1.4 Revisión de literatura y clasificación del IRP.	18
Capítulo 2. Métodos y estrategias de solución del IRP.	25
2.1 Métodos bajo Dominio de Frecuencia.	26
2.1.1 Método de Agregación.	26
2.1.2 Métodos basados en Políticas de Partición Fija (FPP).	30
2.1.3 Métodos para Modelos de Frecuencia Discreta.	32
2.2 Métodos bajo Dominio de Tiempo.	33
2.2.1 Técnicas de Descomposición.	37
2.2.2 Métodos para Modelos de Flujo Dinámico con Demanda Desconocida.	40
2.3 Análisis de los métodos y estrategias de solución del IRP.	42
Capítulo 3. Estrategia propuesta.	47
3.1 Descripción del IRP tratado.	48
3.2 Descripción de la estrategia propuesta.	50
3.3 Problema de Asignación de Vehículos y Distribución (AVD).	52
3.4 Fase I: Aplicación de un esquema de abastecimiento.	55
3.4.1 Sistema de Inventarios de Periodo de Tiempo Fijo.	56
3.4.2 Sistema de Reabastecimiento Opcional (ORS).	58
3.5 Fase II: Asignación de vehículos a clientes y determinación del tamaño de las entregas.	60
3.5.1 Descomposición Cruzada y Descomposición Cruzada Separable.	62
3.5.2 Aplicación de la Descomposición Cruzada Separable al problema AVD.	69
3.5.3 Algoritmo de Descomposición Cruzada Separable para el problema AVD.	79
3.6 Fase III: Definición de la sucesión de visitas.	82
3.6.1 Método Dos Optimal.	83
3.6.2 Método Adaptación Prim.	86
3.6.3 Método Híbrido Dos Optimal-Prim.	87
3.7 Ejemplo de aplicación de la estrategia propuesta.	87

Capítulo 4. Eficiencia computacional.	89
4.1 Complejidad de la estrategia propuesta.	96
4.2 Experiencia computacional.	97
4.2.1 Metodología.	97
4.2.2 Resultados.	98
4.3 Comparación con otras estrategias.	100
Conclusiones y Extensiones.	105
Apéndice A. Complejidad Computacional.	i
Apéndice B. Principios de Descomposición de Benders y Descomposición Dual Lagrangeana.	xi
Apéndice C. Relajación Lagrangeana Separable.	xix
Apéndice D. Comparación de algoritmos para el Problema del Agente Viajero.	xxiii
Referencias.	

Introducción

i. Antecedentes.

La administración logística cambia de manera constante su forma de actuación. Un gran número de compañías y organizaciones del sector privado y de gobierno se están dando cuenta que el valor para un cliente puede, en gran parte, crearse a través de la administración logística. El concepto de valor dirigido al cliente, se puede visualizar a través de la disponibilidad del producto en lugar y tiempo, la oportunidad y consistencia de entregas, y la facilidad de colocación de órdenes, entre otros componentes de los servicios logísticos que forman parte de las cadenas de suministro. De tal forma, que los servicios logísticos se han reconocido actualmente como un elemento esencial de satisfacción del cliente, en un creciente número de mercados de productos y servicios.

La actividad de reabastecimiento de producto que maneja el proveedor, es un ejemplo de combinación de elementos que crean valor logístico. Es una tendencia que está surgiendo en la logística y se refiere a una situación en la cual un proveedor es quien maneja el abastecimiento de inventario de sus clientes. El reabastecimiento manejado por el proveedor crea valor para ambos, tanto para los proveedores como para los clientes. Es una situación de tipo ganar-ganar, debido a que los proveedores ahorran en costo de distribución, estando además en posibilidad de mejorar la coordinación de las entregas; por otro lado, los clientes no tienen que dedicar recursos a actividades de manejo de inventarios.

Una gran variedad de industrias está considerando implementar el reabastecimiento manejado por el proveedor. Tradicionalmente ha sido una situación muy deseada y finalmente implementada por los administradores logísticos, en la industria petroquímica y de gas industrial. Más recientemente, la industria automotriz (distribución de partes) y la industria refresquera (máquinas expendedoras) han entrado en esta dinámica.

Otra razón para que el reabastecimiento manejado por el proveedor sea recibido con gran interés, se debe a la rápida disminución de costos en la tecnología que permite monitorear a distancia los inventarios de los clientes. Además, el conocimiento preciso y oportuno del estado del inventario de los clientes es esencial para la implementación de cada estrategia.

ii. Problemática.

A pesar de que el reabastecimiento manejado por el proveedor es una situación ganar-ganar para proveedores y clientes, y de que una tecnología de monitoreo de inventarios está disponible; la razón para no aplicarla a gran escala es que, contar con una estrategia de distribución que minimice los costos de transporte e inventarios y al mismo tiempo evite el desabasto, una tarea en extremo complicada debido a las características del problema. La tarea de desarrollar tal estrategia de distribución se denomina Problema de Inventario-Distribución o IRP (Inventory-Routing Problem).

El modelo de inventario distribución fue propuesto por Federgruen y Zipkin en 1984 y numerosas modelaciones, así como estrategias de solución han sido propuestas desde entonces. Entre las estrategias de solución propuestas, se encuentran tanto combinaciones de técnicas analíticas con heurísticas como técnicas heurísticas únicamente. Sin embargo, el éxito de estas últimas es limitado, ya que dan una buena solución en un tiempo razonable, pero no garantizan la optimalidad.

En la aplicación práctica, el IRP como muchos otros problemas que surgen en el ámbito logístico, resulta ser un modelo de gran tamaño, asumiendo el concepto de gran tamaño, como modelos con un gran número de variables, así como un gran número de restricciones que pueden ser muy complejas; y también modelos para los que las capacidades de solución de los algoritmos y la velocidad de obtención de resultados, representan un grave problema. De hecho el IRP es un problema NP-duro, y cuando además es grande, la obtención óptima de la solución en tiempos razonables se hace prácticamente imposible.

Por otro lado, las micro, pequeñas y medianas empresas (MPME's) mexicanas, constituyen un subsector de nuestra economía, que juega un papel sumamente importante debido a la generación de un gran número de empleos, así como la distribución del ingreso. Además, tiene un alto potencial de desarrollo y posibilidades competitivas, debido a sus peculiares características de dinamismo y flexibilidad. A pesar de ello, las MPME's enfrentan una compleja problemática que involucra factores tales como: apertura comercial no planificada, falta de estímulos fiscales, escasa incorporación de la tecnología, sistemas de administración deficientes, etc., y el apoyo que el gobierno presta a este subsector es definitivamente insuficiente.

Uno de los obstáculos principales, que detienen el crecimiento de las MPME's, es la falta de recursos y cultura para inversión en proyectos para la optimización de sus actividades de planeación logística. Por lo cual se hace necesario, encontrar estrategias alternativas que logren que dichas empresas sean competitivas.

Por las razones mencionadas, es necesario el desarrollo de estrategias de solución de bajo costo de implementación, para la solución de problemas logísticos, tales como el IRP.

iii. Objetivo general.

Desarrollar una estrategia eficiente para la solución de instancias no triviales del Problema de Inventario-Distribución, basada en la Descomposición Cruzada Separable.

iv. Objetivos específicos.

- ↪ Elaborar el estado del arte del IRP.
- ↪ Conocer el marco teórico de los métodos de solución que existen para el IRP.
- ↪ Especificar el IRP tratado.
- ↪ Diseñar una estrategia de solución basada en la técnica de Descomposición Cruzada Separable y ejemplificar su uso.
- ↪ Mostrar la complejidad computacional de la estrategia.
- ↪ Construir una herramienta de cómputo que permita la aplicación de la estrategia.
- ↪ Evaluar el desempeño computacional de la estrategia mediante su aplicación a diferentes instancias del IRP tratado.

v. Contenido.

Para el cumplimiento de los objetivos mencionados, el documento está conformado de la siguiente manera:

- 📖 Capítulo 1: Se presenta el contexto logístico en el que surge el IRP, la importancia de mejorar estrategias de solución eficientes y poco costosas, la formulación del problema, y la revisión de literatura que concluye en la clasificación de este tipo de problemas.
- 📖 Capítulo 2: Se muestran, clasifican y analizan los métodos y estrategias de solución del IRP reportados en literatura. Este capítulo da una dirección de trabajo permitiendo proponer una estrategia mejorada.

- 📄 Capítulo 3: Se define el IRP tratado en el trabajo, acotándolo y clasificándolo de acuerdo con lo descrito en el Capítulo 1. Se describen detalladamente las fases de la estrategia propuesta y se ejemplifica su aplicación.
- 📄 Capítulo 4: Se define la complejidad computacional de la estrategia y se compara con estrategias reportadas en literatura. Se analiza su desempeño utilizando un programa de cómputo que aplica la estrategia a diferentes instancias del IRP tratado.
- 📄 Se presentan las conclusiones del trabajo de investigación y las extensiones del mismo.
- 📄 Finalmente se encuentran los apéndices necesarios para la comprensión del documento y por último las referencias consultadas.

Capítulo 1

El Problema de Inventario-Distribución (IRP).

El problema de Inventario-Distribución surge en un contexto logístico, y es un modelo que combina a las dos actividades más costosas de la cadena de suministro: el manejo de inventarios y la distribución física de productos. Desafortunadamente, es un problema complejo cuya solución implica una gran dificultad. Sin embargo, dada su importancia, requiere estrategias de solución que logren la optimización de las actividades logísticas que lo conforman.

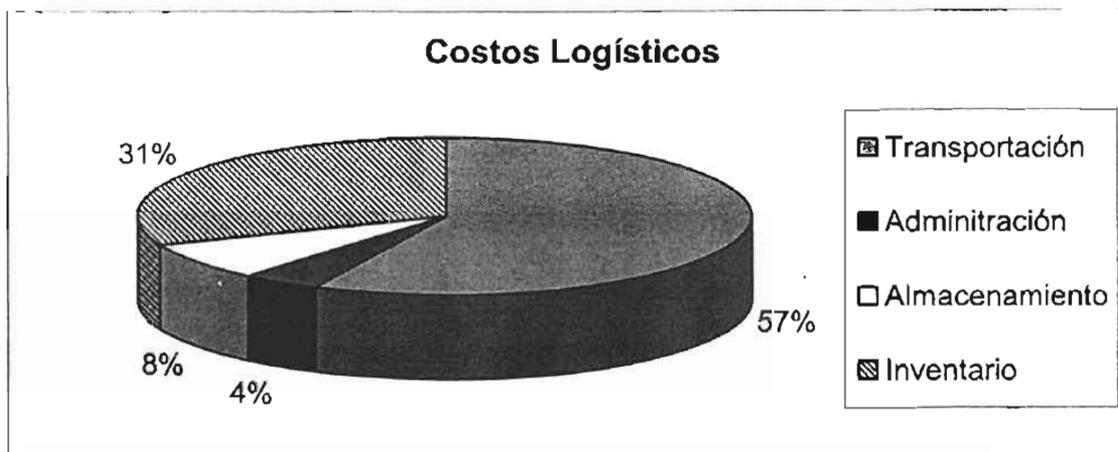
Este capítulo tiene como objetivo, presentar el marco general del problema de Inventario-Distribución por medio de cuatro secciones que muestran: el contexto del problema, su importancia y complejidad, su formulación, y finalmente, sus características de clasificación obtenidas mediante la revisión de literatura.

1.1 Marco contextual.

En la actualidad, los cambios que se llevan a cabo en el ámbito de la administración de las organizaciones, implican la necesidad de permanecer en un estado competitivo. Por tal motivo, es necesario que los sistemas logísticos sean continuamente y significativamente mejorados.

Los procesos de control de inventarios y distribución de productos, forman parte esencial y crítica de los sistemas de cadenas de suministro de las organizaciones dedicadas a la venta de productos y servicios.

Los costos de los factores logísticos de las organizaciones se presentan en las siguientes proporciones:



Fuente: CASS Information Systems

Fig. 1 Proporción de Costos Logísticos.

Es importante tomar en cuenta que hacer llegar los bienes a los consumidores o almacenes, resulta ser el paso más costoso del proceso de distribución. Algunos resultados reportados en literatura, muestran que los costos de la distribución física se estiman alrededor del 16% del valor de un producto (Dror y Trudeau, 1988). Un estudio desarrollado por Data Resources Inc. (Anily *et al.* 1990), especifica que, únicamente los costos de transporte en ciertos sectores de la economía, pueden equivaler a la quinta (madera y derivados) o hasta la cuarta parte (petróleo, rocas, arcilla y productos de vidrio) del valor promedio de las ventas. Consecuentemente, un pequeño porcentaje de ahorro en los gastos de transporte podría resultar en un ahorro global, sustancial.

El Problema de Inventario Distribución o IRP (Inventory-Routing Problem), se refiere a la distribución repetida de un único producto, desde uno o varios almacenes a múltiples consumidores previniendo su desabasto en un predeterminado horizonte de planeación, posiblemente infinito. Por ejemplo, podríamos considerar una compañía de distribución que tiene la tecnología para evaluar la tasa de consumo diario de sus productos. Esta compañía usa su propio sistema de toma de decisiones para todos sus problemas logísticos. De esta forma, los clientes podrían no necesitar hacer órdenes nunca más. La compañía de distribución podría tomar en cuenta el nivel de inventario para decidir cuándo visitarlos y cuánto abastecerles. La distancia entre el cliente y el almacén es un elemento clave. Si el objetivo es minimizar los costos de transporte, clientes lejanos podrían no ser visitados frecuentemente. Información precisa en todo momento de los niveles de inventario y la localización de los clientes, se requiere para tomar decisiones de costo efectivo para la compañía.

Hay muchos sistemas de distribución en los cuales la programación de vehículos y la sincronización y tamaño de las entregas a los clientes son (o deberían ser) simultáneamente determinados. Este es claramente el caso en los sistemas de

distribución interna, en los cuales el almacén y los consumidores representan estratos en la red de distribución de una compañía.

La necesidad de integrar el control de inventarios y decisiones de rutas de vehículos, surge en muchos procesos de distribución en los cuales las entregas deben ser hechas a consumidores externos.

Un ejemplo es la industria del gas, donde los productores instalan tanques en los lugares en que se encuentran sus consumidores y asumen la responsabilidad de mantener un adecuado nivel de inventario, determinando la frecuencia de abastecimiento y el tamaño de entregas.

Proveedores de supermercados y tiendas de departamentos, como otro ejemplo, frecuentemente adquieren espacios en anaqueles y tiene la responsabilidad de abastecer las existencias. De tal forma, ellos adoptan la completa administración del inventario de sus clientes. Se factura al cliente únicamente en el momento en que se realiza una venta al usuario final, los costos asociados con los inventarios del cliente los asume el proveedor. Al proveedor le está dada la responsabilidad de abastecer el inventario del cliente como lo decida, mientras garantice un nivel de abastecimiento, pudiendo considerar alguna pérdida de ventas o planeación de reservas.

A la política anterior se le denomina reabastecimiento de producto manejado por el proveedor (Vendor Managed Inventory, VMI o Vendor Manager Resupply, VMR). El IRP captura las características esenciales de la política de un VMI y las metodologías desarrolladas para este problema.

Esta política difiere del manejo de inventarios convencional de la siguiente forma: en el manejo de inventarios convencional, los clientes monitorean su propio nivel de inventarios, y cuando un cliente considera que es el momento de reordenar, un pedido de cierta cantidad de producto es solicitada al proveedor. El proveedor recibe esas órdenes de los clientes, prepara el producto para su entrega, y realiza dichas entregas utilizando una flotilla de vehículos.

El manejo de inventarios convencional tiene varias desventajas. Una de ellas es que las órdenes no tienen un patrón de ocurrencia uniforme en el tiempo y por lo tanto, tampoco los proveedores pueden utilizar sus recursos, tales como servicios de producción, almacenamiento y transporte, de manera uniforme en el tiempo. Otra desventaja es que algunos clientes ordenan pedidos aparentemente urgentes, cuando existen otros que son realmente urgentes. Como el proveedor no conoce los niveles de inventario de los clientes, la información necesaria para comparar la urgencia real de las diferentes órdenes no está disponible.

El VMI es un arreglo que ahorra al proveedor costosas inversiones en inventario y en las dificultades de su planeación; el proveedor tiene la ventaja de determinar cuándo y qué cantidad se debe distribuir a sus clientes. Más aún, cuando las demandas están sujetas a un considerable grado de incertidumbre, el sistema en

conjunto deriva beneficios adicionales de este arreglo, ya que el proveedor puede enfrentar un determinado nivel de servicio con un nivel de existencias de seguridad, significativamente menor que la suma de las existencias de seguridad requeridas por los vendedores individuales.

Una razón para que la política VMI haya sido recibida con gran interés es el bajo costo actual de disponibilidad de tecnología, que permite monitorear a distancia el inventario de los clientes. El Intercambio Electrónico de Datos o EDI (Electronic Data Interchange) es la comunicación de información estructurada de computadora a computadora y de aplicación a aplicación, sin intervención humana y sin papel, entre corporaciones, instituciones o individuos, utilizando un formato estándar para los documentos de negocios entre las empresas. Mediante su uso es posible obtener catálogos de precios y producto, pedidos, avisos anticipados de embarque, facturación electrónica, avisos de pago, reportes de ventas e inventarios, etc.

Previo al uso de tecnologías EDI, la mayor parte de los IRP consideraban a las demandas de los clientes como variables aleatorias que se convertían en determinísticas en el momento en que el vehículo que debía hacer la entrega, arribaba al lugar en que se localizaba el cliente. Esto implicó que métodos de pronóstico precisos debían ser utilizados para determinar el nivel de producto en el depósito de cada cliente para lograr definir eficientemente qué clientes debían ser visitados. También provocó que la cantidad de producto contenido en el vehículo podía no ser suficiente para completar la ruta programada.

Sin embargo, el uso extensivo de tecnologías EDI, elimina la necesidad de pronosticar la demanda de los clientes. La información de la cantidad de inventario de cada cliente estará disponible para el tomador de decisiones en el momento de planear las rutas de los vehículos, este hecho permitirá que la determinación de los clientes a los que se harán entregas en un día establecido y la cantidad de producto que les será entregada sea una tarea más sencilla de lo que ha sido en el pasado, simplificando en gran medida el IRP (Sarmiento y Nagi, 2003).

Por ejemplo, PRAXAIR (www.praxair.com) una gran compañía productora de gas industrial, usa el sistema Tracker™ como el de la Fig. 2, que es un avanzado microprocesador instalado en el tanque de abastecimiento del cliente, para acceder a lecturas del nivel de inventario en tiempo real. Para estar en posibilidad de practicar esta política, conviene que el proveedor tenga acceso a información precisa y oportuna sobre el estado del inventario de los clientes.



Fig. 2 Sistema Tracker™ utilizado por PRAXAIR.

El IRP está muy cercanamente relacionado con el Problema de Ruteo de Vehículos o VRP (Vehicle Routing Problem), Toth y Vigo (2002). El VRP es fácilmente establecido: dado un conjunto de nodos cada uno con demanda conocida, y un conjunto de vehículos cada uno con capacidad conocida, ¿cuáles son las rutas de distribución desde un almacén central, que minimizan el total de la distancia recorrida? Cuando hay solamente un vehículo de suficiente capacidad, el VRP es equivalente al Problema del Agente Viajero o TSP (Traveling Salesman Problem).

El VRP es un problema NP-duro (Garey y Johnson, 1979) y muchos desarrollos se han hecho para resolverlo, incluyendo algunos métodos exactos como bifurcación y acotación, bifurcación y corte y programación dinámica que, sin embargo, sólo manejan problemas pequeños y poco realistas; también se han desarrollado numerosas heurísticas y metaheurísticas.

De hecho, el IRP se interpreta como un enriquecimiento del VRP que incluye restricciones de inventario. Sin embargo, el IRP es muy diferente al VRP. Un VRP ocurre cuando los clientes hacen órdenes y el proveedor asigna algunas de ellas a las rutas de los vehículos para ser abastecidas en ese día (You call – we haul). En el IRP, el proveedor, no el cliente, decide cuánto hay que entregar a cada cliente cada día (You rely – we supply).

Por otro lado, una visión interesante y novedosa, en la optimización de las cadenas logísticas, es la posibilidad de la combinación de diferentes elementos de dichas cadenas. La creación de modelos que manejen de forma combinada diferentes elementos cuya optimización individual es de importancia vital, crea la posibilidad de una optimización global que podría tener la capacidad de mejorar los niveles de los indicadores de desempeño de las organizaciones y crear más situaciones de tipo ganar-ganar, para las diferentes entidades involucradas. Un enfoque integral logra la modelación de problemas que represente mejor a la realidad, que no sean necesariamente más complejos que sus partes componentes. Sin duda, el IRP es un ejemplo de lo anterior.

1.2 Implicaciones de la solución del problema de Inventario-Distribución.

Si bien el IRP fue primero introducido en la literatura a través de aplicaciones con productos líquidos y gas, muchas industrias diferentes están involucradas con sistemas de inventario manejados por el proveedor o están explorando su uso.

Ejemplos de estas industrias son las siguientes:

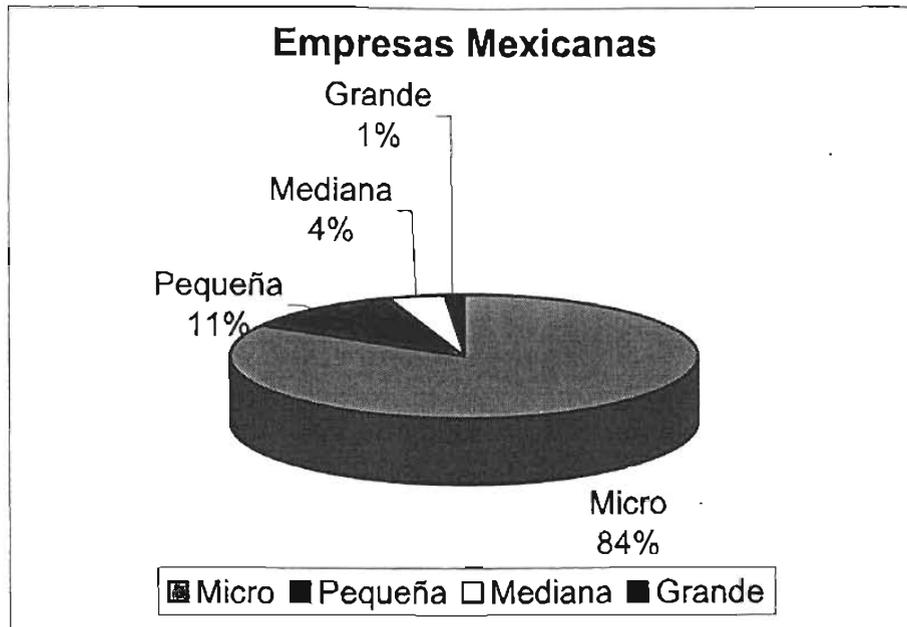
- ▲ Aeroespacial.
- ▲ Agua y desechos.
- ▲ Alimentos y bebidas.
- ▲ Automovilística.
- ▲ Aviación.
- ▲ Celdas de combustible.
- ▲ Cuidado del hogar.
- ▲ Cuidado de la salud/médica.
- ▲ Electrónica/semiconductores.
- ▲ Ensamble de electrónicos.
- ▲ Fabricación con metal.
- ▲ Impresión.
- ▲ Petrolera.
- ▲ Producción de metal.
- ▲ Pulpa y papel.
- ▲ Servicios.
- ▲ Venta a minoristas.
- ▲ Vestido.
- ▲ Vidrio.

En México las micro, pequeñas y medianas empresas (MPME's), constituyen un subsector de la economía que juega un papel sumamente importante, debido a la generación de un gran número de empleos, así como la distribución del ingreso. Además, tienen un alto potencial de desarrollo y posibilidades competitivas debido a sus peculiares características de dinamismo y flexibilidad.

Información generada por la Secretaría de Economía (SE) actualizada al año 2001, reporta que el número total de empresas en México es 774,904, que están distribuidas de la siguiente forma:

- 83.2% de microempresa
- 11.4% de pequeña empresa
- 4.0% de mediana empresa
- 1.4% de gran empresa

Como puede verse, en total el 98.86% de las empresas nacionales pertenece al subsector al que nos referimos. Esta situación puede observarse en la figura 3.



Fuente: Secretaría de Economía.

Fig. 3 Micro, pequeñas y medianas empresas del contexto nacional.

Para la definición y clasificación de este subsector, la SE utiliza a las variables: número de empleados y número de establecimientos. Además, divide en tres sectores: industria, comercio y servicios.

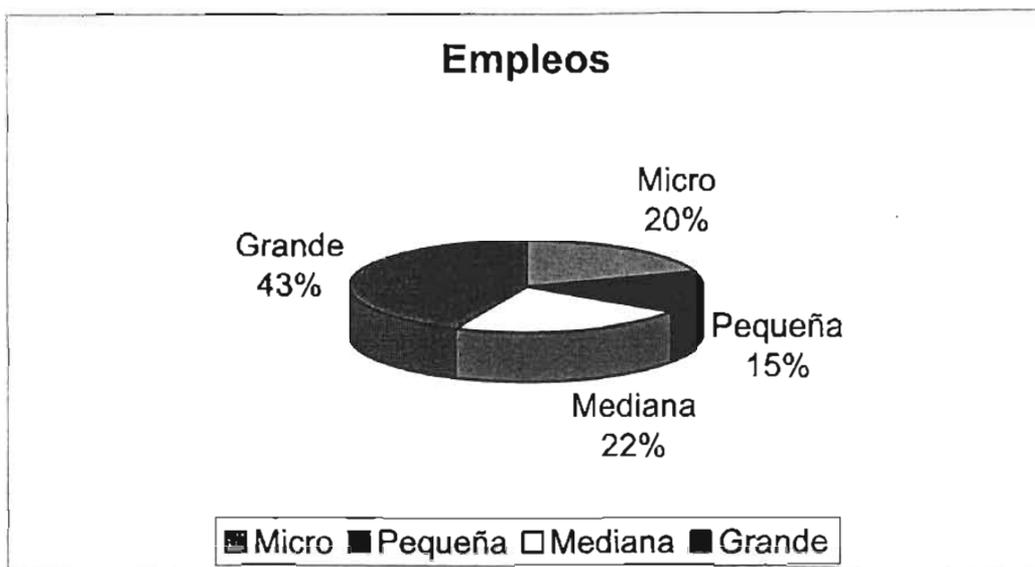
Según el Diario Oficial de la Federación, del 30 de marzo de 1999, la estratificación de empresas micro, pequeña y mediana es como sigue:

Tamaño	Sector Industria	Sector Comercio	Sector Servicios
MICRO	0-30	0-5	0-20
PEQUEÑA	31-100	6-20	21-50
MEDIANA	101-500	21-100	51-100
GRANDE	501 en adelante	101 en adelante	101 en adelante

Fuente: Secretaría de Economía.

Tabla 1 Clasificación por número de empleados.

De los 12,200,369 empleos generados por las empresas del país, las MPME's generan el 56% con la siguiente proporción:



Fuente: Secretaría de Economía.

Fig. 4 Empleos generados por las MPME's.

En países en vías de desarrollo como es el caso de México, este tipo de empresas son una forma de subsistencia, articulada con la economía popular, que excepcionalmente se vinculan al sector moderno como agente innovador. Sin embargo, las estadísticas muestran que existe una gran cantidad de micro, pequeñas y medianas empresas, que pueden llegar a formar parte de las grandes empresas y aún a competir con ellas. Según informaciones de la SE, las MPME's son uno de los sectores de crecimiento más dinámico en el país y pueden competir "exitosamente" en mercados cada vez más exigentes.

Por las razones anteriores es evidente la importancia que tienen las MPME's en la economía nacional. A pesar de ello, las MPME's, enfrentan una compleja problemática que involucra factores tales como: apertura comercial no planificada, devaluaciones sexenales, inflación, deficiente infraestructura regional, falta de estímulos fiscales, escasa incorporación de la tecnología, competencia muy desventajosa, sistemas de administración deficientes, etc, y el apoyo que el gobierno presta a este subsector, es definitivamente insuficiente.

Por otro lado es importante considerar que la competitividad resulta de esencial importancia en el contexto actual de globalización comercial, ya que una empresa, sea cual fuere su dimensión o actividad, que sea competitiva, sería capaz de adaptarse a los cambios, aprender de los obstáculos, aprovechar las oportunidades y reconocer las amenazas. A pesar de esto, uno de los obstáculos principales para la competitividad, que detienen el crecimiento de las MPME's, es la falta de recursos y cultura para invertir en proyectos de desarrollo. Por lo tanto el desarrollo de estrategias poco costosas e implementables que eleven los niveles de sus indicadores de desempeño, les proporcionarán habilidades de sobrevivencia en el inestable, sorpresivo y cambiante entorno económico

mexicano. De tal suerte, que al dirigir los esfuerzos necesarios al auge de las MPME's mexicanas, se contribuye al desarrollo de la economía del país.

El trabajo de investigación desarrollado pretende dar solución al IRP que se presenta en la planeación logística en la industria, el comercio y los servicios, mediante estrategias implementables a bajo costo.

El IRP, como muchos problemas que surgen en el ámbito logístico, es un modelo de gran tamaño. En este contexto, gran tamaño se refiere al gran número de clientes a los que se debe atender. Entendemos también el concepto de gran tamaño, como modelos con un gran número de variables, que igualmente presentan muchas restricciones; y aun muy complejas. Este mismo concepto incluye modelos para los que las capacidades de solución de los algoritmos y la velocidad de obtención de resultados, resultan un grave problema.

El IRP es, por lo tanto, un problema logístico grande y complejo que combina un elemento temporal (el momento en que el reabasto es realizado) con un elemento espacial (el ruteo de vehículos); de hecho, es un problema NP-duro (ver Apéndice A). De tal forma, la obtención óptima de la solución en tiempos razonables es prácticamente imposible.

Es debido a lo anterior y a la consideración de la importancia del problema en el ámbito logístico, que resulta importante y necesario el estudio y desarrollo de estrategias de solución que logren ser eficientes para problemas de gran tamaño, *i.e.*, de tamaño realista que ofrezcan resultados óptimos en los tiempos que se requieren. Numerosas organizaciones planean sus actividades al inicio de cada día y puede entonces necesitarse el tener resultados que permitan realizar rápidamente esta planeación.

Es importante mencionar que, dado el marco contextual de este problema, la implementación de la estrategia es un factor que debe siempre tenerse en cuenta. Así, es necesario no perder de vista que las estrategias logísticas deben ser altamente susceptibles de ser utilizadas, implementándolas como programas de computadora que aprovechen los avances de la investigación teórica, obteniendo en herramientas suficientemente comprensibles para quienes serán sus usuarios.

1.3 Formulación del modelo de Inventario-Distribución.

El modelo de inventario distribución fue propuesto por Federgruen y Zipkin en 1984. A partir de entonces se consideraron diversas variaciones del problema, así como varias estrategias de solución.

Ambos consideran el siguiente modelo de un solo período. En el comienzo de éste, el inventario inicial (que quizá permanece abastecido por un periodo previo) para cada ubicación, es reportado al almacén central. Esta información se utiliza para determinar el nivel de producto disponible. Al mismo tiempo, la asignación de

clientes a vehículos y sus rutas son determinadas. Después, se realizan las distribuciones, las demandas se satisfacen; y los costos de transporte a la localidad del inventario y de escasez de producto son cargados a cada ubicación proporcionalmente, al nivel de inventario, al final del periodo. Observe que en este modelo es posible no visitar alguna de las ubicaciones en el horizonte de tiempo especificado.

La notación que utilizaron es la siguiente:

K = número de vehículos

$F_i(\cdot)$ = función de distribución acumulativa de la demanda de un período en la localidad i

h_i^+ = costo de manejo de inventario por unidad en la localidad i

h_i^- = costo de déficit de inventario por unidad en la localidad i

β_i = inventario inicial en la localidad i

W = cantidad total disponible de producto en el almacén central

b = capacidad del vehículo

A diferencia de VRP, w_i , la cantidad entregada al cliente i , es ahora una variable. Sea $y_{ik} = 1$ si el punto de entrega i es asignado a la ruta k , y $y_{ik} = 0$ en otro caso. Sea $x_{ijk} = 1$ si el vehículo k viaja directamente de la ubicación i a la ubicación j , y $x_{ijk} = 0$ en otro caso.

La función de costo de inventario $s_i(\cdot)$ está dada por:

$$s_i(w_i) = \int_{\beta_i + w_i}^{\infty} h_i^-(u - \beta_i - w_i) dF_i(u) + \int_0^{\beta_i + w_i} h_i^+(\beta_i + w_i - u) dF_i(u)$$

para toda i , mientras su derivada es denotada por $s'_i(\cdot)$ y donde $s_i(\cdot)$ es estrictamente convexa y continuamente diferenciable.

El problema de inventario-distribución de único almacén y de único período, se puede formular de la siguiente manera:

$$\text{Minimizar} \quad \sum_{i,j,k} c_{ij} x_{ijk} + \sum_i s_i(w_i)$$

sujeto a

$$\sum_i w_i y_{ik} \leq b \quad \forall k = 0, 1, \dots, K \quad (1)$$

$$\sum_i w_i \leq W \quad (2)$$

$$w_i \geq 0 \quad (3)$$

$$\sum_{k=1}^K y_{0k} = K \quad (4)$$

$$\sum_{k=0}^K y_{ik} = 1 \quad i = 1, \dots, n \quad (5)$$

$$0 \leq y_{ik} \leq 1 \text{ y entera} \quad (6)$$

$$\sum_i x_{ijk} = y_{jk} \quad \forall j, k \quad (7)$$

$$\sum_j x_{ijk} = y_{ik} \quad \forall i, k \quad (8)$$

$$\sum_{ij \in S} x_{ijk} \leq |S| - 1 \quad \forall S \subseteq \{1, \dots, n\} \quad (9)$$

$$0 \leq x_{ijk} \leq 1 \text{ y entera} \quad (10)$$

La restricción (1) asegura que la carga total asignada a un vehículo no es mayor que su capacidad. La restricción (2) garantiza que la cantidad embarcada no es mayor que W , la cantidad disponible en el almacén, mientras que la restricción (3) garantiza que esta cantidad sea no negativa. Las restricciones restantes aseguran que sean construidas rutas factibles. Observemos que cuando las variables y_{ik} se fijan, el problema se descompone en $(K+1)$ subproblemas simples, uno de localización de inventarios y K problemas de agente viajero. Federgruen y Zipkin (1983) proporcionaron un método eficiente para la solución del subproblema de localización de inventarios.

La modelación anterior es una de las muchas desarrolladas que dependen, tanto de las situaciones especiales que se modelen, *i.e.*, agregando o modificando restricciones o la definición de la función objetivo, así como de los métodos de solución utilizados.

Como se mencionó anteriormente el IRP es NP-duro para un contexto de NP-completez (ver Apéndice A), debido a que el componente de ruteo de vehículos del problema, que ahora incluye restricciones de inventario, es NP-duro (Cousineau-Ouimet, 2002) y generaliza al TSP (Garey y Johnson, 1979). De hecho, todos los problemas no triviales de ruteo de vehículos son NP-duros (Bramel y Simchi-Levi, 1995). Por esa razón es improbable que un algoritmo de

tiempo polinomial sea desarrollado para su solución óptima. Consecuentemente, mucho esfuerzo se ha hecho para el desarrollo de algoritmos heurísticos y estrategias para su solución.

1.4 Revisión de literatura y clasificación del IRP.

El IRP tiene como característica la presencia simultánea de tres aspectos básicos:

Ruteo: entendido como el movimiento físico de bienes entre diferentes sitios geográficos, tales como depósitos, almacenes, puntos de producción, puntos de venta al menudeo, etc.

Inventario: considerando elementos, tales como cantidades y valores de los bienes que son movidos y que son importantes para la definición y evaluación de estrategias de organización y operación.

Dinamismo: los dos primeros aspectos están combinados en un ambiente dinámico de tal forma que, repetidas decisiones son tomadas en diferentes momentos del horizonte de tiempo, y decisiones pasadas influirán en decisiones futuras.

El marco general en el cual se consideran los aspectos mencionados, es la logística de distribución. Además de su prevalencia en la experiencia diaria, la relevancia práctica de la clase problemas que se consideran en esta área, está también probada por la gran cantidad de información reportada en literatura de cada uno de los tres aspectos anteriores, considerados de forma separada, o en pares. De tal forma que muchos modelos se ha propuesto para problemas ordinarios de inventarios y tamaños de lote, donde no se consideran las situaciones de ruteo, o costos de transporte los cuales se les tiene sólo como costos fijos (Chikán, 1990; Salomon, 1995; Hahn y Yano, 1995). Para problemas que involucran aspectos de ruteo únicamente, varios problemas tipo se han estudiado, de los cuales existen excelentes revisiones (por ejemplo Jünger *et al.*, 1995; Fisher, 1995; Assad y Golden, 1995).

Sorpresivamente, existen muy pocos desarrollos que tratan con los tres aspectos en forma integrada: unas pocas docenas contra los miles que tratan de los problemas tipo, *i.e.*, inventarios y ruteo por separado. Más aún, las dos revisiones disponibles en esta área (Daganzo, 1996; Federgruen y Simchi-levi, 1995) están más bien enfocados a los supuestos de específicos métodos desarrollados. Los artículos considerados en esas dos revisiones, ni siquiera citan artículos del otro grupo, a pesar de que los problemas considerados son totalmente similares, aunque son abordados de manera diferente.

Una primera clasificación del IRP determina dos variantes importantes: problemas de tipo determinístico, en el cual se asume que conocemos el promedio de consumo de los clientes; y problemas de tipo estocástico, en los cuales se asume

que conocemos la distribución de probabilidad de consumo de los clientes (Sarmiento y Nagi, 2003). Sin embargo, una clasificación más detallada y completa se desarrolla a continuación.

El IRP debe decidir la mejor planeación de actividades de abastecimiento de (uno o varios bienes) desde (uno o varios) orígenes a (uno o varios) destinos durante algún horizonte de tiempo (finito o infinito), considerando problemas de ruteo y de inventario. Para efectos de simplicidad, nos referiremos a envíos utilizando vehículos, para los cuales el tiempo de partida, sus cargas y rutas, pueden ser elegidas libremente, a menos que se especifique lo contrario. Además, actividades de distribución serán consideradas en lugar de actividades de recolección, aunque es necesario aclarar que, la diferencia es irrelevante dados los métodos de solución revisados.

Las decisiones que deben tomar los encargados de estas actividades son:

- **cuándo** deben ser hechos los envíos; *i.e.*, cuándo deben ser cargados los vehículos y cuándo deben ser visitados los clientes;
- **cuánto** se debe cargar en cada vehículo, en términos de la cantidad de cada producto bajo consideración, y cuánto de tal carga debe distribuirse entre los clientes que la requieren;
- **qué ruta** debe seguir cada vehículo para visitar a sus clientes.

A través del tiempo se han desarrollado modelos con diferentes características para hacer frente a estas decisiones, las cuales están siempre presentes, aunque en diferentes escenarios o formulaciones y el objetivo de su revisión es proporcionar un marco general de trabajo para la presente investigación, y contar con una herramienta útil para conocer las direcciones tomadas por las metodologías, en las diversas situaciones específicas del IRP, que permita dejar clara la dirección de la estrategia propuesta.

Una clasificación desarrollada por Baita, Walter, Pesenti y Favareto (1998), distingue siete grupos principales de elementos mediante los cuales se puede clasificar al IRP:

1. Topología.

La relación origen-destino puede ser uno a uno, uno a mucho, muchos a uno o muchos a muchos.

2. Número de productos.

Uno o muchos productos.

3. Tipo de demanda.

Puede ser totalmente conocida o incierta y entonces será descrita por una variable aleatoria, de hecho, puede ser totalmente desconocida, si no hay una

representación estocástica asociada a ella. Cuando se consideran varios clientes o productos, la demanda puede estar uniformemente distribuida o no.

4. Dominio de decisión.

Se consideran dos casos bajo los cuales se toman las decisiones: frecuencia o tiempo. En el dominio de frecuencia las variables de decisión son las frecuencias de reabastecimiento o tiempo entre envíos. En el dominio de tiempo la programación de envíos está decidida o, con modelos de tiempo discreto, las cantidades y rutas son decididas para intervalos fijos de tiempo.

5. Restricciones/objetivo.

Las relaciones entre variables de decisión se pueden expresar como restricciones o como objetivos. Las capacidades de los vehículos, de los inventarios y de los puntos de abastecimiento pueden estar presentes, tanto en las restricciones como en los objetivos o pueden no estarlo. De igual forma el número de vehículos podría considerarse como una variable, un valor predeterminado, o que el número de vehículos no esté restringido. Es posible también, que la distribución sea realizada por instancias externas.

6.- Costos.

El objetivo es siempre la minimización de costos relacionados con el transporte, que incluyen costos fijos por viajes y costos variables relacionados con la distancia o el número de clientes visitados; y costos relacionados con el inventario, que involucran el almacenamiento de bienes, costos por desabasto y costos fijos por ordenación de pedidos.

7.- Métodos de solución propuestos.

Entre los métodos propuestos, se clasifica la descomposición del horizonte de tiempo o descomposición del tipo "agrupar primero, rutear después" o viceversa; y para determinar cómo agrupar a los clientes en dichas rutas se toma en cuenta su posición geográfica, su frecuencia de visitas estimadas, o el tiempo estimado de partida. También puede identificarse el uso del cálculo del tamaño económico de lote o EOQ (por sus siglas en inglés), para obtener frecuencias de entregas, así como cantidades de entrega. De manera más general, los métodos de solución son aproximados o exactos. Además, es útil distinguir si los modelos de programación matemática utilizados son lineales, enteros o no lineales.

La Tabla 2 muestra los elementos de clasificación, atributos y alternativas para el IRP desarrollada por Baita *et al.* (1998). A partir de ella es posible clasificar a los artículos que tratan este problema, para que en lugar de analizar a cada uno por separado, sea posible un análisis comparativo entre diferentes métodos que han sido utilizados para modelos similares o iguales métodos para diferentes problemas. El nivel de clasificación principal se hace al considerar el dominio de decisión. La Tabla 3 muestra la clasificación realizada a partir de los artículos revisados con base en Baita *et al.* (1998).

Elementos de clasificación	Atributo	Alternativas		
<i>Topología</i>	Puntos finales	Uno a uno	Uno a muchos	Muchos a muchos
<i>Productos</i>	Número	Uno	Muchos	
<i>Demanda</i>	Conocimiento Comportamiento en el tiempo Distribución	Conocida Constante Uniforme	Incierta (v.a.) Variable No uniforme	
<i>Tipo de decisión</i>	Dominio de decisión	Tiempo	Frecuencia	
<i>Restricciones/Objetivo</i>	Capacidad del vehículo Capacidad de inventario Capacidad de abastecimiento Número de vehículos	Igual Presente Presente Dado	Diferente No presente No presente Var. de dec.	No restringido
<i>Costos</i>	Inventario Transportación	Almacenamiento Fijo	Desabasto Relacionado con la distancia	Ordenación Relacionado con el número de paradas
<i>Método de solución</i>	Descomposición Agrupación Modelo/algorithmo Modelos de MP ¹	Tiempo Tiempo Exacto LP ²	Agrupar-Rutear Frecuencia Aproximado IP ³ o MIP ⁴	Distancia NLP ⁵

Fuente: Baita *et al.* (1998).

- 1 Programación Matemática
- 2 Programación Lineal
- 3 Programación Entera
- 4 Programación Entera Mixta
- 5 Programación No Lineal

Tabla 2 Elementos de clasificación del IRP.

Autor	Año	Estructura de nodos terminales	Prod.	Demanda				Capacidades				Costos de transportación			Método de solución			Otros modelos
				Conocida	Tiempo	Dist. uniforme	Dominio de decisión	Vehículo	Invent.	Sum.	No. de vehículos	Fijo	Dist.	Parad.	Costos de inventario	Descomposición	Agrupar	
Federgruen y Zipkin	1984	Uno a muchos	Uno	Incierta	Constante	Diferente	Tiempo	Diferente	SI	SI	Dado	No	SI	No	Inventario del vendedor + desabasto	Agrupar I - Rutear II	Mínimo costo	MP
Golden, Assad y Dahl	1984	Uno a muchos	Uno	Conocida	Constante	Diferente	Tiempo	Igual	SI	No	Var. de dec.	SI	SI	No	SI	Rutear I - Agrupar II	Razón de urgencia y tiempo	Heurístico
Blumenfeld et al.	1985	Muchos a muchos	Muchos	Conocida	Constante	Igual	Frecuencia	Igual	No	No	SI	SI	SI	SI	No	No	EOQ	-
Burns et al.	1985	Uno a muchos	Uno	Incierta	Constante	Diferente	Frecuencia	Igual	No	No	SI No restringido	SI	SI	SI	Agrupar I - Rutear II	Distancia	EOQ	-
Daganzo	1985	Muchos a uno	Uno	Conocida	Constante	Igual	Frecuencia	Igual	No	No	Var. de dec.	No	SI	No	SI	Agrupar I - Rutear II	Capacidad de carga	EOQ
Daganzo y Newell	1985	Uno a uno	Uno	Conocida	Variable	Igual	Frecuencia	Igual	No	No	Var. de dec.	No	SI	SI	SI	Agrupar I - Rutear II	Distancia	EQD
Hall	1985	Muchos a uno	Uno	Conocida	Constante	Igual	Frecuencia	Igual	No	No	No restringido	SI	SI	SI	Agrupar I - Rutear II	Distancia	EOQ	-
Daganzo y Newell	1986	Uno a muchos	Uno	Conocida	Constante	Diferente	Frecuencia	Diferente	No	No	No restringido	No	SI	No	SI	Agrupar I - Rutear II	Distancia	EOQ
Dror y Levy	1986	Uno a muchos	Uno	Conocida	Variable	Diferente	Tiempo	Igual	No	No	Dado	No	SI	No	SI	Tiempo	Tiempo	heurístico
Federgruen, Prastacos y Zipkin	1986	Uno a muchos	Dos	Incierta	Constante	Diferente	Tiempo	Diferente	SI	SI	Dado No restringido	No	SI	No	Inventario del vendedor + desabasto	Agrupar I - Rutear II	Mínimo costo	PM y heurístico
Blumenfeld et al.	1987	Muchos a muchos	Muchos	Conocida	Constante	Diferente	Frecuencia	Igual	No	No	No restringido	No	SI	No	SI	Agrupar I - Rutear II	Distancia	EOQ
Daganzo	1987	Muchos a muchos	Uno	Conocida	Constante	Igual	Frecuencia	Igual	No	No	No restringido	No	SI	No	SI + desabasto	No	No	EOQ
Dror y Ball	1987	Uno a muchos	Uno	Incierta	Constante	Diferente	Tiempo	Igual	SI	No	Dado	No	SI	No	SI	Agrupar I - Rutear II	Tiempo	MIP
Miller	1987	Uno a uno	Muchos	Conocida	Variable	Diferente	Tiempo	Diferente	SI	No	Dado No restringido	Por viaje	SI	SI	SI	No	No	MIP
Daganzo	1988	Uno a uno	Muchos	Conocida	Constante	Diferente	Frecuencia	Igual	No	SI	No restringido	No	SI	No	SI	Agrupar I - Rutear II	Distancia	EOQ
Benjamin	1989	Muchos a muchos	Uno	Conocida	Constante	Diferente	Frecuencia	No	No	SI	No restringido	No	SI	No	Orden/dep. del vendedor	No	No	EOQ
Chen, Balakrishnan y Wong	1989	Uno a muchos	Uno	Conocida	Constante	Diferente	Tiempo	Diferente	No	SI	Dado	Por viaje	SI	No	SI + desabasto	No	No	Heurístico y MIP
Anily y Federgruen	1990	Uno a uno	Uno	Conocida	Constante	Diferente	Frecuencia	Diferente	SI	No	No restringido	Por viaje	SI	No	SI	Agrupar I - Rutear II	Distancia	Heurístico asintóticamente óptimo
Gallego y Simchi-Levi	1990	Uno a uno	Uno	Conocida	Constante	Diferente	Frecuencia	Igual	No	No	No restringido	Por viaje	SI	SI	Dependiente del vendedor	Ruta directa	No	Colas
Kincewicz	1990	Muchos a muchos	Muchos	Conocida	Constante	Igual	Frecuencia	No	No	No	No restringido	SI	No	No	SI	No	No	EOQ
Anily y Federgruen	1991	Uno a muchos	Uno	Conocida	Constante	Diferente	Frecuencia	Igual	No	No	Dado	Por viaje	SI	No	SI	Agrupar I - Rutear II	Distancia	Heurístico asintóticamente óptimo

Fuente: Elaboración propia con base en Batta et al. (1998).

Tabla 3 Clasificación de artículos relativos al IRP.

Autor	Año	Estructura de nodos terminales	Prod.	Demanda				Capacidades				Costos de transportación			Método de solución			
				Conocida	Tiempo	Dist. uniforme	Dominio de decisión	Vehículo	Invent.	Sum.	No. de vehículos	Fijo	Dist.	Parad.	Costos de inventario	Descomposición	Agrupar	Otros modelos
Trudeau y Dror	1992	Uno a muchos	Uno	Incierta	Variable	Diferente	Tiempo	Igual	SI	No	Dado	No	SI	No	SI	Agrupar I - Rutear II	Tiempo	Heurístico y MIP
Anily y Federgruen	1993	Uno a muchos	Uno	Conocida	Constante	Diferente	Frecuencia	Igual	No	SI	No restringido	Por viaje	SI	No	SI + desabasto	Agrupar I - Rutear II	Distancia	Heurístico asintóticamente óptimo
Campbell	1993	Uno a uno	Uno	Conocida	Constante	Igual	Frecuencia	Diferente	No	No	No restringido	No	SI	SI	SI	Agrupar I - Rutear II	Distancia	EOQ
Daganzo y Newell	1993	Uno a uno	Muchos	Conocida	Constante	-	Frecuencia	No	No	No	No restringido	Por viaje	No	No	SI	No	No	EOQ
Ernst y Pike	1993	Uno a uno	Uno	Incierta	Variable	-	Frecuencia	Igual	SI	No	No restringido	SI	SI	SI	SI + desabasto	No	No	EOQ
Farvolden, Laporte y Xu	1993	Uno a muchos	Muchos	Conocida	Variable	Diferente	Tiempo	Igual	No	No	Dado	Por viaje	No	No	SI	Agrupar I - Rutear II	Tiempo	MIP y heurístico
Anily	1994	Uno a muchos	Uno	Conocida	Constante	Diferente	Frecuencia	Igual	No	No	No restringido	Por viaje	SI	No	Dependiente del vendedor	Agrupar I - Rutear II	Distancia/razón de costo	Heurístico asintóticamente óptimo
Chandra y Fisher	1994	Uno a muchos	Muchos	Conocida	Variable	Diferente	Tiempo	Igual	No	No	No restringido	Por viaje	SI	No	SI	No	No	Heurístico y MIP
Chen, Egbelu y Wu	1994	Muchos a muchos	Muchos	Conocida	Constante	Diferente	Tiempo	No	No	SI	No restringido	Por viaje	SI	No	SI	No	No	Heurístico y MLP
Popken	1994	Muchos a muchos	Uno	Conocida	Constante	Igual	Frecuencia	Igual	No	No	Var. de dec.	No	SI	No	SI	No	No	EOQ
Speranza y Ukovich	1994	Uno a uno	Muchos	Conocida	Constante	Diferente	Frecuencia	Diferente	No	No	No restringido	Por viaje	No	No	SI	No	No	MIP
Bramel y Simchi-Levi	1995	Uno a muchos	Uno	Conocida	Constante	Diferente	Frecuencia	Igual	No	No	No restringido	Por viaje	SI	No	Dependiente del vendedor	Agrupar I - Rutear II	Distancia	Heurístico asintóticamente óptimo
Bertazzi, Speranza y Ukovich	1996	Uno a muchos	Muchos	Conocida	Constante	Diferente	Frecuencia	Diferente	No	No	Var. de dec.	Por viaje	SI	No	SI	Agrupar I - Rutear II	Frecuencia	MIP
Blanchini <i>et al.</i>	1996	Muchos a muchos	Uno	Desconocida	Variable	Diferente	Tiempo	Diferente	SI	-	No	No	No	No	No	No	No	Programación dinámica
Blanchini, Rinaldi y Ukovich	1996	Muchos a muchos	Uno	Desconocida	Variable	Diferente	Tiempo	Diferente	SI	SI	-	No	No	No	No	No	No	Programación dinámica
Dror y Trudeau	1996	Uno a muchos	Uno	Incierta	Variable	Diferente	Tiempo	Igual	SI	No	Dado	Por viaje	SI	No	SI + desabasto	No	Tiempo	Heurístico
Federgruen y Tzur	1996	Uno a muchos	Uno	Incierta	Variable	Diferente	Tiempo	Diferente	SI	SI	Dado	Por viaje	SI	No	Dependiente del vendedor + desabasto	Agrupar I - Rutear II	Distancia	Programación dinámica
Herer y Levy	1997	Uno a muchos	Uno	Incierta	Constante	Diferente	Frecuencia	Igual	SI	No	Dado	SI	SI	SI	SI + desabasto	Distancia	Frecuencia	heurísticas, simulación
Blanchini, Rinaldi y Ukovich	1997	Muchos a muchos	Uno	Desconocida	Variable	Diferente	Tiempo	Diferente	No	SI	-	No	No	No	No	No	No	Programación dinámica
Bard, Huang, Jaillet y Dror	1998	Muchos a muchos	Uno	Incierta	Variable	Diferente	Tiempo	Igual	No	No	Dado	No	SI	No	No	Tiempo	Tiempo	Heurístico
Chan, Federgruen y Simchi	1998	Uno a muchos	Uno	Conocida	Constante	Diferente	Frecuencia	SI	SI	No	Dado	Por viaje	SI	No	No	No	Distancia	EOQ

Fuente: Elaboración propia con base en Batta *et al.* (1998).

Tabla 3 Clasificación de artículos relativos al IRP (continuación).

Autor	Año	Estructura de nodos terminales	Prod.	Demanda				Capacidades				Costos de transportación			Método de solución			
				Conocida	Tiempo	Dist. uniforme	Dominio de decisión	Vehículo	Invent.	Sum.	No. de vehículos	Fijo	Dist.	Parad.	Costos de inventario	Descomposición	Agrupar	Otros modelos
Christiansen	1999	Muchos a muchos	Uno	Conocida	Constante	Diferente	Tiempo	SI	SI	SI	Dado	SI	SI	SI	No	No	Tiempo	Dantzig-Wolfe
Reiman, Rubio y oEIN	1999	Uno a muchos	Uno	Incierta	Variable	Diferente	Frecuencia	SI	SI	No	Dado (uno)	SI	SI	No	No	No	Frecuencia	Prog. dinámica est.
Qu, Bookbinder e Iyoung	1999	Uno a muchos	Muchos	Incierta	Constante	Diferente	Tiempo	no	No	No	Dado	Por viaje	SI	SI	SI + desabasto	Agrupar I - Rutear II	Distancia	Heurístico de descomposición
Hwang	1999	Muchos a muchos	Uno	Conocida	Constante	Diferente	Tiempo	SI	SI	SI	No restringido	No	SI	SI	No	Agrupar I - Rutear II	Distancia	Heurístico
Achulan	2000	Uno a muchos	Uno	Conocida	Constante	Diferente	Tiempo	SI	SI	SI	Var. de dec.	Por viaje	SI	No	SI	No	No	Bifurcación y Corte (MIP)
Berman y Larson	2001	Uno a muchos	Uno	Incierta	Constante	Diferente	Tiempo	SI	SI	No	No restringido	SI	SI	No	No	No	Frecuencia	Programación dinámica est.
Chan, Carter y Burnes	2001	Muchos a muchos	Uno	Conocida	Constante	Diferente	Tiempo	SI	No	No	Var. de dec.	Por viaje	SI	SI	No	Agrupar I - Rutear II	Distancia	Heurístico y IP
Coustineau-Ouimet	2002	Uno a muchos	Uno	Conocida	Constante	Diferente	Frecuencia	SI	SI	No	Dado	No	SI	SI	No	Agrupar I - Rutear II	Distancia	Búsqueda Tabú
Kleywet, Nori y Savelsbergh	2002	Uno a muchos	Uno	Desconocida	Constante	Diferente	Tiempo	SI	SI	No	Dado	Por viaje	SI	No	SI + desabasto	No	No	Proc. de Dec. de Markov
Bertazzi, Paletta y Speranza	2002	Uno a muchos	Uno	Conocida	Variable	Diferente	Frecuencia	SI	SI	SI	Uno	No	SI	No	SI	No	No	Heurístico
Jaillet, Bard, Huang y Dror	2002	Uno a muchos	Uno	Desconocida	Variable	Diferente	Frecuencia	SI	SI	SI	Dado	SI	SI	No	SI	No	No	Sim. Monte Carlo
Campbell, Clarke y Savelsbergh	2002	Uno a muchos	Uno	Conocida	Constante	Diferente	Tiempo	SI	SI	No	Dado	Por viaje	SI	No	No	Agrupar I - Rutear II	Distancia	Heurístico y IP
Adelman	2003	Uno a muchos	Muchos	Conocida	Constante	Diferente	Frecuencia	SI	SI	SI	No restringido	SI	SI	No	No	No	No	Proc. de Dec. Semi-Markoviano y Prog. Dinámica
Kleywet, Nori y Savelsbergh	2004	Muchos a muchos	Uno	Desconocida	Variable	Diferente	Frecuencia	SI	SI	SI	Dado	Por viaje	SI	No	SI	No	No	Mod. de Dec. de Markov y Prog. Din. Heurísticos: inserción, Búsqueda Tabú
Rudriansyah y Tsao	2004	Uno a muchos	Uno	Conocida	Constante	Diferente	Frecuencia	SI	No	SI	Dado	No	SI	No	SI	No	No	Probabilidad de abasto y tasas de demanda
Anily y Bramel	2004	Uno a muchos	Uno	Conocida	Constante	Diferente	Tiempo	SI	No	No	Ilimitado	SI	SI	SI	SI	Agrupar I - Rutear II	Tasas de demanda	FFP Simulación y procesos de dec. de Markov
Cheng y Duran	2004	Uno a muchos	Uno	Incierta	Variable	Diferente	Tiempo	SI	SI	SI	Var. de dec.	SI	SI	No	SI + desabasto	Agrupar I - Rutear II	Tasas de demanda	Programación de entregas con heurístico global
Campbell y Hardin	2005	Uno a muchos	Uno	Conocida	Constante	Diferente	Frecuencia	No	No	No	Var. de dec.	No	SI	No	No	Subproblemas de entregas directas	Tasas de demanda	

Fuente: Elaboración propia con base en Batta et al. (1998).

Tabla 3 Clasificación de artículos relativos al IRP (continuación).

Capítulo 2

Métodos y estrategias de solución del IRP.

Mediante la extensa revisión de literatura relativa al IRP, se conocen sus diversas formulaciones y técnicas de solución. Esta información es muy valiosa para definir claramente las direcciones que se deben tomar, al proponer formas nuevas o mejoradas de solución del problema.

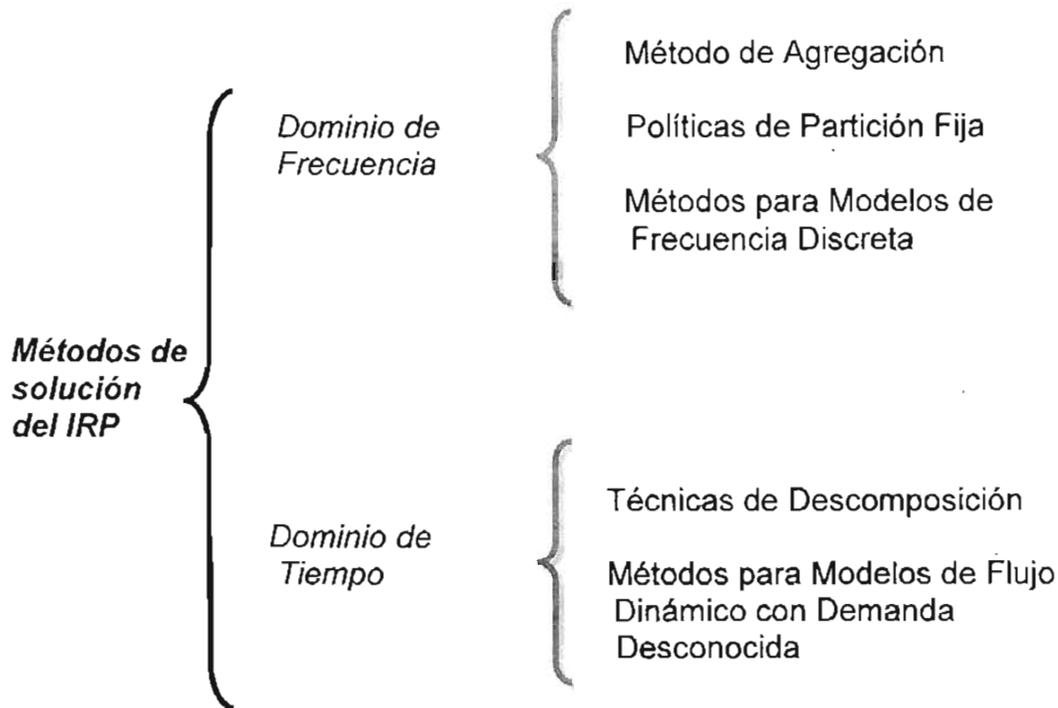
Debido a que la modelación del IRP es relativamente reciente, no es posible decir aún que existe una formulación general estándar. Si bien este problema integra consideraciones de abastecimiento de inventario al VRP en un único modelo, las decisiones de ruteo dependen de la selección de clientes, y la selección de clientes, a su vez, depende de la cantidad de producto que un vehículo (o vehículos) puede transportar. Esta interdependencia entre selección de clientes y ruteo de vehículos es la esencia del IRP.

El contenido de este capítulo se construye con base en la revisión de múltiples trabajos relacionados con el IRP y muestra el panorama completo de los métodos y estrategias de solución existentes hasta el momento.

Con base en la clasificación desarrollada por Baita *et al.* (1998), es posible dividir dichos métodos en dos grupos principales: los métodos bajo dominio de frecuencia, que incluyen a los métodos de agregación, métodos de políticas de partición fija y métodos para modelos de frecuencia discreta; y los métodos bajo dominio de tiempo que engloban a las técnicas basadas en métodos de solución de VRP, técnicas de descomposición y los métodos para modelos de flujo dinámico con demanda desconocida. La clasificación anterior puede verse gráficamente en la figura 5. Como parte final del capítulo se presenta un análisis de las ventajas y desventajas de los métodos y estrategias de solución clasificados.

2.1 Métodos bajo Dominio de Frecuencia.

En los métodos bajo dominio de frecuencia las variables de decisión son precisamente las frecuencias de reabastecimiento o tiempo entre envíos. En este rubro se identifican tres líneas: 1. Métodos de Agregación, 2. Métodos basados en Políticas de Partición Fija y 3. Métodos para modelos de Frecuencia Discreta.



Fuente: Elaboración propia con base en Baita et al. (1998).

Fig. 5 Métodos de solución del IRP.

2.1.1 Método de Agregación.

Es un método jerárquico enfocado a problemas estratégicos y tácticos (Daganzo, 1996). Parte de la utilización de procedimientos tradicionales de Investigación de Operaciones (I. de O.), tratando de evitar descripciones detalladas, tanto de los problemas como de sus soluciones. Para un problema típico, en lugar de buscar una última solución basada en grandes cantidades de datos detallados y en largos tiempos consumidos en análisis numéricos, la meta es presentar soluciones razonables con tan poca información como sea posible.

El método asume el principio de que mucha información detallada puede ser pasada por alto, sin que los modelos analíticos pierdan su capacidad de visualizar soluciones implementables. De tal forma, los modelos resultantes son fáciles de resolver, logrando visualizar factores cualitativos y tomando en cuenta sólo a los más importantes que además, sean fáciles de comunicar a administradores que estén poco familiarizados matemáticamente.

Este método se aplica en general, a problemas con muchos orígenes o muchos destinos o ambos. Problemas complejos se resuelven en dos fases: primero se determinan los intervalos entre partidas de vehículos (cada cuánto deben partir los vehículos) y después las rutas (en qué secuencia visitar a los que se requiere servir). Cuando hay varios vehículos con diferentes rutas, la segunda fase se divide en: primero agrupar y después rutear, en el caso de que cada vehículo debe servir a varios clientes; en cambio, se aplica primero rutear y a agrupar después, cuando cada vehículo debe visitar a un número pequeño de clientes.

Los supuestos principales del método de agregación son:

- i) las demandas de los clientes varían lentamente en el tiempo y
- ii) los clientes están distribuidos con baja densidad de variación en el espacio, *i.e.*, se encuentran dispersos.

Asumiendo los dos supuestos anteriores se pueden incluir situaciones completamente generales que dan buenos resultados en problemas reales.

El método de agregación proporciona un detallado análisis de los costos por políticas potenciales. El objetivo minimiza los costos ocasionados durante el recorrido desde el origen hasta el destino. Dichos costos se van adicionando a medida que se avanza en la ruta de distribución, tomando en cuenta a los actores involucrados. Si no es posible obtener la minimización global, el costo agregado será, tarde o temprano, distribuido entre los involucrados.

En estos modelos el objetivo es cóncavo a partir de supuestos aceptables que permiten la naturalidad de la descomposición incluyendo las decisiones de todo o nada entre alternativas de políticas de distribución.

Un resultado utilizado comúnmente para encontrar los tamaños de los intervalos entre salidas de vehículos, es la generalización de la fórmula del tamaño económico de lote (Economic Order Quantity) o EOQ, donde dichos intervalos son una función de $\sqrt{1/d}$, donde d es la demanda de los clientes.

Los lineamientos básicos del método de agregación se discuten ampliamente en Daganzo (1990), como una alternativa a métodos numéricos de solución. El método fue introducido primero por Newell (1971) y consiste principalmente en sustituir los conjuntos de entidades discretas (tiempos de atención, intervalos de

salidas, etc.) por la búsqueda de funciones continuas, las cuales dan los valores deseados discretos con costos cercanos al costo óptimo.

Blumenfeld *et al.* (1985) y Burns *et al.* (1985) analizan la negociación entre costos de inventario y transportación para i) el caso envíos directos y ii) el caso de envíos a puntos intermedios, a través de métodos analíticos sencillos que minimizan funciones cóncavas sin restricciones, que no requieren del uso de MP.

Una contribución interesante es la que se relaciona con la división de la red en subredes separadas para facilitar el problema de ruteo, definidas por diferentes tasas de consumo para varios productos, decidiendo así, el número de puntos intermedios asociados con subsistemas de transportación, Daganzo y Newell (1986).

Blumenfeld *et al.* (1987) muestran una aplicación real de gran escala General Motors Corporation, frecuentemente mencionada en artículos. Un gran número de productos se envían desde 20,000 puntos proveedores a 160 plantas. Basado en el análisis de la negociación entre costos de inventario y de transportación, los autores desarrollaron una herramienta computacional de decisión, obteniendo un ahorro potencial estimado en 26%. La herramienta se utiliza actualmente.

Algunos autores tratan el uso de heurísticas para la definición de frecuencias óptimas, entre ellos Hall (1985) y Ernst y Pyke (1993) quienes manejan demanda estocástica. Benjamin (1989) propone una heurística para la optimización simultánea del EOQ y del problema de transporte.

Daganzo (1987), Campbell (1993) hacen un análisis de las ventajas de situaciones de transbordo y Klinecicz (1990) y Popken (1994) consideran el caso de funciones objetivo no lineales.

Un concepto interesante introducido por Bard *et al.* 1998, es el de "horizonte cíclico" (rolling horizon) y el uso de servicios satélite, en los cuales los vehículos se abastecen, para cumplir con las entregas programadas, utilizando heurísticas clásicas modificadas.

Reiman *et al.* (1999) trabajan con envíos directos en un sistema de distribución dinámico estocástico, que se define a través de tres problemas controlados dinámicamente para decidir si un vehículo en el almacén debe permanecer sin actividad o debe cargarse completamente para realizar entregas. El punto interesante son los problemas que se resuelven: en el primero, el vehículo debe recorrer una ruta de agente viajero a través de todos los clientes, especificada con anterioridad, y el controlador decide cuántas unidades debe entregar a cada cliente; en el segundo el vehículo entrega una carga completa a un cliente, lo que es un envío directo, y el controlador decide a cuál cliente se debe visitar a continuación; el tercero permite la elección dinámica entre el envío directo y el agente viajero.

En su trabajo, Jaillet *et al.* (2002) continúa con el concepto de horizonte cíclico y servicios satélite de Bard *et al.* (1998), trabajando un problema de la repetida distribución de aceite para un largo período de tiempo y un gran número de clientes. Se tiene un almacén central, pero también varios servicios satélite que los vehículos pueden visitar durante su viaje para recarga. El consumo del producto varía diariamente y no se puede predecir de manera determinística, de tal forma se manejan consumos esperados. Una de las políticas aplicadas es que en caso de desabasto, se realiza una entrega directa, pero se incurre en un costo de penalización. Se utilizan aproximaciones de costos incrementales para usarse en el horizonte cíclico, tratando de minimizar el costo total anual de entregas. En este trabajo se pretendía justificar las derivaciones de costo incrementales usadas en el marco de horizonte cíclico y una de las principales aportaciones fue la habilidad de reducir el problema de una base de período de tiempo anual a un período de planeación cíclico de dos semanas.

Un modelo y solución para el IRP con ventanas de tiempo, denominado Integrated Inventory and Periodic Vehicle Routing Problem with Time-windows, es desarrollado en Rusdiansyah y Tsao (2004). Se obtiene la frecuencia de visitas a cada vendedor construyendo un conjunto de tours para minimizar el promedio de los costos de transportación e inventario en un horizonte cíclico dado de m -días. Para este complejo caso (ventanas de tiempo) se desarrolla un heurístico que dividido en dos fases, una de solución inicial y otra de mejora, utiliza a su vez cuatro heurísticas: de selección, de inserción, intercambio y Búsqueda Tabú. Se obtuvieron ahorros sustanciales comparados con la mejor solución conocida para IRP con ventanas de tiempo.

El trabajo desarrollado en Campbell y Hardin (2005), resuelve un problema basado en un IRP en el que los clientes consumen bienes de tal forma que es posible minimizar el número de vehículos requerido para hacer entregas directas y periódicas a un solo cliente a la vez. Así, es posible descomponer en varios subproblemas que tiene a un solo punto de entrega, asumiendo que la demanda del cliente es cuando mucho la capacidad del vehículo y que la entrega requiere de no más de un día para realizarse. La ventaja de entregas directas mostrada por Gallego y Simchi-Levi (1990) es importante de la justificación de este trabajo. A través de un algoritmo de programación heurístico glotón, se obtiene una programación de entregas que resulta se óptima para algunos casos tratados.

En particular el éxito del método de agregación puede verificarse en problemas de gran escala donde la variabilidad local es compensada por comportamientos comunes promedio. Como consecuencia de lo anterior, el método de agregación resulta más ventajoso para el tipo de situaciones donde métodos más detallados son inconvenientes debido a la gran cantidad de datos y recursos de cómputo requeridos. El método produce modelos analíticos, que aunque no requieren demasiados recursos de computación, sí requieren habilidades para construir e interpretar los modelos. La forma de utilizar el método de agregación, adaptándose a diferentes situaciones de manera muy específica, muestra ser más

un arte que un método completamente formalizado, resultando muy ventajoso en circunstancias particulares.

2.1.2 Métodos basado en Políticas de Partición Fija (FPP).

La estrategia de Políticas de Partición Fija o FPP (Fixed Partition Policies), realiza anticipadamente una partición de los clientes en regiones.

Sea Π el conjunto de todas las particiones. Una estrategia en Π se considera como óptima; si, con respecto a la capacidad de los vehículos, ésta define regiones que minimizan el promedio de la suma de los costos de mantenimiento de inventarios y de transportación. Determinar una estrategia óptima en Π es claramente un problema NP-duro, ya que requiere definir el óptimo ciclo Hamiltoniano entre la región considerada y el almacén central. Sin embargo, la estructura a largo plazo de los costos promedio de cada región es una generalización directa del EOQ, cuando se expresa en términos del número de clientes y la longitud del ciclo Hamiltoniano. A largo plazo las particiones resultan asintóticamente óptimas.

Una vez definida la partición de los puntos de demanda en regiones, se determina el ciclo óptimo y la frecuencia de visitas para cada región por separado, tomando en cuenta que cada vez que un vehículo visita una región, todos los clientes o vendedores minoristas pequeños que lo forman serán atendidos. Sin embargo, no hay coordinación entre los tiempos de visita para las diferentes regiones (Bramel y Simchi-Levi, 1995).

En forma particular, el ciclo óptimo se determina con alguna heurística para el TSP, y la frecuencia de visitas óptima se puede determinar con la fórmula para EOQ, considerando a los costos de transportación como los costos fijos componentes. Uno de los méritos principales de la estrategia de FPP, al menos desde el punto de vista de los administradores, es que resulta fácil de implementar (Bramel y Simchil Levi 1995, Anily y Federgruen, 1990a).

Anily y Federgruen (1990a) buscan la estrategia de reabastecimiento óptima en Π , para el caso de horizonte infinito con demandas constantes. El problema se acota a la situación de todos los clientes que tienen la misma tasa de demanda, asumiendo posiblemente que ésta se constituye por una serie de distribuidores que están localizados en el mismo punto. Un cliente podría ser asignado a diferentes regiones. Las regiones se obtienen heurísticamente con el llamado esquema modificado circular de partición de regiones, que se comporta asintóticamente óptimo en Π cuando las distancias entre los clientes y el almacén son aleatorias, están independiente e idénticamente distribuidas (i.i.d.) y el número de puntos de distribución crece al infinito. Un caso en el que el número de regiones es fijado *a priori* se considera en Anily y Federgruen (1991b), no hay almacén central, sino varios distribuidores.

Anily y Federgruen (1993) generalizan el problema anterior considerando un almacén central desde el que se distribuyen los productos. El almacén paga costos de mantenimiento de inventario y tiene cantidades de inventario limitadas, por ello debe ser reabastecido continuamente enfrentando costos, tanto fijos como proporcionales. En una estrategia FPP similar, aplicada en Anily y Federgruen (1990a), los intervalos entre dos visitas a la regiones se redondean a un vector de tiempo de segundo orden. Con el usual supuesto de que las posiciones de los clientes tienen una distancia i.i.d. desde el almacén y el número de puntos de distribución crece hasta el infinito, tal estrategia está asintóticamente en el 6% de optimalidad en Π .

El método FPP permite una modelación extremadamente general de los costos de ruta, tales como costos por secuencias de arcos y tipo de vehículos. Por ejemplo, los costos por parada no se pueden tomar en cuenta, ya que la venta de un cliente puede estar subdividida en distribuidores (muchos a uno). Esta es una diferencia importante con respecto al método de agregación de Danganzo (1996), donde lo que él hace es aplicar estrategias que minimicen el número de paradas por entregas y no permite la división de carga entre vehículos, a pesar de ello, en algunos casos puede ser posible reducir el número de tours (ciclos) y la distancia viajada por cargas divididas. Como consecuencia de lo mencionado, en las FPP la división de carga entre vehículos es permitida.

Hall (1991) muestra en el uso de FPP, que asignando a diferentes regiones los puntos de distribución correspondientes a un único cliente y coordinando entregas en tales regiones, se puede inducir una reducción de costos adicional. Apunta, además, que la debilidad de coordinación de las entregas, no minimiza los costos relacionados con el espacio y servicios necesarios para mantener el máximo nivel de existencias en el almacén de un cliente.

Simchi-Levi (1995), considera FPP donde una heurística de localización es utilizada para determinar las regiones. Este método consiste en buscar una solución aproximada en tiempos razonables para un VRP capacitado que es NP-duro. Los costos considerados son costos de mantenimiento del cliente, costos fijos de ordenación, y costos de transportación proporcionales a la distancia. Todos los costos y demandas son dependientes de ese cliente y son constantes. No existe límite en el monto del inventario de los clientes, ni en la capacidad de abastecimientos del almacén central.

Utilizando la misma hipótesis al considerar los costos, demandas y capacidades, Gallego y Simchi-Levi (1990) estudian cotas de efectividad para envíos directos contra estrategias de paradas múltiples. Claramente se observó que la política de envíos directos está muy cercana al óptimo, cuando el tamaño económico de lote de los vendedores minoristas es cercana a las capacidades de los camiones, por ejemplo, un error menor al 6% se obtiene cuando el tamaño del lote es al menos 71% de la capacidad del camión. Sin embargo, Hall (1992) muestra que los envíos directos proporcionan ese buen desempeño únicamente cuando los costos fijos de

transportación son insignificantes. En otras circunstancias, la estrategias de parada múltiple pueden significar ahorros.

Herer y Levi (1997) presentan una heurística que combina versiones modificadas de la heurística clásica de Clarke y Wriqth, con el objetivo de determinar cuándo servir a cada usuario, así como la ruta que debe recorrer cada camión para minimizar los costos. El algoritmo desarrolla el concepto de distancias temporales, que es el costo de agrupar a dos consumidores dentro de un período común.

Desde el punto de vista computacional, el método de FPP es útil para el manejo de problemas de tamaño importante (más de miles de clientes). Más aún, algunas variantes del modelo han sido introducidas (Anily y Federgruen, 1990a, Chan *et al.* 1998), tales como vehículos no capacitados, vehículos no idénticos, restricciones de volumen de venta y opciones de pedidos pendientes.

Es importante tomar en cuenta que las FPP sólo proporcionan optimalidad asintótica. De hecho esto puede proporcionar casi triviales y claramente malas soluciones para instancias de tamaño modesto. Además, ningún problema de tamaño mínimo es claramente señalado, para el cual se hayan dado soluciones aceptables en la práctica. Trabajar en esta interesante línea de investigación podría dar resultados útiles para tales problemas (Baita *et al.* 1998).

La limitante principal del método FPP es mostrada por Hall (1992): el método no puede modelar el hecho de que las entregas estén coordinadas. Así, una consecuencia es que los resultados que se dan son válidos únicamente en el caso de entregas independientes, o pueden ser apenas considerados como cotas superiores de los costos reales. La modelación de la coordinación de entregas es uno de los factores más complejos en los métodos de frecuencia del IRP.

2.1.3 Métodos para Modelos de Frecuencia Discreta.

La principal desventaja de los métodos basados en el EOQ es el hecho de que proporcionan intervalos de entrega que pueden ser no enteros. Algunos autores, por ejemplo Hall (1985), Maxwell y Muckstadt (1985) y Jackson *et al.* (1988), reconocen lo impráctico de este resultado.

Hall (1985) sugiere redondear el valor de EOQ al valor de frecuencia factible más cercana. Esta es una aproximación de aparente sentido común, sin embargo, puede llevar a costos que son mucho más grandes de lo necesario, Speranza y Ukovich (1994a) proporcionan un ejemplo simple donde el redondeo eleva los costos 20% por encima del mínimo actual.

Una posibilidad generalizada es presentada por Speranza y Ukovich (1994b), para la distribución de diferentes productos desde un origen a un destino usando camiones de capacidad dada. El mismo producto puede ser dividido en diferentes frecuencias de entregas y diferentes productos pueden compartir la misma

frecuencia. Los envíos que tienen la misma frecuencia se suponen síncronos, *i.e.*, realizados en el mismo momento. Un modelo de MIP es propuesto para asignar productos a frecuencias. Si bien el problema con frecuencias discretas es NP-duro, un algoritmo para él se formula en Speranza y Ukovich (1996) el cual muestra eficiencia computacional en la práctica.

El caso uno a uno con frecuencias discretas se ha generalizado al caso de varios destinos por Bertazzi *et al.* (1997). Ellos proponen un conjunto de heurísticas de descomposición que prueban arista por arista de la red de distribución (*i.e.* envíos directos) obteniendo una solución al problema y entonces realizan una búsqueda local para mejorar la solución por medio de la consolidación, agrupando con base en frecuencias de envío. Después, en una segunda fase, los clientes visitados en la misma frecuencia son considerados por agregación en esa misma ruta. En una tercera fase, clientes con la misma frecuencia son consolidados de nuevo y las rutas son designadas. Diferentes variantes del esquema anterior se utilizan en otros desarrollos y se evalúa su eficiencia.

Considerando la aplicabilidad práctica de los modelos de frecuencia discreta, resulta claro que su principal mérito es considerar únicamente aquellos valores de frecuencia que son implementables. En general los algoritmos que se han propuesto, muestran buen desempeño en términos de eficiencia (tiempo computacional) y en términos de efectividad para problemas de tamaño considerable.

La principal desventaja de los modelos de frecuencia discreta es la misma que para FPP, *i.e.*, la imposibilidad de tratar con entregas coordinadas. Para este caso, el considerar coordinación podría requerir la introducción de más variables que deberían tomarse en cuenta para la fase de coordinación de frecuencias, pero los modelos obtenidos resultan ser extremadamente complejos e intratables.

2.2 Métodos bajo Dominio de Tiempo.

En los métodos de dominio de tiempo para el IRP, las cargas, rutas y frecuencias de visitas, son decididas en una actitud de planeación *a priori* para intervalos fijos de tiempo. Como resultado, operaciones periódicas pueden ser propuestas.

Contrariamente al dominio de frecuencia, las operaciones son programadas a través de un horizonte posiblemente infinito. En principio las decisiones se tomarían una sola vez. Sin embargo, el caso más interesante es cuando con un modelo de tiempo discreto, las decisiones se toman al inicio de cada espacio de tiempo (por ejemplo cada día o cada semana), conociendo el estado del sistema (los niveles de inventario). Este es un método de loop cerrado, en el sentido de que hay una retroalimentación de las decisiones tomadas en un período, que afectan las decisiones del siguiente período.

El método de dominio de tiempo es idóneo para situaciones en las cuales las condiciones pueden variar significativamente en el tiempo. Sin embargo, tal flexibilidad es contrarrestada por la necesidad de recalcular diferentes soluciones en cada intervalo de tiempo.

Como se mencionó anteriormente se puede interpretar al IRP como un enriquecimiento del VRP que incluye restricciones de inventario. En este sentido, una de las áreas más desarrolladas e interesantes en la solución del IRP, es la aplicación de técnicas que han sido ampliamente utilizadas para solucionar el VRP (Toth, 2002). De tal forma, una amplia variedad de técnicas que incluyen métodos de optimización exactos, métodos heurísticos clásicos y modificados, metaheurísticas e híbridos que combinan técnicas exactas con heurísticas.

Federgruen *et al.* (1986), presenta el caso donde productos de dos diferentes niveles de calidad son considerados. En este caso, los costos, pero no las restricciones, dependen del nivel de calidad de los productos. Se proponen métodos de solución basados en heurísticas y programación matemática, trabajando juntos con resultados computacionales que involucran más de 75 clientes. Más recientemente Federgruen y Tzur (1996) estudian el mínimo horizonte de pronóstico para la solución de horizonte cíclico de problemas generales de programación dinámica.

Golden *et al.* (1984) desarrolló una heurística para la optimización de un sistema de planeación de entregas integrado para una gran compañía de energéticos de distribución de propano líquido. El estudio fue hecho para un distrito de distribución que servía aproximadamente a 3,000 clientes. El propósito del estudio fue comparar las reglas de distribución usadas por la compañía con el algoritmo heurístico propuesto por los autores. Las reglas de distribución de la compañía se basaban en los puntos de reabastecimiento para cada cliente. El horizonte de tiempo se ajustó a un solo día. Introdujeron el concepto de "urgencia de un cliente", para decidir a qué clientes incluir en la planeación de ese período. La urgencia de un cliente fue definida como la razón del nivel actual del tanque y el tamaño del mismo. Con base en datos históricos, la compañía calculaba una tasa promedio de consumo y mantenía un registro de la última fecha de abastecimiento, para cada cliente. Esta información fue usada para decidir el momento en que el cliente llegaría al punto de reabastecimiento y el siguiente reabasto era programado. Los autores desarrollaron una heurística que incluía un algoritmo de selección de clientes que elegía el conjunto de clientes que debían ser visitados cada día a un costo efectivo. La heurística también obtenía la asignación de rutas y camiones, usando un esquema de inserción similar a los utilizados para VRP. Los clientes eran insertados en un creciente tour, considerando el valor de la razón de urgencia, así como el tiempo de viaje. El gigantesco tour se dividía en partes para cada vehículo. Cada uno de los componentes del problema se resolvió de forma efectiva y realista y un experimento de simulación fue usado para evaluar el desempeño integrado de los componentes. Los resultados de la comparación simulada de la heurística propuesta con las reglas de distribución utilizadas por la compañía, mostraron que

la heurística tenía un desempeño superior. El número de galones/hora enviados fue mejorado en 8.4%, el faltante de producto se redujo en 50% y los costos totales se redujeron en 23%.

El método de dominio de tiempo se utiliza en una serie de artículos por Dror y Ball (1987), Dror y Levy (1986), Dror y Trudeau (1996), y Trudeau y Dror (1992). En cada intervalo de tiempo, únicamente los clientes que alcanzan su nivel de seguridad de existencias se visitan. Un sólo producto se debe entregar desde un almacén a varios clientes, cuyas demandas son diferentes en cada período y son determinísticas en Dror y Ball (1987) y Dror y Levy (1986), estocásticas en Trudeau y Dror (1996), y tanto determinísticas como estocásticas en Dror y Trudeau (1996). Dror y Levy (1986) y Dror y Trudeau (1996) consideran el problema de distribución de propano en un período largo, desde una perspectiva de valor presente de flujo de efectivo. Ellos se enfocan en la maximización de la eficiencia operacional (cantidad promedio de entrega en una hora de operación) y la minimización de la cantidad promedio de producto no abastecida en un período. Dror y Ball (1987) proponen un modelo de MIP donde el efecto de las presentes decisiones en subsecuentes períodos son tomados en cuenta como factores de penalizaciones e incentivos. Trudeau y Dror (1992) desarrollan algunas heurísticas, las cuales usan submodelos lineales de MIP, para resolver su problema. Basado en las ideas anteriores, un sistema de distribución computarizada para un producto (aceite para calentador) evaluado con datos actuales para más de 2000 clientes, se presenta en Dror y Trudeau (1988).

Miller (1987) elige los clientes por visitar en cada ruta de acuerdo con sus tiempos de desabasto. El problema es cómo programar cada envío para que las necesidades de inventario de todas las terminales sean satisfechas durante un horizonte de planeación predefinido, con el objetivo de minimizar los costos de transportación y de inventario. Se propuso un modelo de programación binaria para la solución del modelo.

Farvolden *et al.* (1993) tratan con un problema actual de distribución a gran escala, desde un almacén a varios clientes (alrededor de 165 productos diferentes y 400 clientes, algunos de los cuales tienen restricciones de ventanas de tiempo). El problema es eventualmente reducido a un problema de ruteo de vehículos y es resuelto adaptando las heurísticas descritas por Salomon (1987) y Potvin *et al.* (1992).

Los artículos de Chen *et al.* (1994) y Chandra y Fisher (1994) coordinan la planeación de la producción y la distribución. Proponen dos modelos integrados de NLP basados en dos alternativas de estrategias de compra de materia prima y desarrollan algoritmos para resolver los modelos. Algunos resultados computacionales son proporcionados para probar la efectividad de los algoritmos propuestos. En especial, Chandra y Fisher (1994) siguen dos métodos: uno en el cual la programación de la producción y los problemas de ruteo de vehículos se resuelven por separado y otro método en el cual están coordinados en un solo

modelo. Algunos experimentos computacionales muestran que tal coordinación puede reducir costos.

Hwang (1999) muestra una aplicación muy interesante de distribución de alimentos para la mitigación de la hambruna en Corea del Norte, que resuelve independientemente los problemas de abastecimiento de inventario, distribución a grupos de sectores y programación de rutas, utilizando heurísticas modificadas (Clarke y Wright y Sweep).

Un trabajo presentado por Achutan (2000) en APORS2000, desarrolla un algoritmo de Bifurcación y Corte para resolver un MIP de un solo producto. El modelo se construye para un número máximo de 4 períodos de tiempo (días), utilizando variables binarias para la atención a clientes y uso de vehículos en un periodo, además de variables reales para cantidades de abastecimiento y niveles de inventario finales, también por cada período. El algoritmo fue desarrollado en lenguaje C utilizando CPLEX para resolver las respectivas relajaciones de LP. Las violaciones ocurren cuando se excede la capacidad del vehículo o cuando la solución óptima de un problema relajado contiene un subtour. Los problemas de prueba fueron generados aleatoriamente. Aunque el modelo obtuvo la optimalidad para este problema, no es posible todavía resolver problemas grandes, de hecho, el número mayor de nodos solucionado fue de 5.

Bertazzi, Paletta y Speranza (2002) consideran un modelo multiperíodo con demanda determinística, en el cual un producto debe ser enviado desde un proveedor a varios puntos de venta en un horizonte de tiempo dado. Estos envíos se realizan en un vehículo con un costo y capacidad dados. Cada punto de venta determina un nivel mínimo y máximo de inventario y debe ser visitado antes de que se alcance el nivel mínimo. Cada vez que se visita al vendedor, la cantidad de producto entregada es tal que se llega al nivel máximo de inventario, esto con base en la política clásica order-up-to, ampliamente estudiada en la teoría de inventarios. Lo que se busca es determinar para cada instante discreto de tiempo, qué vendedores serán visitados y la ruta del vehículo. Los objetivos del trabajo son resolver el problema y determinar el impacto que diferentes funciones objetivo (únicamente costos de transporte, únicamente costos de inventario de los vendedores y suma de costos de transporte más costos de inventario del proveedor) en la solución del problema. Utilizan una heurística constructiva basada en ruta más corta, que en cada iteración inserta un vendedor en la solución, evaluando el costo incremental de la inserción. Aunque el problema tratado es para un producto y un solo vehículo, la heurística propuesta permite fácilmente la extensión a múltiples productos y vehículos. Los problemas prueba son generados aleatoriamente.

En otra interesante aplicación, Campbell *et al.* (2002) muestran un método de solución que resuelve por separado los problemas de atención a clientes y de programación y ruteo, para un horizonte de un día, utilizando una combinación de la heurística GRASP e IP, para una aplicación real de distribución de gas industrial de la compañía PRAXAIR.

Cousineau-Quimet (2002) aplica la conocida metaheurística Búsqueda Tabú al IRP. Búsqueda Tabú ha mostrado dar muy buenos resultados para VRP (Toth, 2002). La aplicación para problemas generados aleatoriamente mostró flexibilidad y eficiencia para problemas de múltiples clientes y múltiples tamaños de flota. Sin embargo, resultó limitado para la longitud del período de planeación y realmente se considera en estado inicial de investigación.

Para el dominio de decisión de tiempo, Anily y Bramel (2004) presentan un excelente trabajo en el que definen una cota inferior asintótica 98.5%-efectiva en políticas de partición fija para minimizar los costos promedio de inventario y transportación IRP. Un análisis probabilístico del desempeño de esta cota demuestra que al ser 98.5%-efectiva, a medida que el número de vendedores se incrementa, la cota inferior será cada vez más cercana al costo de la política de partición fija óptima.

Una aplicación del transporte mundial de petróleo crudo es desarrollada por Cheng y Duran (2004) que por medio de la combinación de simulación de eventos discretos con el control óptimo estocástico (un tipo de proceso de decisión de Markov), apoya la toma de decisiones para mejorar el desempeño del sistema. Debido al gran espacio de estados del problema de control óptimo, se utilizaron rigurosos métodos de programación dinámica, pero en muchos casos resultaron intratables. El modelo de simulación que representa el comportamiento del sistema real de inventario-transportación se construyó en Arena junto con módulos de interfase en Visual Basic y Excel.

2.2.1 Técnicas de Descomposición.

Dentro de los métodos bajo dominio de tiempo, se identifica una clasificación que aplica técnicas de descomposición en la solución del problema. Como se verá en el Capítulo 3, una técnica de descomposición es utilizada en la estrategia propuesta. De tal forma que es necesario hacer un análisis de los desarrollos de solución del IRP que utilizan estas técnicas.

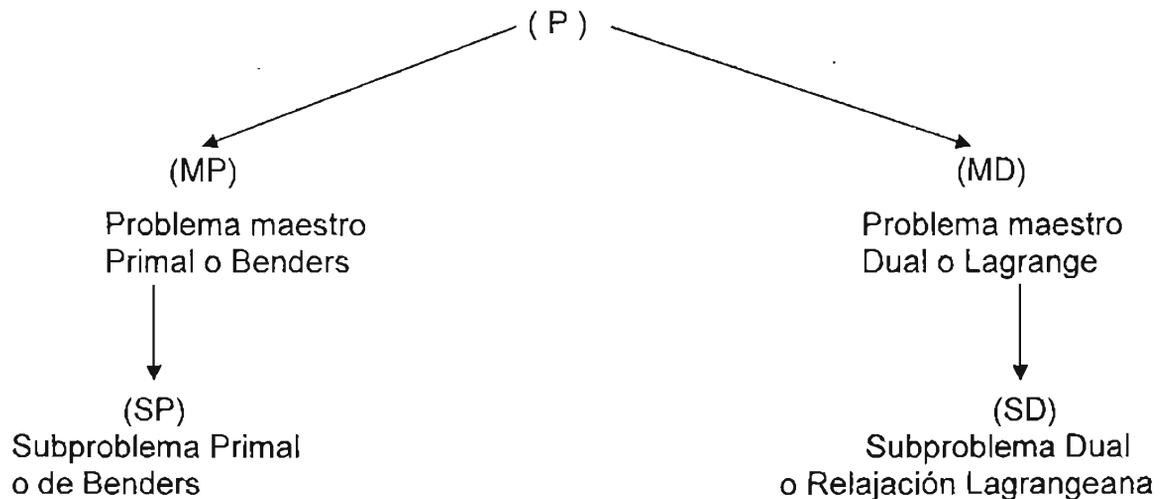
Las llamadas técnicas de descomposición, se utilizan para resolver problemas de gran escala, así como problemas de LP con una estructura especial. Estas técnicas, se caracterizan por una descomposición del sistema original en subsistemas, cada uno menor e independiente. Los subproblemas se consideran manejables a partir de su tamaño, pero juntos constituyen un problema que excede la capacidad de cómputo disponible o tiene una gran complejidad.

El principio de descomposición, incluye las formas primal y dual, llamadas Descomposición de Benders y Descomposición Dual Lagrangeana (Dantzig-Wolfe), respectivamente. Estas técnicas permiten tomar ventaja de la estructura especial de un problema al resolver una secuencia de problemas más simples. De esta forma, se explota la subestructura primal o dual del problema. De hecho, tales

estrategias son consideradas duales una de otra. En la Fig. 6 se puede observar el mapa general de transformaciones de la técnica de descomposición.

Para ambos esquemas el ciclaje se da entre el problema maestro y el subproblema.

Entre las técnicas de solución para el IRP, las técnicas de descomposición fueron propuestas y utilizadas por autores tales como Federgruen y Zipkin (1984); Chien, Balakrishnan y Wong (1989) y Christiansen (1999).



Fuente: Van Roy (1983).

Fig. 6 Mapa de transformaciones de la Técnica de Descomposición.

Como se mencionó anteriormente, Federgruen y Zipkin (1984) consideran un solo período, las demandas son aleatorias y las decisiones se toman con base en los inventarios iniciales y distribuciones de probabilidad de la demanda. Los costos se relacionan con la transportación, pero también con el nivel de inventario y escasez al final del período. Formulan un problema de MIP no lineal que se resuelve de manera exacta por medio del método de descomposición generalizada de Benders, que además proporciona una cota inferior para el costo óptimo en cada iteración.

Esta técnica puede utilizarse para IRP, considerándolo como un caso de VRP a partir de que el TSP puede ser visto como un LP, cuyo conjunto factible se define implícitamente como la cubierta convexa de todas sus soluciones factibles (llamado politopo de agente viajero). Dado que el VRP es un programa lineal entero mixto, donde las variables enteras son las y_{ik} (variable binaria que indica si el cliente i es asignado a la ruta k o no), al fijar estas y , el VRP se reduce a K problemas de agente viajero. El VRP puede entonces ser resuelto

exactamente con el procedimiento de descomposición de Benders (Federgruen y Zipkin, 1984).

La descomposición de Benders requiere soluciones duales para el o los subproblemas lineales con el propósito de generar cortes. Sin embargo, el método utilizado no requiere de planos de corte para resolver cada TSP, a menos que los (relativamente pocos) planos de corte ajustantes sean generados y ponderados una vez que la solución sea encontrada.

Si y se fija, obtenemos entonces los mismos K problemas de agente viajero más el problema de abastecimiento de inventario. Éste último es un programa no lineal, así que la descomposición de Benders no se puede utilizar. La descomposición generalizada de Benders (desarrollado por Geoffrion en 1972), fue adaptada adecuadamente para el problema.

El método puede ser resumido como sigue: por dualización se deriva un problema maestro en las variables y , equivalente al original. Una secuencia de problemas maestros relajados se resuelven. Cada uno de esos problemas da una solución tentativa para y , la cual define a los subproblemas. Los subproblemas son solucionados o determinados como infactibles. Las soluciones duales o rayos extremos definen entonces una o más restricciones, *i.e.*, los cortes, del problema maestro. Estos cortes son añadidos al problema maestro relajado previo, y el proceso continúa.

El éxito de la descomposición generalizada de Benders para un problema particular depende de varios factores. Los subproblemas deben ser relativamente fáciles de resolver; los algoritmos para ellos deben producir multiplicadores óptimos (o rayos extremos, cuando sea apropiado); y no deben tener brechas de dualidad (si el problema maestro es equivalente a la versión original del problema). Los problemas de agente viajero asumidos como programas lineales satisfacen esas condiciones.

Para probar el algoritmo, se usaron dos problemas como base para los experimentos, introducidos por Christofides y Eilon (1969). Se compararon los tiempos de computadora utilizados y los costos para el problema combinado de abastecimiento/ruteo y para el VRP, únicamente. Los resultados mostraron ahorros sustanciales de 6 a 7% que pueden ser alcanzados en costos operativos, mientras que la reducción del número de camiones requerido no es menor del 20%.

Chien *et al.* (1989) proponen un método integrado para un único período del IRP y lo formulan como un programa entero mixto. Hay un almacén central con capacidades fijas de abastecimiento y muchos clientes con demanda determinística. No es necesario satisfacer la demanda completa y hay un costo de penalización por unidad de demanda no satisfecha. Se utiliza un procedimiento

basado en relajación Lagrangeana que explota la estructura especial del modelo generando buenas cotas superiores, además de soluciones heurísticas.

Qu *et al.* (1999) resuelven un modelo dinámico multiproducto con demandas aleatorias, utilizando una técnica de descomposición que separa en un problema maestro que se refiere al inventario y subproblemas que se refieren a la transportación (agente viajero). El problema de inventario trabaja con un producto a la vez para la transportación y una ruta óptima es obtenida para cada período. Esta situación hace que se puedan resolver problemas de gran tamaño.

Una interesante aplicación real en Europa se desarrolló en Christiansen (1999), en la cual se resuelve un problema de planeación de manejo de inventarios de amoníaco y de ruteo de barcos con ventanas de tiempo, utilizando el método de descomposición de Dantzig-Wolfe que define el ruteo y el manejo de inventarios como los subproblemas, y que resuelve al problema maestro con Branch and Bound (B-B).

2.2.2 Métodos para Modelos de Flujo Dinámico con Demanda Desconocida.

Una nueva y promisoría línea de investigación en el método de dominio de tiempo para el IRP trata con modelos de sistemas dinámicos con demanda desconocida.

Los enfoques iniciales consideran un sistema dinámico general, sujeto a perturbaciones desconocidas de la demanda, pero que pueden ser acotadas y situadas en intervalos conocidos. El problema ha sido determinar una estrategia de control de retroalimentación apta para mantener el estado del sistema dentro de cotas dadas en un horizonte dado, para alguna perturbación admisible. La idea manejada fue la de una secuencia de conjuntos que en cada momento contiene a cada estado, cuya futura trayectoria resultante pueda mantenerse dentro de las restricciones, para todas las perturbaciones permitidas. Como esta estrategia de control, un denominado tubo objetivo (tubo direccional) puede ser determinado mediante una adecuada técnica iterativa, reminiscencia de algunos conceptos de programación dinámica. Sin embargo, pronto se observó que la definición de este tubo objetivo (y sobre todo su descripción) era muy complejo en el manejo de problemas de cualquier tamaño realista (Glover y Schweppe, 1971).

La aplicación al IRP es inmediata y de manera general se da de la siguiente forma: las perturbaciones pueden representar demandas, y también instancias abastecidas en intervalos inciertos; la estrategia de operación es expresada por la ley de control; las restricciones de estado representan capacidades de almacenamiento mientras que las restricciones de control representan capacidades de transportación.

Blanchini *et al.* (1996a) ha estudiado un modelo más detallado que tiene la forma de una red dinámica, *i.e.*, una red cuyos nodos tienen capacidad de almacenamiento y los flujos pueden variar a través del tiempo. Así, los nodos

corresponden a los almacenes y los arcos a las ligas de transportación; de esta forma, un sistema de distribución de red (muchos a muchos) para un único bien se presenta. Las demandas se asumen periódicas en el sentido en el que pueden tomar cualquier valor en un intervalo conocido, cuyas cotas son funciones periódicas de tiempo. El primer problema es determinar, en cada momento, los niveles de inventario que permitan enfrentar cualquier demanda futura sin desabasto o sobrantes en ningún almacén. Un problema de múltiples almacenes con estructura similar es trabajado por Chan *et al.* (2001).

Sin embargo, para el mismo tipo de sistemas, demandas con diferentes propiedades han sido estudiadas en Blanchini *et al.* (1996b). La demanda agregada, sobre todos los puntos cliente, es conocida, si bien la demanda local en cada punto es únicamente acotada superiormente. En términos prácticos, esta es la situación donde los clientes requieren cantidades conocidas de un bien, pero no conocemos el lugar específico donde ellos pueden querer ir a comprarlos. También en este caso, las capacidades de almacenamiento y transportación varían a través del tiempo de acuerdo con un patrón conocido, así como las cotas superiores para la demanda. El problema de determinar los niveles de almacenamiento inicial que puedan satisfacer cualquier nivel de demanda, se resuelve tanto para el caso de horizonte finito como infinito, con un comportamiento periódico de capacidades de almacenamiento y transportación y de cotas de demanda.

Un problema de diseño del mismo tipo de sistemas se presenta en Blanchini *et al.* (1997) para el caso estacionario, *i.e.*, cuando las cotas en las cuales la demanda puede estar, son constantes en el tiempo. El problema consiste en determinar el mínimo costo de almacenamiento y capacidades de envío que permitan la satisfacción de cualquier demanda factible sobre un horizonte infinito. La solución de este problema se encontró resolviendo dos problemas de programación convexa independientes, uno para las capacidades de almacenamiento y otro para las capacidades de envío. Aunque la formulación de los problemas es muy sencilla, el último resulta ser NP-duro, además de que puede tener un número exponencial de restricciones lineales. Una heurística es propuesta por los autores para la solución de este problema.

Como parte de las técnicas de descomposición se mencionó el trabajo de Qu *et al.* (1999), quienes resuelven un modelo dinámico multiproducto con demandas aleatorias, mediante una técnica de descomposición.

Una visión interesante y que había sido descrita en Campbell *et al.* (1998) y que desarrollan Kleywegt *et al.* (2002), es el desarrollo del proceso de decisiones como un proceso de decisión de Markov de tiempo discreto en donde se debe decidir el itinerario óptimo, entendiendo como itinerario a la decisión que se toma partiendo de la observación del inventario de cada cliente al principio del período (un día). Las decisiones incluyen a qué cliente abastecer, cuánto entregar a cada cliente, cómo combinar clientes en rutas de vehículos y qué rutas de vehículos asignar a cada vehículo. Adelman (2004) presenta una investigación en la cual, para

instancias de casos reales de PRAXAIR, compara la política de precio directo (price-directed), tratadas con un proceso de decisión semi-Markoviano con las políticas de partición fija y políticas de valoración.

Sin embargo, estos últimos métodos son prácticos únicamente si el espacio de estados es pequeño y si la función de costo óptimo esperado puede obtenerse eficientemente.

Es evidente que una de las motivaciones del origen de los modelos de flujo con demanda desconocida fue proporcionar herramientas concretas, capaces de vencer las limitaciones y las desventajas de modelos basados en una descripción estocástica de esa demanda. Según los trabajos revisados, es importante mencionar que para aplicar cualquier modelo estocástico a una situación real, un análisis preliminar de la demanda es generalmente requerido, el cual debe estar basado en al menos cuatro elementos: 1. Una gran y confiable base de datos de registros históricos, 2. Apropriados supuestos acerca del comportamiento futuro de la demanda, 3. Apropriados supuestos acerca de los procesos estocásticos que podrían representar adecuadamente la demanda, y 4. La estimación de los parámetros de aquellos procesos estocásticos, usando herramientas estadísticas.

A pesar de todo es necesario tomar en cuenta que la aplicabilidad práctica de los modelos de demanda desconocida se encuentra aún muy limitada, debido a la alta complejidad computacional de sus modelos. Aun cuando también se ha hecho uso de modelos de Programación Dinámica, requieren de excesivos recursos computacionales.

2.3 Análisis de los métodos y estrategias de solución del IRP.

Una vez terminada la revisión de los trabajos reportados en las fuentes de información y desarrollada la clasificación de métodos y estrategias de solución, se hizo un análisis de ellos.

Como se mencionó, la clasificación en dominio de tiempo y dominio de frecuencia está dada por las necesidades del decidor y el contexto del problema.

En la tabla 4 puede observarse el comportamiento de cada dominio de decisión para los factores relevantes: i) horizonte de planeación, ii) complejidad de los modelos y métodos y iii) representación de la realidad.

En este punto es conveniente resaltar que el método de solución propuesto en el presente trabajo es bajo dominio de tiempo, toma decisiones al inicio de cada día y además, presenta sencillez en los cálculos gracias a la estrategia de descomposición utilizada.

	<i>Dominio de frecuencia</i>	<i>Dominio de tiempo</i>
Horizonte de planeación	<ul style="list-style-type: none"> ⊕ Permite la planeación adecuada sólo para periodos cortos. 	<ul style="list-style-type: none"> ⊕ Es un método de loop cerrado en el que existe una retroalimentación de las decisiones tomadas en un periodo, que afectan las decisiones del siguiente, permitiendo así la corrección de errores periodo a periodo. ⊕ Permite la programación de operaciones a través de un horizonte posiblemente infinito.
Complejidad de los métodos	<ul style="list-style-type: none"> ⊕ Utiliza modelos complejos que frecuentemente deben ser sobresimplificados perdiendo precisión en los resultados. 	<ul style="list-style-type: none"> ⊕ Se pueden utilizar diversas técnicas basadas en VRP, que es un área ampliamente desarrollada e investigada.
Modelación de la realidad	<ul style="list-style-type: none"> ⊕ Los métodos basados en EOQ tienen la principal desventaja de que proporcionan intervalos de entrega que pueden ser no enteros. ⊕ Generalmente se aplica a situaciones en las que las condiciones no tengan cambios importantes a través del tiempo. 	<ul style="list-style-type: none"> ⊕ Un caso interesante y ventajoso en cuanto a representación de la realidad y precisión ocurre, cuando las decisiones se toman al inicio de cada intervalo de tiempo, conociendo el estado del sistema. ⊕ Es idóneo para situaciones en las cuales las condiciones pueden variar de manera significativa a través del tiempo.

Fuente: Elaboración propia.

Tabla 4 Dominio de frecuencia vs. Dominio de tiempo.

Finalmente, mediante el análisis de los métodos pertenecientes a cada dominio de decisión se construyeron las tablas 5 y 6 que muestra un resumen de sus ventajas y desventajas.

Método	Ventajas	Desventajas	Características especiales
Método de Agregación	<ul style="list-style-type: none"> ☆ Obtienen modelos fáciles de resolver. ☆ Presentan soluciones razonables con la menor información posible ☆ Trabajan mejor en situaciones donde métodos más detallados son inconvenientes por la gran cantidad de datos y recursos requeridos. 	<ul style="list-style-type: none"> ✗ Obtienen modelos muy generales. ✗ Es más un arte que un método completamente formalizado y sólo es muy ventajoso en casos particulares. 	<ul style="list-style-type: none"> <input type="checkbox"/> Generalmente se aplican a muchos orígenes muchos destinos. <input type="checkbox"/> No permiten la división de carga entre vehículos. <input type="checkbox"/> Casos en que las demandas varían lentamente en el tiempo y los clientes se encuentran geográficamente dispersos.
Políticas de Partición Fija	<ul style="list-style-type: none"> ☆ Puede manejar problemas muy grandes (miles de clientes). ☆ Presenta flexibilidad en el uso de vehículos capacitados, vehículos no idénticos, restricciones de volumen de ruta y opciones de pedidos pendientes. 	<ul style="list-style-type: none"> ✗ A largo plazo las particiones resultan asintóticamente óptimas. ✗ Da casi triviales y claramente malas soluciones para problemas prácticos de tamaño modesto. ✗ No se puede modelar la coordinación entre los tiempos de visita para las diferentes regiones y los resultados se consideran sólo como cotas superiores de los costos reales. ✗ Modela de manera demasiado general a los costos de ruteo. 	<ul style="list-style-type: none"> <input type="checkbox"/> Determinar un estrategia óptima de particiones \square es un problema NP-duro. <input type="checkbox"/> Permite la división de carga entre vehículos.
Métodos de Frecuencia Discreta	<ul style="list-style-type: none"> ☆ Muestran buen desempeño en eficiencia y efectividad para problemas de tamaño considerable. ☆ Tiene aplicabilidad práctica. 	<ul style="list-style-type: none"> ✗ Se basan casi completamente en heurísticas. ✗ Imposible modelar entregas coordinadas. 	<ul style="list-style-type: none"> <input type="checkbox"/> Pretende evitar frecuencias de entrega continuas.

Fuente: Elaboración propia.

Tabla 5 Ventajas y desventajas de los métodos bajo dominio de frecuencia.

Método	Ventajas	Desventajas	Características especiales
Técnicas de Descomposición	<ul style="list-style-type: none"> ☆ Está diseñado para manejar problemas muy grandes. ☆ Pueden aprovechar estructuras especiales de la matriz de coeficientes. 	<ul style="list-style-type: none"> ✎ Modela de manera lineal a los costos de ruteo. ✎ Al definir los subproblemas puede ser necesario elegir entre la calidad de la cota que se obtienen y el grado de dificultad del subproblema. 	<input type="checkbox"/> Es necesario un amplio conocimiento de la técnica.
Métodos para Flujo Dinámico con Demanda Desconocida	<ul style="list-style-type: none"> ☆ Modela situaciones más reales. 	<ul style="list-style-type: none"> ✎ Obtiene modelos muy complejos para problemas de tamaño realista. ✎ Su aplicabilidad práctica es limitada. ✎ Deben satisfacerse requerimientos de datos históricos y comportamiento estadístico originados por la naturaleza estocástica de la demanda. ✎ Suele requerir excesivos recursos computacionales. 	<input type="checkbox"/> Es una promisoría y abierta línea de investigación.

Fuente: Elaboración propia.

Tabla 6 Ventajas y desventajas de los métodos bajo dominio de tiempo.

Capítulo 3

Estrategia propuesta.

Debido al contexto logístico en el que se desarrolla el problema de Inventario-Distribución, su formulación generalmente implica un modelo de gran tamaño en términos del número de clientes involucrados. Adicionalmente el problema es NP-duro, y al combinar estas dos características, gran tamaño y complejidad, nos enfrentamos a un problema imposible de solucionar con métodos de optimización exactos. Sin embargo, dadas las ventajas que el IRP representa en el contexto logístico, diversos autores han desarrollado múltiples estrategias de solución.

Al analizar la amplia variedad de trabajos que existen relativos al IRP, se pueden especificar varios elementos de clasificación tales como topología de la red, tipo de decisión, restricciones y estructura de costos; éstos permiten tipificar y comprender el IRP al que nos enfrentamos, y en consecuencia, diseñar estrategias de solución adecuadas. La primera sección de este capítulo clasifica al problema tratado mediante dichos elementos.

Las siguientes cinco secciones del capítulo detallan la estrategia propuesta, que conformada por tres fases, fortalece los puntos en los cuales las estrategias desarrolladas hasta el momento, presentan debilidades. Se advierte que la mayor dificultad en la solución del IRP tratado, es el enfrentarse a la asignación de clientes en subgrupos, para ser visitados por los vehículos que les distribuyen producto. Generalmente, ese problema resulta ser entero mixto, con la inherente complejidad combinatoria para su solución. La estrategia que se propone en este trabajo ataca principalmente esa dificultad, por medio de una técnica de descomposición denominada Descomposición Cruzada Separable, que aplicada al problema de asignación mencionado, logra un algoritmo de solución polinomial cuyos resultados son óptimos y que además, es factible de ser implementada a bajo costo.

El séptimo y último apartado del capítulo, desarrolla un ejemplo de aplicación para la mejor comprensión de la estrategia.

3.1 Descripción del IRP tratado.

El IRP que se tratará en la presente investigación tiene las siguientes características:

Elementos de clasificación	Atributo	Valor
Topología	Puntos finales	Uno a muchos
Productos	Número	Uno
Demanda	Conocimiento Comportamiento en el tiempo Distribución	Conocida Constante Uniforme
Tipo de decisión	Dominio de decisión	Tiempo
Restricciones/Objetivo	Capacidad del vehículo Capacidad de inventario Capacidad de abastecimiento Número de vehículos	Sí Sí No Variable de dec.
Costos	Inventario Transportación	Sí Fijo + rel. con dist.

Fuente: Elaboración propia.

Tabla 7 Problema de Inventario-Distribución tratado en la presente investigación.

Las características mostradas son las que se han considerado más comunes en los sistemas de inventario-distribución de las micro, pequeñas y medianas empresas mexicanas.

En adelante, a quien hace las decisiones centrales se le llamará almacén y las localidades de inventario serán llamadas clientes. El almacén es responsable de mantener un nivel de inventario ideal para cada cliente, y decide qué clientes deberían ser abastecidos en el período de tiempo, que en este caso es de un día, y con qué cantidad de producto.

La definición general del problema se da a continuación.

Topología.-

Desde un almacén central hacia varios clientes se distribuye un solo producto. La posición geográfica del almacén y de los clientes es conocida, así como las distancias entre el almacén y los clientes y las distancias entre todos los clientes.

Producto.-

En este punto es importante considerar que el manejo de un solo producto responde a la "unitarización de la carga", que es la combinación de pequeños componentes o unidades de un cargamento en una unidad más grande. En este contexto, una caja de 24 refrescos, un pallet, un contenedor, una caja de FedEx, se consideran un paquete unitario, y por lo tanto un único producto. De manera generalizada, lo que las empresas buscan mover son "paquetes unitarios", para optimizar su esfuerzo de carga y descarga, transportación y almacenamiento.

Demanda.-

Se asume que la demanda del producto de cada cliente es conocida y constante en el tiempo, de hecho se conocen las necesidades en tiempo real, situación que permite una estrategia reactiva, basada en el conocimiento preciso y a tiempo de todos los inventarios.

Tipo de decisión.-

El dominio de decisión del problema es de tiempo, ya que se definen periodos de tiempo fijos de un día.

Restricciones/Objetivo.-

Se asume que la demanda diaria del cliente, no excede la capacidad del inventario. No se tiene limitación en la capacidad proveedora del almacén central, pero sí se limita la capacidad de los inventarios de los clientes. En el almacén central se dispone de una flotilla de vehículos que atenderá a los clientes en el período de tiempo definido. Se tienen limitaciones para la capacidad de los vehículos de la flotilla, y el tamaño de ella es una variable de decisión que estará involucrada en el proceso de optimización.

Sin embargo, para construir el modelo se requiere conocer el mínimo número de vehículos necesario para atender al conjunto de clientes. Esto puede hacerse resolviendo un problema de cajas de empaquetamiento o BPP (Bin Packing Problem), o utilizando la cota inferior trivial (Toth y Vigo, 2002) que está dada por:

$$\lceil d(S)/C \rceil$$

donde

$S = V \setminus \{0\}$: es el conjunto de los nodos de la red de distribución sin incluir al almacén (o nodo fuente), esto es, únicamente los clientes.

$d(S)$: es la demanda de todos los clientes

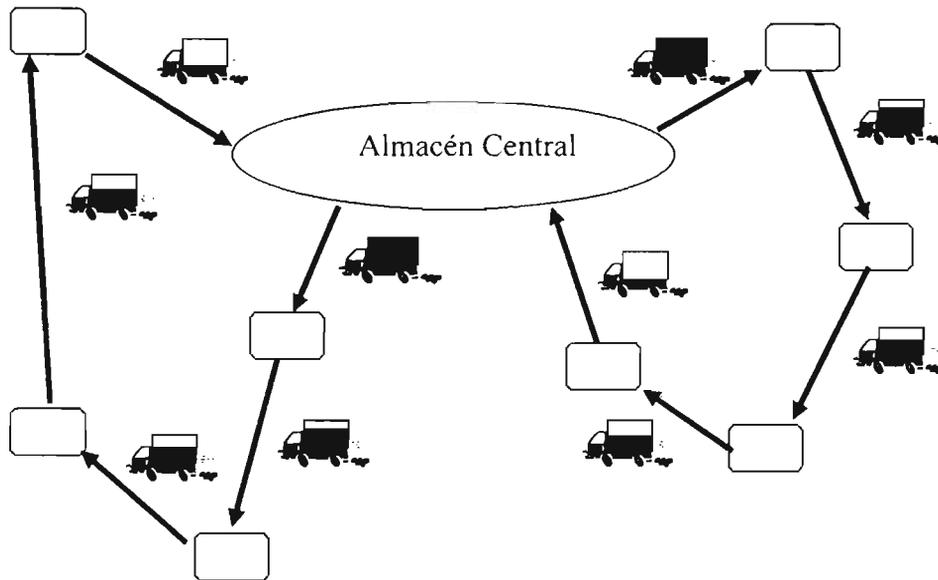
C : es la capacidad del vehículo, asumiendo que los vehículos tienen la misma capacidad. En caso de no ser así, puede utilizarse el promedio de capacidad de los vehículos.

El objetivo es satisfacer la demanda de los clientes al menor costo.

Costos.-

El costo que debe minimizarse se compone por costos de manejo de inventario de los clientes, así como costos de transportación que incluyen costos fijos más costos relacionados con la distancia recorrida por los vehículos utilizados.

En la Fig. 7 se muestra el esquema del problema.



Fuente: Elaboración propia.

Fig. 7 Esquema del IRP.

Así, el problema es encontrar las políticas de abastecimiento y distribución, que minimicen los costos relacionados con el inventario y la transportación. Las políticas mencionadas se definen a partir de la respuesta a las cuatro preguntas siguientes que deben contestarse al inicio de cada día:

1. **¿A qué clientes se debe atender en el día?**
2. **¿Qué vehículo va a atender a cada cliente?**
3. **¿Qué cantidad de producto entregar al cliente cuando sea visitado?**
4. **¿Cuál es la ruta de visitas de cada vehículo?**

3.2 Descripción de la estrategia propuesta.

Para tener una dirección inicial y adecuada al proponer una estrategia que fuese eficiente, se analizaron las estrategias de solución reportadas en literatura, que se utilizaron para resolver el IRP, esto fue hecho en el Capítulo 2.

Como puede verse, el uso de técnicas heurísticas es muy frecuente. Esto es evidente dada la compleja naturaleza del problema. Sin embargo, la utilización de técnicas de optimización en subprocesos específicos de los métodos de solución, fortalecen los resultados.

Una estrategia que es muy utilizada y que ha mostrado buenos resultados es la de "estrategia de dos fases". El problema es descompuesto en sus componentes naturales, aglomeración (agrupación) de los nodos (vértices) en rutas factibles y

construcción de las rutas (ruteo). La estrategia de dos fases se divide en dos clases:

- i) Agrupar I – Rutear II
- ii) Rutear I – Agrupar II

En Agrupar I – Rutear II, que es la más utilizada de las dos, los nodos son primero organizados en grupos factibles, y una ruta de vehículos es construida para cada uno de ellos. En ambas fases se pueden utilizar, tanto algoritmos heurísticos como exactos (Toth y Vigo, 2002). Fue precisamente la estrategia Agrupar I – Rutear II una de las ideas iniciales en el diseño de la presente propuesta.

Como se mencionó, el IRP es un problema que integra restricciones de abastecimiento de inventario al VRP en un solo modelo, las decisiones de ruteo dependen de la selección de clientes, y la selección de clientes a su vez depende de la cantidad de producto que los vehículos pueden transportar. Dicha interdependencia entre selección de clientes y ruteo de vehículos representa el punto clave del problema y con respecto a él se desarrollan diversas estrategias de solución.

Con base en toda la información obtenida en el Capítulo 2, así como el detallado estudio del problema, considerando las herramientas de optimización susceptibles de aplicarse, es como se llega a proponerse una estrategia de solución que logre buenos resultados de manera eficiente (Ghiani,2004). A continuación se describe de manera general la estrategia y más adelante se hará una descripción detallada.

Para la solución del IRP se diseñó una estrategia integrada por tres fases, que responde a las cuatro preguntas planteadas de la siguiente forma:

La **Fase I** contesta la pregunta: *1. ¿A qué clientes se debe atender en el día?*, por medio de la aplicación de un esquema de abastecimiento que utiliza un sistema de inventarios específico para cada cliente, y decide qué clientes, de todos los existentes, son los que requieren ser atendidos en ese día. Una vez contestada la pregunta se pasa a la siguiente fase.

La **Fase II** responde las preguntas: *2. ¿Qué vehículo va a atender a cada cliente?* y *3. ¿Cuánto entregar al cliente cuando sea visitado?* Estas preguntas se contestan aplicando una técnica de descomposición denominada Descomposición Cruzada Separable al problema de Asignación de Vehículos y Distribución o AVD, que al ser solucionado define la asignación de clientes a vehículos y la cantidad de producto que cada vehículo entregará a cada cliente. A partir de este resultado se pasa a la siguiente fase.

La **Fase III** responde a la pregunta: *4. ¿Cuál es la ruta de visitas de cada vehículo?*, la cuál es respondida al resolver un Problema de Agente Viajero o TSP (Traveling Salesman Problem), para cada uno de los vehículos utilizados.

El uso de la Descomposición Cruzada Separable de la Fase II, representa una de las aportaciones más importantes del presente trabajo de investigación. Las razones para considerar su uso responden a la estructura especial que tiene el problema que debe ser resuelto en esa fase. De tal forma es posible aprovechar los trabajos iniciados por Van Roy (1983), resolviendo en esa fase al problema AVD que utilizará a las variables binarias, como las decisiones combinatorias de asignación de vehículos a clientes. En la sección 3.4 se detalla el uso de esta técnica.

3.3 Problema de Asignación de Vehículos y Distribución (AVD).

Supongamos que la Fase I ha respondido ya, qué clientes deben ser atendidos en el día, entonces la Fase II debe responder las preguntas 2. ¿Qué cliente debe asignarse a cada vehículo? y 3. ¿Cuánto entregar a cada cliente? Estas preguntas pueden ser contestadas por medio de un modelo que tiene como base a uno de FL y en el que ahora, los servicios son los vehículos que conforman la flotilla que atenderá a los clientes y no se elabora producto si no que se distribuye. El nombre de este nuevo problema es Asignación de Vehículos y Distribución o AVD y al darle solución, se determinará qué vehículo atiende a cada cliente y qué cantidad de producto le debe abastecer.

Sea entonces el problema AVD en el cual un número finito m de vehículos disponibles, distribuyen un solo tipo de producto, para atender la demanda de una cierta población de usuarios, concentrada en n puntos discretos, cada uno con demanda d_j . Cuando un vehículo en particular es seleccionado, un costo fijo f_i incurre por la utilización del servicio y un costo variable $c_{ij}x_{ij}$ que está en función de la distancia recorrida es también incurrido por atender la fracción de la demanda x_{ij} , del cliente j . El costo $c_{ij} = x_{ij}^*$ el costo unitario de viaje del origen i al destino j .

Suponiendo que la capacidad de cada vehículo es limitada y uno o más de los vehículos pueden ser requeridos para satisfacer la demanda total, el problema consiste en decidir cuáles de los posibles vehículos serán utilizados, de tal manera que sus capacidades no sean excedidas y las demandas satisfechas; así como qué patrón de distribución deberá utilizar; tal que el costo total de establecer las unidades en servicio conformado por pagos al chofer, costos de mantenimiento, etc, más el costo total de atender la demanda, conformado por costos de transportación como una función de la demanda; sea minimizado en un horizonte finito de planeación. El problema AVD se formula como sigue:

$$\text{Minimizar} \quad \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} + \sum_{i=1}^m f_i y_i$$

sujeto a

$$\sum_{i=1}^m x_{ij} = 1, \quad \forall j \quad (3.1)$$

$$x_{ij} \leq y_i, \quad \forall i, j \quad (3.2)$$

$$\sum_{i=1}^m a_i y_i \geq \sum_{j=1}^n d_j \quad (3.3)$$

$$\sum_{j=1}^n d_j x_{ij} \leq a_i y_i, \quad \forall i \quad (3.4)$$

$$x_{ij} \geq 0, \quad y_i = 0,1, \quad \forall i, j.$$

donde

m : número de vehículos disponibles

n : número de clientes

d_j : demanda del cliente j

f_i : costo fijo para la ruta i

a_i : capacidad del vehículo i

c_{ij} : costo (en función de la distancia recorrida) de distribución al cliente j utilizando el vehículo i .

x_{ij} : fracción de la demanda total atendida del cliente j utilizando el vehículo i .

$y_i = 1$ si se usa el vehículo i , 0 en otro caso.

La restricción (3.1) asegura la atención total de la demanda, (3.2) establece la distribución sólo con vehículos activos, (3.3) considera el uso de suficientes vehículos para atender la demanda y (3.4) considera no exceder la capacidad del vehículo.

Este tipo de problemas presenta varios desafíos para su optimización, ocasionados por su tamaño y por su estructura combinatoria. De hecho, los

problemas de naturaleza logística, tales como el AVD y el IRP, suelen ser de gran tamaño, considerando un gran número de variables y de restricciones. El AVD por sí solo es muy complicado, puesto que la sola decisión básica de utilizar o no un vehículo da a los problemas una compleja estructura combinatoria. Por ejemplo, si se tiene un problema con a lo más 30 vehículos, puede existir más de un millón de combinaciones potenciales para la elección de ellos. Cuando se trata con problemas que asocian esa estructura combinatoria con un gran tamaño, las complicaciones aumentan, por lo que la MP se apoya en técnicas clave para este tipo de desafíos, tales como las técnicas de acotamiento, la teoría de dualidad y las técnicas de descomposición.

El AVD tienen dos tipos de decisiones inherentes: elegir los vehículos que se utilizarán y la forma de distribuir mejor la demanda hacia los clientes. Esta característica los hace un atractivo campo para el uso de técnicas de descomposición, ya que si la decisión discreta de elegir el vehículo se ha tomado, el problema continuo de distribución generalmente es más fácil de resolver. Aun si no es posible aprovechar esta característica en el diseño de algoritmos de solución, el utilizar la descomposición puede seguir siendo muy atractivo, ya que si el AVD no fuese complicado por la decisión discreta de elección de vehículos y fuera formulado como un problema de LP (relajando las restricciones de integralidad de las variables del problema), aún podría ser muy grande y difícil de resolver. Afortunadamente, el AVD tiene una estructura especial que puede ser explotada por las técnicas de descomposición. (Aceves, 1996).

La revisión de los métodos de solución existentes y las consideraciones anteriores dieron como resultado la utilización de la Descomposición Cruzada Separable en la estrategia de solución de IRP.

Al resolver el problema AVD se obtendrá:

- i) Qué vehículo atiende a cada cliente, y
- ii) Cuál es la cantidad de producto que cada vehículo debe entregar a cada cliente.

Entonces, si ahora sabemos:

- ➔ qué clientes visitar en el día
- ➔ cuánto entregar a cada cliente y
- ➔ qué cliente debe asignarse a cada vehículo

Lo que resta es saber las secuencias de cada vehículo al visitar a los clientes que se le han asignado, partiendo del almacén central y regresando al mismo. Es evidente que este es un TSP que, pudiendo ser de gran tamaño, será resuelto usando técnicas heurísticas.

Resolviendo este último problema se soluciona el IRP planteado, ya que se contesta a las cuatro preguntas fundamentales que resuelven el problema, posibilitando la definición de las políticas de manejo de inventarios y distribución, que minimizan los costos totales asociados.

La estrategia mencionada, que consta de tres etapas, puede verse en la Fig. 8.

A continuación se describirán detalladamente las tres fases.



Fuente: Elaboración propia.

Fig. 8 Fases de la estrategia de solución propuesta.

3.4 Fase I: Aplicación de un esquema de abastecimiento.

Consideremos entonces el supuesto del VMI y la utilización de tecnología para conocer el estado del inventario de los clientes de manera precisa y en tiempo real.

En la primera fase de la estrategia se debe aplicar un esquema de abastecimiento de inventarios para seleccionar un subconjunto de clientes, formado por aquellos que serán visitados por los vehículos en el período de tiempo fijado (un día). El criterio de decisión acerca de si un cliente debe ser visitado o no en el día, depende del sistema de inventario particular de cada cliente, pero que es manejado por el proveedor. Ahora bien, un sistema de inventario es el conjunto de políticas y controles que monitorean el comportamiento de los inventarios y determinan qué niveles deben ser mantenidos, cuándo deben reabastecerse las existencias, y qué tan grandes deberían ser las órdenes de reabastecimiento.

Si bien algunos de los propósitos de los inventarios son mantener independencia de operaciones, enfrentar variación en la demanda del producto, permitir flexibilidad en la programación de la producción; el mantener inventarios conlleva también costos asociados, tales como costos de mantenimiento por almacenamiento, manipulación, robo, daños, depreciación, etc.; y costos de abastecimiento, como son: conteo de existencias, cálculo de requerimientos, transportación, etc.

La definición de políticas de inventario óptimas implica, mediar entre el hecho de que muy poca reserva puede ocasionar costosas situaciones de desabasto de producto y demasiada reserva puede arruinar la ventaja competitiva y el margen de ganancias del negocio. Considerando lo anterior la mejor forma de manejar los inventarios es minimizar su impacto adverso, encontrando el "justo medio" entre los dos casos extremos.

El esquema que se utilizará será el Sistema de Reabastecimiento Opcional u ORS (Optional Replenishment System). El ORS es una variación del Modelo de Período de Reorden de Tiempo Fijo (también llamado Modelo P o Modelo de Revisión Periódico). Este modelo es adecuado para el caso del IRP tratado, ya que como se ha indicado, el horizonte de planeación es de un día, *i. e.*, es un periodo de tiempo fijo. Además, este sistema se considera adecuado, ya que se cumple con los supuestos para su uso y se ajusta a las necesidades y características del IRP planteado. Aunado a esto, no es un modelo complejo que da buenos resultados y es fácil de implementar.

Para comprender mejor lo anterior, a continuación se describe el funcionamiento del modelo de inventario P básico.

3.4.1 Sistema de Inventarios de Período de Tiempo Fijo.

El modelo de inventarios de período de tiempo fijo se caracteriza porque es activado por el tiempo. Está limitado a hacer reabasto hasta el final de un período de tiempo predeterminado y únicamente el paso del tiempo activa al modelo.

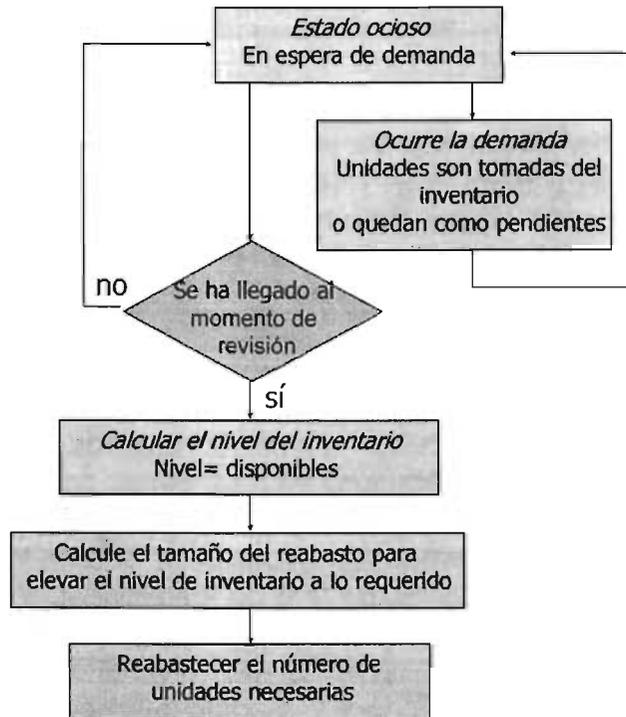
- i) el reabastecimiento se realiza hasta que se cumple el tiempo de revisión T ,
- ii) tiene un tamaño de inventario promedio grande ya que éste debe proteger también contra desabasto durante el período de revisión T ,
- iii) la cantidad q que debe reabastecerse, varía cada vez que se hace el reabasto,
- iv) el conteo se hace sólo en el momento de revisión (generalmente al inicio del período).

La Fig. 9, muestra el diagrama de flujo del proceso del Modelo P.

Como puede observarse, una decisión de reabasto se hace después de que el nivel de existencias se ha contado o revisado. La realización de un reabasto depende de la posición del inventario en el tiempo.

El modelo P es utilizado en situaciones en las que el proveedor quiere combinar entregas para ahorrar en costos de transportación y/o en las que los vendedores hacen visitas de rutina a sus clientes (Chase *et al.* 2001), ambas situaciones están presentes en la operación del IRP.

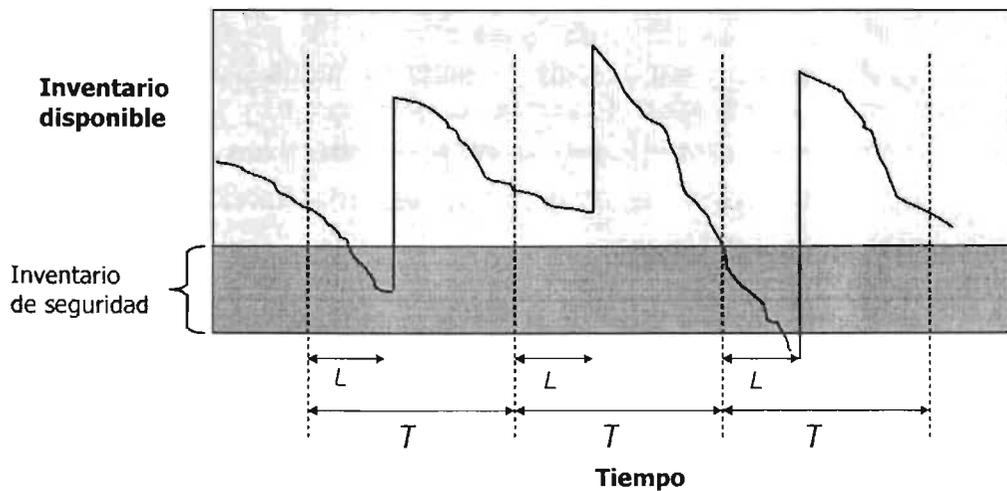
Modelo P
Sistema de Período de Reabasto
de Tiempo Fijo



Fuente: Chase et al. (2001).

Fig. 9 Diagrama del proceso del Modelo P básico.

En la Fig. 10 se puede ver la gráfica de un Sistema de Período de Tiempo Fijo P con un ciclo de revisión de T y un tiempo de reabasto constante L . En este caso las demandas se distribuyen aleatoriamente con una media \bar{d} .



Fuente: Chase et al. (2001).

Fig. 10 Gráfica del Modelo P.

Los modelos P generan tamaños de reabasto que varían de período a período dependiendo de la tasas de uso. Esto requiere generalmente mantener un nivel de existencias de seguridad. Como se ve en la Fig. 10, es posible que alguna gran demanda haga que las existencias descendan hasta cero, inmediatamente después de realizar un reabasto.

Estas condiciones podrían no conocerse hasta que llegue el siguiente período de revisión. De tal forma es posible tener desabasto por casi todo el período. Estas situaciones hacen necesario contar con un nivel de existencias de seguridad que sea calculado cuidadosamente y realmente proteja contra el desabasto del inventario.

Para el IRP tratado, es necesaria la decisión de abastecer o no a un cliente en el día, por tanto se tiene un período de tiempo fijo. Por otro lado, como indica Chase *et al.* (2001), obtener en la práctica los costos actualizados de ordenar, preparar las órdenes, transporte y almacenamiento es muy difícil, de hecho en ocasiones es imposible y los supuestos suelen ser muy poco realistas. Además, el mismo autor indica que todos los sistemas de inventarios tienen las siguientes dos problemáticas: control adecuado de cada uno de los productos del inventario y la seguridad del registro preciso del inventario disponible (a la mano). Por lo tanto, se recomienda el uso de sistemas simples de inventarios. Considerando lo anteriormente mencionado se decidió que el modelo utilizado en esta Fase II sería el modelo ORS, una variación del modelo básico P, que se adecua a las necesidades de la estrategia.

3.4.2 Sistema de Reabastecimiento Opcional (ORS).

Un ORS fuerza la revisión del nivel de inventario en una frecuencia de tiempo fija y ordena un reabastecimiento, si el nivel ha caído por debajo de alguna cantidad determinada. De tal forma, éste es básicamente un modelo P. Un nivel de demanda D se pronostica para ese día, a partir de datos históricos de demanda diaria. Debido a que la actividad de abasto cuesta tiempo y recursos monetarios, un tamaño mínimo de orden Q puede establecerse. Una buena opción para calcular Q consiste en aplicar la ecuación de tamaño económico de lote EOQ (Economic Order Quantity), que es:

$$Q = \sqrt{\frac{2DS}{H}}$$

donde

D = demanda

S = costo por reabastecer

H = costo de mantenimiento y almacenamiento por unidad del inventario promedio

Entonces, cuando el producto se revisa, la posición del inventario I , se resta del nivel de reabasto requerido D . Estableciéndolo formalmente,

$$q = D - I$$

donde

D : nivel de demanda pronosticada

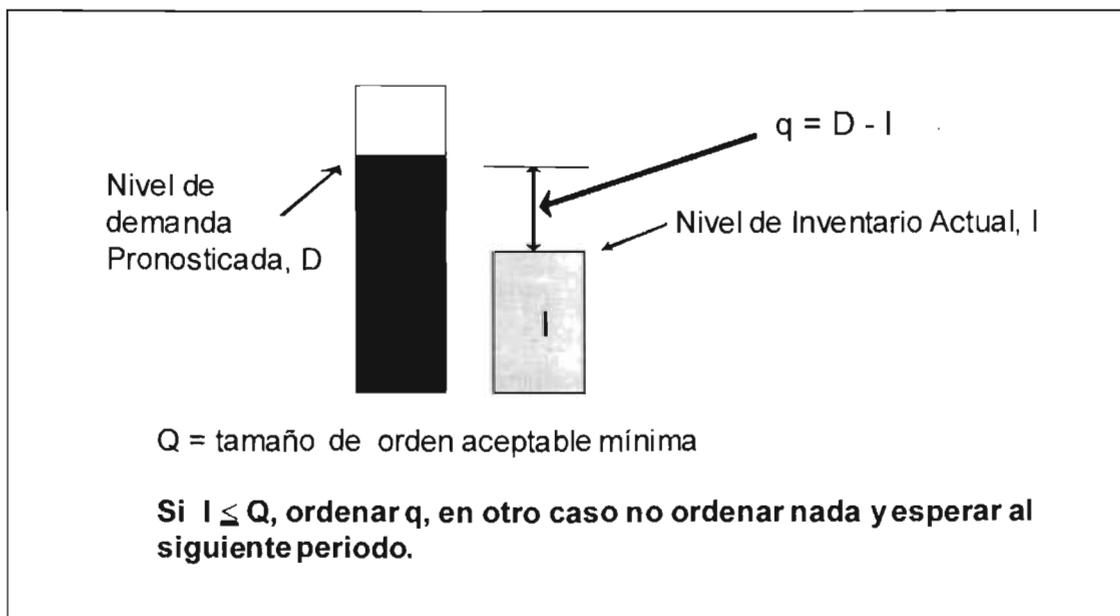
I : nivel de inventario actual

q : cantidad faltante de inventario para alcanzar el nivel de demanda pronosticada

Q : cantidad de orden aceptable mínima

y la regla es que sólo si $I \leq Q$, se ordena q ; si $I > Q$ o $q = 0$ no se atiende al cliente.

La Fig. 11 muestra el proceso del ORS.



Fuente: Elaboración propia con base en Chase et al. (2001).

Fig. 11 Proceso del modelo ORS.

Utilizando un ORS para cada cliente, dependiendo de sus circunstancias específicas, el proveedor decide si se abastece a un cliente en el día o se debe esperar la siguiente revisión. Así, es posible obtener el subconjunto de clientes que se visitarán en ese día.

La información necesaria en esta fase son los D, l, q y Q particulares para cada cliente, dependiendo de su demanda, capacidad, niveles de inventario, costos de reabastecimiento, de desabasto y de almacenamiento.

Esta fase contesta la pregunta 1. De tal manera, que ahora se sabe a qué clientes visitar en el día y por lo tanto se pasa a la Fase II.

3.5 Fase II : Asignación de vehículos a clientes y determinación del tamaño de las entregas.

Retomando el modelo de AVD planteado en la sección 3.3:

$$\begin{aligned}
 & \text{Minimizar} && \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} + \sum_{i=1}^m f_i y_i \\
 & \text{sujeto a} && \\
 & && \sum_{i=1}^m x_{ij} = 1, \quad \forall j \\
 & && x_{ij} \leq y_i, \quad \forall i, j \\
 & && \sum_{i=1}^m a_i y_i \geq \sum_{j=1}^n d_j \\
 & && \sum_{j=1}^n d_j x_{ij} \leq a_i y_i, \quad \forall i \\
 & && x_{ij} \geq 0, \quad y_i = 0,1, \quad \forall i, j.
 \end{aligned}$$

Recordemos que la solución de este problema definirá qué vehículo atiende a cada cliente y cuál es la cantidad de producto que se le entrega al ser visitado.

Como se mencionó antes, este problema es por sí solo muy complicado debido a su naturaleza y cuando su estructura combinatoria se une con un gran tamaño, el problema se hace excesivamente difícil de resolver, razón por la cual la MP se ha apoyado en técnicas especiales, tales como las de Descomposición.

También se indicó que el AVD, tiene dos tipos de decisiones inherentes: elegir los vehículos que se utilizarán y la mejor forma de distribuir la demanda hacia los vehículos. Esta peculiaridad hace a este tipo de problema un atractivo campo para el uso de técnicas de descomposición, ya que si la decisión discreta de elegir el vehículo se ha realizado, el problema continuo de distribución generalmente es más fácil de resolver. Un hecho ventajoso y parte esencial de este trabajo es que el problema AVD tiene una estructura especial que puede ser explotada por las técnicas de descomposición.

Las técnicas llamadas de descomposición fueron descritas de forma general en la sección 2.2.1. Como se mostró en esa sección, muchos problemas de MIP se basan en descomposición primal o en descomposición dual, lo cual da respectivamente, un algoritmo de descomposición de Benders y un algoritmo de enumeración implícita con cotas calculadas vía relajación Lagrangeana estrechamente relacionada con el principio de descomposición de Dantzig-Wolfe. Estos métodos explotan la estructura primal o dual del problema (ver el Apéndice B).

La descomposición primal o descomposición de Benders fue introducida poco después de que se desarrolló el método de descomposición de Dantzig-Wolfe. El subproblema de Benders es una restricción del problema original donde algunas de las variables primales tienen valores fijos. En cada iteración de la descomposición de Benders, estos valores son ajustados por el problema maestro. El método fue propuesto para diseño de redes de distribución, programación de trenes, ruteo de aviones, ruteo de vehículos, etc.

La descomposición dual para LP es ampliamente conocida como la descomposición de Dantzig-Wolfe. El algoritmo de descomposición dual resuelve sucesivamente un "subproblema dual" y un problema maestro hasta que se alcanza el óptimo y se verifica. Geoffrion en 1972, mostró que el método se puede generalizar vía relajación Lagrangeana para tratar con MIP. Así, se le llama descomposición dual o método de relajación Lagrangeana.

El subproblema dual se obtiene tomando la relajación lagrangeana del problema original relativo a algunas restricciones. Cada iteración consiste de:

1. el problema maestro dual selecciona un nuevo conjunto de multiplicadores de Lagrange,
2. se resuelve el subproblema dual para los valores dados de los multiplicadores.

El algoritmo se vuelve computacionalmente atractivo si las restricciones "complicadas" se relajan, de tal forma que se obtenga un subproblema relativamente fácil de resolver. El algoritmo de descomposición dual resuelve efectivamente el dual Lagrangeano formal, relativo a los subconjuntos de restricciones dados. Así que, cuando se aplica a MIP, una brecha de dualidad puede surgir y usualmente es cerrada por un esquema de enumeración del tipo Bifurcación y Acotación. Cotas inferiores en el valor del problema (minimización) se obtienen en cada iteración del algoritmo de descomposición dual, de tal forma, no se requiere resolver el dual Lagrangeano formal, óptimamente. La técnica fue propuesta para TSP, scheduling, redes multiproducto, localización de servicios, tamaño de lote, partición de conjuntos, pareamiento, etc.

El Apéndice B muestra más detalles de la Descomposición de Benders y de la Descomposición dual Lagrangeana.

El principio de descomposición en su forma primal o dual ha generado algoritmos muy eficientes para muchos problemas de MIP. La técnica permite tomar ventaja de la estructura especial del problema resolviendo una secuencia de subproblemas "sencillos". A partir de ello es que se elige explotar una u otra estructura, la dual o la primal. Sin embargo muchos problemas tienen ambos subproblemas fáciles de resolver. En particular, el problema de FL se reduce a un problema de transporte cuando las variables binarias de localización se fijan en cero o uno; y por otro lado, el subproblema dual por relajación de las restricciones de requerimientos de los clientes, reduce a un número de problemas mochila continuos, uno por cada servicio (Geoffrion, 1974), como lo muestra la Fig. 12.

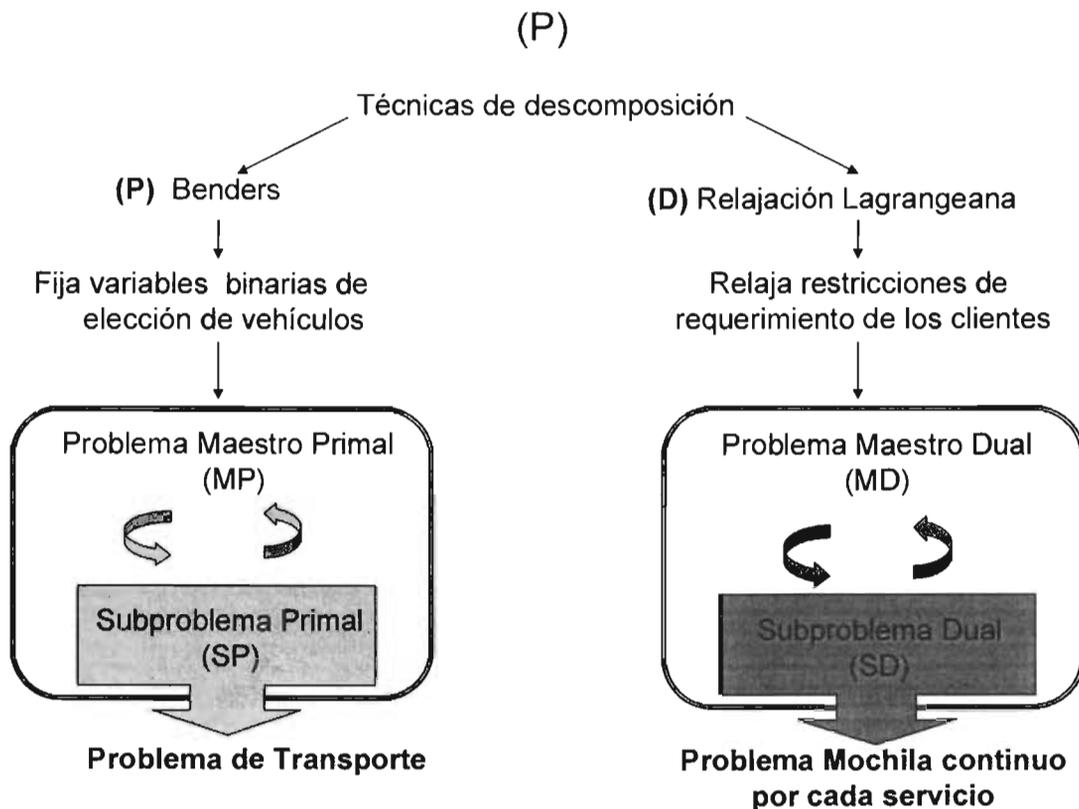


Fig. 12 Técnicas de descomposición para el problema de Localización de Servicios (FL).

3.5.1 Descomposición Cruzada y Descomposición Cruzada Separable.

A partir de lo anterior, Van Roy (1983) propone un nuevo método de descomposición llamado Descomposición Cruzada, que explota el hecho de que para algunos casos ambos subproblemas son sencillos de resolver, aprovechando además las estrechas relaciones entre la descomposición primal y dual.

Es bien conocido que ambos métodos son pares duales, *i.e.*, si el algoritmo de Benders es aplicado a un problema de LP, coincide con el algoritmo de descomposición de Dantzig-Wolfe que es aplicado al dual de ese mismo problema de LP (Lasdon, 1970).

El principio de Descomposición Cruzada se basa, precisamente, en la relación entre ambos principios de descomposición, Benders y Dantzig-Wolfe. De hecho, es posible establecer que el subproblema dual Lagrangeano es un problema maestro relajado en la descomposición de Benders y paralelamente, el subproblema de Benders puede ser considerado como un problema maestro relajado para la descomposición dual. De tal forma, que la idea básica de la Descomposición Cruzada es usar los subproblemas de ambos procedimientos en un solo procedimiento de descomposición. El algoritmo de Descomposición Cruzada para resolver MIP, ha sido desarrollado ampliamente por Van Roy (1982). Cabe mencionar que dicho algoritmo, aplicado a un conjunto de problemas prueba estándar, dio una reducción promedio de tiempo de solución en CPU de 20% con respecto a los mejores algoritmos conocidos hasta ese momento.

El procedimiento general se muestra a continuación:

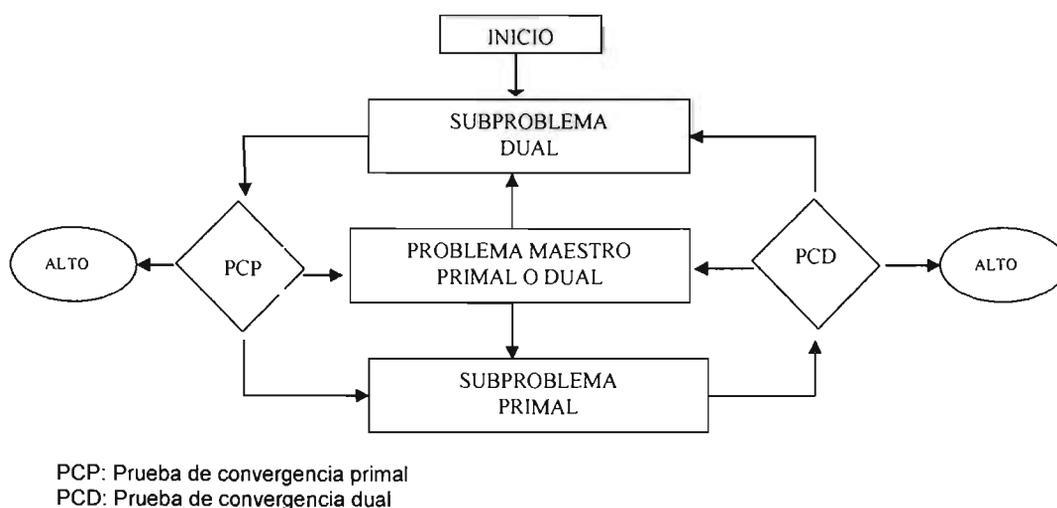


Fig. 13 Procedimiento general de la Descomposición Cruzada.

Las pruebas de convergencia PCP y PCD, controlan la secuencia entre los subproblemas primal y dual y los problemas maestros. Se ve cómo, en ocasiones, es posible pasar de un subproblema primal a un subproblema dual sin pasar por algún problema maestro que generalmente es muy complejo.

El siguiente concepto de equivalencia muestra lo anterior:

Definición 3.1. Un problema Q es equivalente al problema Q' con respecto a un subconjunto Z (primal o dual) de variables, si la solución óptima de Q para Z , es óptima y viceversa.

Ejemplos del concepto de equivalencia anterior pueden ser, el problema (P) y el problema maestro (MP) en la descomposición primal, con respecto a la variable x_2 (por construcción). Similarmente (D) y (MD) son equivalentes con respecto a la variable u_2 (ver Apéndice B).

Así, se tienen las siguientes proposiciones:

Proposición 3.1. El problema maestro primal (MP_{u_2}) es equivalente al problema (SD_{u_2}) con respecto a x_2 y los valores de las funciones objetivo son iguales $v(MP_{u_2}) = v(SD_{u_2})$.

Prueba. El problema (MP_{u_2}) se puede escribir como:

$$\begin{aligned} & \text{Min}_{x_2 \in Z, x_0} \quad x_0 \\ & \text{s.a.} : \\ & \quad u' b + (C^2 - u' A^2) x_2 \leq x_0, \quad t \in T_{u_2} \end{aligned}$$

igualmente como

$$\text{Min}_{x_2 \in Z} \quad \text{Sup} \left\{ u' b + (C^2 - u' A^2) x_2, t \in T_{u_2} \right\}$$

o también

$$\text{Min}_{x_2 \in Z} \left\{ \begin{array}{l} \text{Max}_{u_1 \geq 0} \quad u' b + (C^2 - u A^2) x_2, \\ \text{s.a.} : \\ \quad u_1 A_1^1 \leq C^1 - u_2 A_2^1 \end{array} \right\},$$

puesto que la solución óptima del problema de maximización interior, es una solución básica que corresponde a un punto extremo del poliedro y además se tienen fijas x_2 y u_2 , el dual del problema interior queda como

$$\text{Min}_{x_2 \in Z} \left\{ \begin{array}{l} \text{Min}_{x_1 \geq Z} \quad (C^1 - u_2 A_2^1) x_1 + u_2 b_2 + (C^2 - u_2 A_2^2) x_2 \\ \text{s.a.} : \\ \quad A_1^1 x_1 \leq b_1 - A_1^2 x_2 \end{array} \right\}$$

que resulta ser (SDu_2) con x_2 fija,

$$\text{Min } Cx + u_2(b_2 - A_2x)$$

s.a :

$$A_1^1 x_1 \leq b_1 - A_1^2 x_2$$

$$x_1 \geq 0$$

Termina la prueba.

Paralelamente a la proposición anterior, se tiene el uso del subproblema de Benders (SPx_2) como un problema maestro para la descomposición dual. Para lo cual consideramos la siguiente proposición:

Proposición 3.2. El problema maestro dual (MDx_2) es equivalente al subproblema primal (SPx_2) con respecto a u_2 y los valores de las funciones objetivo son iguales $v(MDx_2) = v(SPx_2)$.

Prueba. El problema (MDx_2) puede escribirse de la siguiente forma

$$\text{Max}_{u_2 \geq 0, u_0} u_0$$

s. a :

$$Cx^t + u_2(b_2 - A_2x^t) \geq u_0, \quad t \in T_{x_2}$$

también como

$$\text{Max}_{u_2 \geq 0} \inf \{ Cx^t + u_2(b_2 - A_2x^t), \quad t \in T_{x_2} \}$$

igualmente

$$\text{Max}_{u_2 \geq 0} \left\{ \begin{array}{l} \text{Min}_{x_1 \geq 0} \quad Cx + u_2(b_2 - A_2x) \\ \text{s.a :} \\ A_1^1 x_1 \geq b_1 - A_1^2 x_2 \end{array} \right\}$$

ya que la solución del problema de minimización interior es un punto extremo de la región factible que se obtiene de la solución del dual del problema interior con u_2 y x_2 fijos, es posible escribir a (MDx_2) como

$$\begin{aligned} \underset{u \geq 0}{\text{Max}} \quad & C^2 x_2 + u_2 b_2 - u_2 A_2^2 x_2 + u_1 (b_1 - A_1^2 x_2) \\ \text{s.a:} \quad & u_1 A_1^1 \leq C^1 - u_2 A_2^1 \end{aligned}$$

o también

$$\begin{aligned} \underset{u \geq 0}{\text{Max}} \quad & C^2 x_2 + u b - u A^2 x_2 \\ \text{s.a:} \quad & u A^1 \leq C^1 \end{aligned}$$

que es el problema dual de

$$\begin{aligned} \underset{x_1 \geq 0}{\text{Min}} \quad & C^2 x_2 + C^1 x_1 \\ \text{s.a:} \quad & A^1 x_1 \geq b - A^2 x_2 \end{aligned}$$

que resulta ser (SP_{x_2}) . Termina la prueba.

Las proposiciones 3.1 y 3.2 implican que los subproblemas (SP_{x_2}) y (SD_{u_2}) son maestros uno del otro.

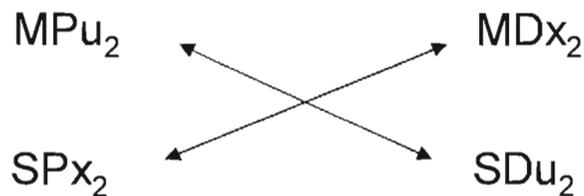


Fig. 14 Los subproblemas (SP_{x_2}) y (SD_{u_2}) son maestros uno del otro.

Investigaciones de Van Roy (1984), indican que cuando se aplica la estrategia de descomposición al problema de FL y en consecuencia al AVD, se produce una solución sucesiva de dos subproblemas, incorporando un proceso de ida y vuelta entre ellos, reduciendo el número de problemas maestros a resolver.

Tres hechos son básicos para el algoritmo de Descomposición Cruzada:

1. la relación que existe entre la descomposición primal (Benders) y la dual (vía relajación Lagrangeana);
2. los subproblemas $(SP x_2)$ y $(SD u_2)$ pueden ser considerados problemas maestros uno del otro;
3. consideraciones bajo las cuales (P) se puede resolver iterando únicamente entre ambos subproblemas.

Los primeros dos se han mostrado ya, las consideraciones del tercer hecho se muestran a continuación.

Para saber cuándo se alcanza el óptimo de (P) , obsérvese el *Lema 3.1*.

Lema 3.1.

- i) Sea u° solución dual óptima de $(SP x_2)$ y sea x^+ una solución óptima de $(SD u_2)$. Entonces $x_2^+ \neq x_2^\circ$ a menos que sus valores objetivo sean iguales, $v(SP x_2^\circ) = v(P)$.
- ii) Sea x° solución óptima de $(SD u_2)$ y u^+ solución dual óptima de $(SP x_2)$. Entonces $u_2^+ \neq u_2^\circ$, a menos que sus valores objetivo sean iguales, $v(SD u_2^\circ) = v(P)$.

Prueba.

- i) Supongamos que sucede lo opuesto, $v(SP x_2^\circ) > v(P)$ y que $x_2^+ = x_2^\circ$. Entonces como u° es solución dual óptima de $(SP x_2)$, u_2° será óptima en MDx_2° por la proposición (3.2) y el hecho de que, $Cx^+ + \underbrace{u_2^\circ(b_2 - A_2x^+)}_{\text{Dual de } SPx_2^\circ} \geq v(SP x_2^\circ) = v(MDx_2^\circ)$. Lo cual contradice que x^+ sea solución óptima de $(SD u_2^\circ)$ puesto que, $Cx^+ + u_2^\circ(b_2 - A_2x^+) = v(SD u_2^\circ) \leq v(P) \leq v(SP x_2^\circ)$.
- ii) Supongamos que sucede lo contrario $v(SD u_2^\circ) < v(P)$ y que $u_2^+ = u_2^\circ$. Entonces como x° es una solución óptima de $(SD u_2^\circ)$ y por la proposición (3.1) también será óptima en (MDx_2°) así que,

$u^+ b + \underbrace{x_2^\circ (C^2 - u^+ A^2)}_{\text{Dual de } SDu_2^\circ} \leq v(SDu_2^\circ) = v(MPu_2^\circ)$. Lo cual contradice que

u^+ sea una solución óptima de (SPx_2) , entonces,
 $u^+ b + x_2^\circ (C^2 - u^+ A^2) = v(SPx_2^\circ) \geq v(p) \geq v(SDu_2^\circ)$.

Y la prueba termina.

Para garantizar que la estrategia de Descomposición Cruzada no se ciclará en la etapa de solución de los subproblemas, considérese el *Lema 3.2*, que da las condiciones necesarias para mejorar los valores de la solución actual, donde T_D, T_P son los conjuntos de cortes duales y primales, respectivamente y $v(\bullet)$ indica la solución del problema, además de que $\bar{v}_{(\bullet)}$ indica la solución actual.

Lema 3.2.

- i) Si $u_2 \geq 0$ satisface $v(SDu_2) > \bar{v}_D$ entonces,
 $Cx^t + u_2(b_2 - A_2x^t) \geq \bar{v}_D, t \in T_D$.
- ii) Si $x_2 \in Z$ satisface $(SPx_2) < \bar{v}_P$, entonces;
 $u^t b + x_2(C^2 - u^t A^2) < \bar{v}_P, t \in T_P$.

Prueba.

- i) Como x^t es factible en (SDu_2) para algún u_2 seleccionado, se tiene que $Cx^t + u_2^\circ(b_2 - A_2x^t) \geq v(SDu_2), t \in T_D$, y como $v(SDu_2) > \bar{v}_D$, la prueba termina ($\bar{v}_D =$ valor de la solución dual actual).
- ii) Como u^t es factible en el dual de (SPx_2) , para algún x_2 seleccionado, se tiene que $u^t b + x_2(C^2 - u^t A^2) \leq v(SPx_2), t \in T_P$ y como $v(SPx_2) \leq \bar{v}_P$ por suposición, entonces la prueba termina ($\bar{v}_P =$ valor de la solución primal actual).

Termina la prueba.

Finalmente, el hecho de que el procedimiento de Descomposición Cruzada es finito, se establece con la siguiente proposición:

Proposición 3.3. El algoritmo generado de la estrategia de Descomposición Cruzada, es finito y resuelve a (P) .

Prueba. La prueba se establece utilizando los *Lemas* (3.1), (3.2) y el hecho de que, los algoritmos de descomposición en su forma "pura" son finitos.

Ahora bien, la investigación desarrollada por Aceves (1996) incorpora la estrategia de relajación Lagrangeana Separable (ver Apéndice C) a este proceso cruzado, estableciendo que no es necesario utilizar el problema maestro en la solución, *i.e.*, es posible resolver a (P) iterando únicamente entre subproblemas, evitando por completo a los problemas maestros. A este procedimiento se le denominó Descomposición Cruzada Separable y es el que se aplica al problema AVD en esta fase.

3.5.2 Aplicación de la Descomposición Cruzada Separable al problema de Asignación de Vehículos y Distribución (AVD).

Partamos nuevamente del problema AVD anterior, en su forma primal (P) :

$$\text{Minimizar} \quad \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} + \sum_{i=1}^m f_i y_i$$

sujeto a

$$\sum_{i=1}^m x_{ij} = 1, \quad \forall j \quad (3.1)$$

$$x_{ij} \leq y_i, \quad \forall i, j \quad (3.2) \quad (P)$$

$$\sum_{i=1}^m a_i y_i \geq \sum_{j=1}^n d_j \quad (3.3)$$

$$\sum_{j=1}^n d_j x_{ij} \leq a_i y_i, \quad \forall i \quad (3.4)$$

$$x_{ij} \geq 0, \quad y_i = 0,1, \quad \forall i, j.$$

donde (3.1) asegura la atención total de la demanda, (3.2) establece el envío sólo con vehículos en uso, (3.3) considera el uso de suficientes vehículos para atender la demanda y (3.4) considera enviar solamente la capacidad del vehículo.

Para comenzar a aplicar la Descomposición Cruzada Separable se debe obtener a partir de (P) , el subproblema primal por Benders y el subproblema dual por relajación Lagrangeana Separable.

El Problema Primal o de Benders.

Aplicando el proceso de descomposición de Benders al problema anterior (P) , se fijan las variables primales y_i en los valores de 0 ó 1 y el subproblema que se obtiene es uno de transporte con variables acotadas. La ventaja de esto es que existen métodos de solución muy eficientes para este tipo de problema.

Así, el subproblema primal SP queda como:

$$\begin{aligned}
 & \text{Minimizar} && \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} + \sum_{i=1}^m f_i y_i \\
 & \text{sujeto a} && \\
 & && \sum_{i=1}^m x_{ij} = 1, \quad \forall j \quad (SP) \\
 & && \sum_{j=1}^n d_j x_{ij} \leq a_i y_i, \quad \forall i \\
 & && 0 \leq x_{ij} \leq y_i, \quad \forall i, j
 \end{aligned}$$

que es un problema de transporte.

El dual de SP se plantea como:

$$\begin{aligned}
 & \text{Maximizar} && \sum_j v_j + \sum_i (f_i - a_i u_i - \sum_j w_{ij}) y_i \\
 & \text{sujeto a} && \\
 & && v_j - u_i - w_{ij} \leq c_{ij}, \quad \forall i, j \quad (DSP) \\
 & && v_j, u_i \geq 0, \quad \forall i, j
 \end{aligned}$$

y el problema maestro de Benders o primal se expresa como:

$$\begin{aligned}
& \underset{y \in \{0,1\}, \rho}{\text{Minimizar}} \quad \rho \\
& \text{sujeto a} \\
& \sum_j v_j^t + \sum_i (f_i - a_i u_i^t - \sum_j w_{ij}^t) y_i \leq \rho \quad (MP) \\
& t \in T_p
\end{aligned}$$

donde T_p es el conjunto de índices de todas las soluciones duales factibles básicas del conjunto de restricciones del problema (DSP), es decir, $t \in T_p$ son los cortes primales o de Benders. Además, u_i^t, v_j^t para $t \in T_p$ determinan todos los puntos extremos de la región factible del dual del subproblema primal.

El problema (MP), que es el problema maestro primal, es un MIP, con una sola variable continua, ρ .

El Problema Dual Lagrangeano Separable.

Como se mencionó anteriormente el esquema de relajación Lagrangeana utilizado es la relajación Lagrangeana Separable, cuyas ventajas se citan en el Apéndice C. Entonces, el proceso de análisis con esa estrategia para el problema (P) planteado, se establece de la manera siguiente:

Propósito:

- i. Inducir la separación del problema en subproblemas independientes.
- ii. Capturar las diferentes características estructurales del problema.
- iii. Obtener cotas más fuertes que con los otros esquemas de relajación.

Por medio de:

- i. Identificar las partes del problema que pueden ser separadas.
- ii. Reemplazar las variables en cada parte por copias o sustituir nuevas expresiones.
- iii. Dualizar la copia de las variables o la expresión de sustitución.

Investigaciones previas obtuvieron un buen esquema de relajación Lagrangeana para (P), (Chen y Guignard, 1992), sin embargo, Aceves (1996) obtuvo en su investigación, un esquema de relajación mejorado y simplificado que hace innecesario partir del anterior.

De tal forma, el problema dual Lagrangeano Separable *RLS* (ver Apéndice C), se puede formular como la suma de los siguientes dos problemas (Aceves, 1996):

1. Un problema Q para la variable y_i donde se incluye al valor v_i que se obtiene por Linearización entera (Aceves, 1996), al separar en un problema para cada vehículo i , considerando a cada $\{y_i = 1\}$, para $i=1, \dots, m$, y valores conocidos de λ_i , con $i=1, \dots, m$, teniendo entonces que

$$\{y_i = 1\},$$

$$v_i = \underset{x}{\text{Minimizar}} \left\{ \lambda_i \sum_j d_j x'_{ij} \mid \sum_j d_j x_{ij} \leq a_i, 0 \leq x_{ij} \leq 1, \lambda_i \text{ no restringida} \right\},$$

que resulta ser un problema tipo mochila continuo, para cada i , y que por su estructura, la solución óptima para cada v_i con $i=1, \dots, m$ está dada por

$$v_i = \begin{cases} \lambda_i \text{ Min}\{a_i, D\} & \text{si } \lambda_i < 0 \\ 0 & \text{si } \lambda_i \geq 0 \end{cases}$$

donde $\sum_j d_j x_{ij} = a_i, \forall i$ y el parámetro $\sum_j d_j = D$, es útil en problemas con capacidad limitada. De tal forma en la construcción del problema Q , el problema v_i contiene ya a las restricciones (3.4) y (3.2) para $y_i = 1$ y se agrega la restricción (3.3). Entonces, el problema Q es

$$Q = \underset{y}{\text{Minimizar}} \sum_i (f_i + v_i) y_i$$

sujeto a

$$\sum_i a_i y_i \geq D, \quad \forall j$$

$$y_i = 0, 1, \quad \forall i$$

y se requeriría resolver un problema v_i con $i=1, \dots, m$, para cada vehículo y después minimizar la contribución total que produce cada uno de ellos.

El problema Q es también un problema tipo mochila 0-1 con $i=1, \dots, m$.

2. Un segundo problema S para la variable x_{ij} con las restricciones (3.1) y (3.2):

$$S = \underset{x}{\text{Minimizar}} \sum_i \sum_j (C_{ij} - d_j \lambda_i) x_{ij}$$

sujeto a

$$\sum_i x_{ij} = 1, \quad \forall j$$

$$0 \leq x_{ij} \leq y_i, \quad \forall i, j$$

$$\lambda_i \text{ no restringida}, y_i = 0, 1, \quad \forall i$$

El problema S es un problema del tipo de selección múltiple, que por su estructura se puede establecer como uno con $x_{ij} = 0$ ó 1 , $\forall i, j$.

Ambos problemas, Q y S , para variables $y_i = 0$ ó 1 , $\forall i$. La solución para RLS está dada por $v(RLS) = v(Q) + v(S)$, que puede escribirse como:

$$\text{Minimizar} \sum_i \sum_j C_{ij} x_{ij} + \sum_i f_i y_i + \sum_i \left(v_i y_i - \lambda_i \sum_j d_j x_{ij} \right)$$

sujeto a

$$\sum_i x_{ij} = 1, \quad \forall j \quad (RLS)$$

$$x_{ij} \leq y_i, \quad \forall i, j$$

$$\sum_i a_i y_i \geq D$$

$$x_{ij} \geq 0, y_i = 0, 1, \lambda_i \text{ no restringida}, \forall i, j$$

Un hecho importante es: dado que $v_i = a_i \lambda_i$, podemos escribir a RLS como:

$$\text{Minimizar } \sum_i \sum_j C_{ij} x_{ij} + \sum_i f_i y_i + \sum_i \lambda_i \left(a_i y_i - \sum_j d_j x_{ij} \right)$$

sujeto a

$$\sum_i x_{ij} = 1, \quad \forall j \quad (LD)$$

$$x_{ij} \leq y_i, \quad \forall i, j$$

$$\sum_i a_i y_i \geq D$$

$$x_{ij} \geq 0, \quad y_i = 0,1, \quad \lambda_i \text{ no restringida}, \quad \forall i, j$$

y resulta ser equivalente al problema (P), con la restricción (3.4) relajada. Lo que implicaría una relajación Lagrangeana tradicional, (LD), para la cual $(P) \geq (LD)$.

Si consideramos una solución factible para (LD) que además cumpla con la restricción (3.4), vamos a tener que el producto

$$\lambda_i \left(a_i y_i - \sum_j d_j x_{ij} \right) \leq 0$$

para $y_i = 1$ y $\lambda_i < 0$, con $i = 1, \dots, m$.

Así que, si queremos que se cumpla con la restricción (3.4) y se agrega de nuevo como restricción, el problema queda como

$$\text{Minimizar } \sum_i \sum_j C_{ij} x_{ij} + \sum_i f_i y_i + \sum_i \lambda_i \left(a_i y_i - \sum_j d_j x_{ij} \right)$$

sujeto a

$$\sum_i x_{ij} = 1, \quad \forall j$$

$$x_{ij} \leq y_i, \quad \forall i, j \quad (LD1)$$

$$\sum_i a_i y_i \geq \sum_j d_j$$

$$\sum_j d_j x_{ij} \leq a_i y_i, \quad \forall i$$

$$x_{ij} \geq 0, \quad y_i = 0,1, \quad \forall i, j$$

y dada la estructura de la función objetivo de (P) se va a cumplir nuevamente que $(P) \geq (LD1)$, para $y_i = 1$ y $\lambda_i < 0$, con $i = 1, \dots, m$.

Dado este resultado, es posible reforzar al problema S , ya que es ese el problema que trabaja con las variables x_{ij} , incorporando la restricción (3.4) y formularlo como

$$\begin{aligned}
 S = \underset{x}{\text{Minimizar}} \quad & \sum_i \sum_j (C_{ij} - d_j \lambda_i) x_{ij} \\
 \text{sujeto a} \quad & \\
 & \sum_i x_{ij} = 1, \quad \forall j \\
 & \sum_j d_j x_{ij} \leq a_i, \quad \forall i \\
 & x_{ij} \geq 0, \forall i, j; \quad \lambda_i \text{ no restringida}, \forall i
 \end{aligned}$$

que es nuevamente un problema de transporte para el que existen algoritmos de solución muy eficientes.

Por otra parte, tomando en cuenta que los problemas Q y S son función de λ_i para $i = 1, \dots, m$ y retomando el valor de $v_i = a_i \lambda_i$, el problema Q se puede escribir como

$$\begin{aligned}
 Q = \underset{y}{\text{Minimizar}} \quad & \sum_i (f_i + \lambda_i a_i) y_i \\
 \text{sujeto a} \quad & \\
 & \sum_i a_i y_i \geq D, \quad \forall j \\
 & y_i = 0, 1, \quad \lambda_i \text{ no restringida}, \forall i
 \end{aligned}$$

en el cual, para $\lambda_i = -\frac{f_i}{a_i}$ con $i = 1, \dots, m$ tendremos que

$$Q = \underset{y}{\text{Minimizar}} \sum_i \left[f_i - \left(\frac{f_i}{a_i} \right) a_i \right] y_i$$

sujeto a

$$\sum_i a_i y_i \geq D, \quad \forall j$$

$$y_i = 0, 1, \quad \forall i$$

y cuya solución, como puede observarse, es $Q = 0$. De tal forma que, cualquier valor de $y_i \neq 0, i = 1, \dots, m$, cumple para obtener el mínimo del problema dual Lagrangeano completo, *RLS*. De hecho, considerando la solución del problema Q que se obtuvo al aplicar el principio de Linearización entero,

$$v_i = \lambda_i \text{Min}\{a_i, D\} \text{ si } \lambda_i < 0$$

donde

$$\sum_j d_j x_{ij} = a_i, \quad \forall i,$$

podemos formular de manera más general a la función objetivo del problema Q , como sigue

$$Q = \underset{y}{\text{Minimizar}} \sum_i \left(f_i + \lambda_i \sum_j d_j x_{ij} \right) y_i$$

y establecer que para cualquier valor de lambda, $\lambda_i = -\frac{f_i}{\sum_j d_j x_{ij}}$, se tendrá nuevamente que $Q = 0$.

Como consecuencia de esto, el problema Q se formula de la siguiente forma

$$Q = \text{Minimizar} \sum_j \left[f_i - \left(\frac{f_i}{\sum_j x_{ij} d_j} \right) \sum_j x_{ij} d_j \right] y_i$$

sujeto a

$$\sum_j a_j y_j \geq D, \quad \forall j$$

$$x_{ij} \geq 0, y_i = 0,1, \quad \forall i, j$$

A partir de lo cual podemos establecer que los vehículos $y_i \neq 0$ para $i = 1, \dots, m$ (activos) que satisfarán la demanda, obtenidos a partir de la solución del problema de transporte S , minimizarán también al subproblema dual Lagrangeano completo RLS . Por lo tanto es suficiente con resolver al problema S para obtener la solución del subproblema dual Lagrangeano (SD), en este nuevo esquema de Descomposición Cruzada.

El algoritmo para resolver el problema de transporte minimiza costos al saturar sucesivamente las rutas de distribución, que son más económicas; así, para un determinado origen, t , que tenga costos de transporte más económicos, y para valores de $\lambda_t \leq 0$ se tendrá que en sucesivas iteraciones

$$\sum_j x_{ij}^{k-1} \leq \sum_j x_{ij}^k \leq \sum_j x_{ij}^{k+1} \leq \dots \text{ con } \lambda_t^{k-1} \leq \lambda_t^k \leq \lambda_t^{k+1} \leq \dots$$

y consecuentemente ocurrirá que para los costos de distribución,

$$(c_{ij} - \lambda_t)^{k-1} \geq (c_{ij} - \lambda_t)^k \geq (c_{ij} - \lambda_t)^{k+1} \geq \dots$$

esto significa que las rutas más baratas se harán aún más económicas. Así, los flujos $\sum_j x_{ij}^k$ para las rutas t con los costos más bajos, se incrementarán

sucesivamente hasta alcanzar su límite, el cual está definido por su demanda o por su capacidad. De tal forma ocurrirá que

$$\sum_j x_{ij}^{k-1} \leq \sum_j x_{ij}^k \leq \sum_j x_{ij}^{k+1} \leq \dots \leq a_i.$$

Contrariamente para las rutas r más caras, la $\sum_j x_{ij}^k$ irá disminuyendo

sucesivamente hasta llegar a cero y se harán más costosas en cada iteración.

Una situación interesante es que a los valores de λ_i^k para $i = 1, \dots, m$, se les puede interpretar como la memoria del algoritmo, ya que contienen información de la iteración precedente, indicando así, cuáles fueron los caminos de distribución más baratos, así como los servicios con el costo fijo más bajo.

De tal forma que para las trayectorias más económicas se tiene que

$$\left(\frac{f_i}{\sum_j x_{ij}^{k-1}} \right)^k \geq \left(\frac{f_i}{\sum_j x_{ij}^k} \right)^{k+1} \geq \left(\frac{f_i}{\sum_j x_{ij}^{k+1}} \right)^{k+2} \geq \dots$$

El algoritmo para resolver el problema de transporte, también requiere que $\sum_i a_i \geq \sum_j d_j$ para que exista solución factible y en la solución óptima requiere

que se cumpla que $\sum_i x_{ij} = d_j$. Entonces podemos establecer que los orígenes i

que se obtienen de la solución del problema de transporte S , para los cuales $\sum_j x_{ij}^* = d_j$, $i = 1, \dots, m$, serán las $y_i^* = 1$, $i = 1, \dots, m$ de la solución del problema

tipo mochila Q . Es decir, los orígenes i que satisfagan la demanda del problema de transporte, serán los vehículos utilizados en el subproblema dual Lagrangeano (SD_λ) y como consecuencia en el problema de localización (P) .

Consecuentemente debe mencionarse que un aspecto importante obtenido por Aceves (1996), es la posibilidad de resolver el problema AVD, únicamente dando solución a una serie de problemas tipo transporte, i.e., el problema S , con la siguiente estructura:

$$\begin{aligned} & \underset{x}{\text{Minimizar}} \quad \sum_i \sum_j (C_{ij} - \lambda_i) x_{ij} \\ & \text{sujeto a} \\ & \quad \sum_i x_{ij} \geq d_j, \quad \forall j \\ & \quad \sum_j x_{ij} \leq a_i, \quad \forall i \\ & \quad x_{ij} \geq 0, \quad \lambda_i < 0, \quad \forall i, j \end{aligned}$$

y de hecho es el subproblema (SD) que se resuelve en el algoritmo de Descomposición Cruzada Separable, con valores de λ_i , para $i = 1, \dots, m$ de

$$\lambda_i^0 = -\frac{f_i}{a_i}, \quad \forall i \text{ como valor inicial, y}$$

$$\lambda_i^k = -\frac{f_i}{\sum_j x_{ij}}, \quad \forall i \text{ para los valores subsecuentes.}$$

Este resultado ha permitido desarrollar un algoritmo más sencillo que los obtenidos hasta el momento (Aceves, 1996).

3.5.3 Algoritmo de Descomposición Cruzada Separable para el problema AVD.

Utilizando a los subproblemas primal SP_y y dual SD_λ , iterando entre sus soluciones al fijar los valores de las variables primales y_i o las duales λ_i e iniciando con alguna suposición para dichas variables es posible pasar de una a otra de las siguientes etapas:

- i) Fijar y_i es su valor actual y resolver el subproblema de Benders SP_y , que es un problema de transporte, para generar un nuevo valor para la cota superior de P .
- ii) Fijar λ_i en su valor actual y resolver el subproblema Lagrangeano SD_λ , para generar un nuevo valor para la cota inferior de P .

Como puede observarse, estas dos etapas producen de forma sucesiva cotas inferior y superior para el valor óptimo de la solución de (P).

A continuación se muestran las etapas del algoritmo:

Algoritmo de Descomposición Cruzada Separable.

1.- Iniciar.

$$v_D(-\infty), \quad v_P(+\infty); \quad Y_i^0 = 1, \text{ para } i = 1, \dots, m; \quad \lambda_i^0 = -\frac{f_i}{a_i}, \text{ para } i = 1, \dots, m.$$

2.- Resolver. Subproblema dual SD_{λ^k} (problema de transporte S), para obtener

$$y_i^1 = 1, \quad \sum_j x_{ij} \quad \text{y} \quad v(SD_{\lambda^k}).$$

ESTA TESIS NO SALE
DE LA BIBLIOTECA

2.1.- **Calcular.** $\lambda_i^k = -\frac{f_i}{\sum_j x_{ij}}$, para $i = 1, \dots, m$.

3.- **Probar.** Si $\lambda_i^{k-1} = \lambda_i^k$ para $y_i^k = 1$, entonces terminar. De otra forma, identificar cuáles $y_i^k = 1$ para $i = 1, \dots, m$.

4.- **Resolver.** Subproblema primal $SP_{y_i^k}$ (problema de transporte por Benders), para obtener $v(SP_{y_i^k})$.

5.- **Probar.** Si $v(SP_{y_i^k}) = v(SD_{\lambda^k})$, entonces terminar. De otra forma regresar a la etapa 2 pero ahora con λ_i^k .

El diagrama de flujo del algoritmo se muestra en la Fig. 15.

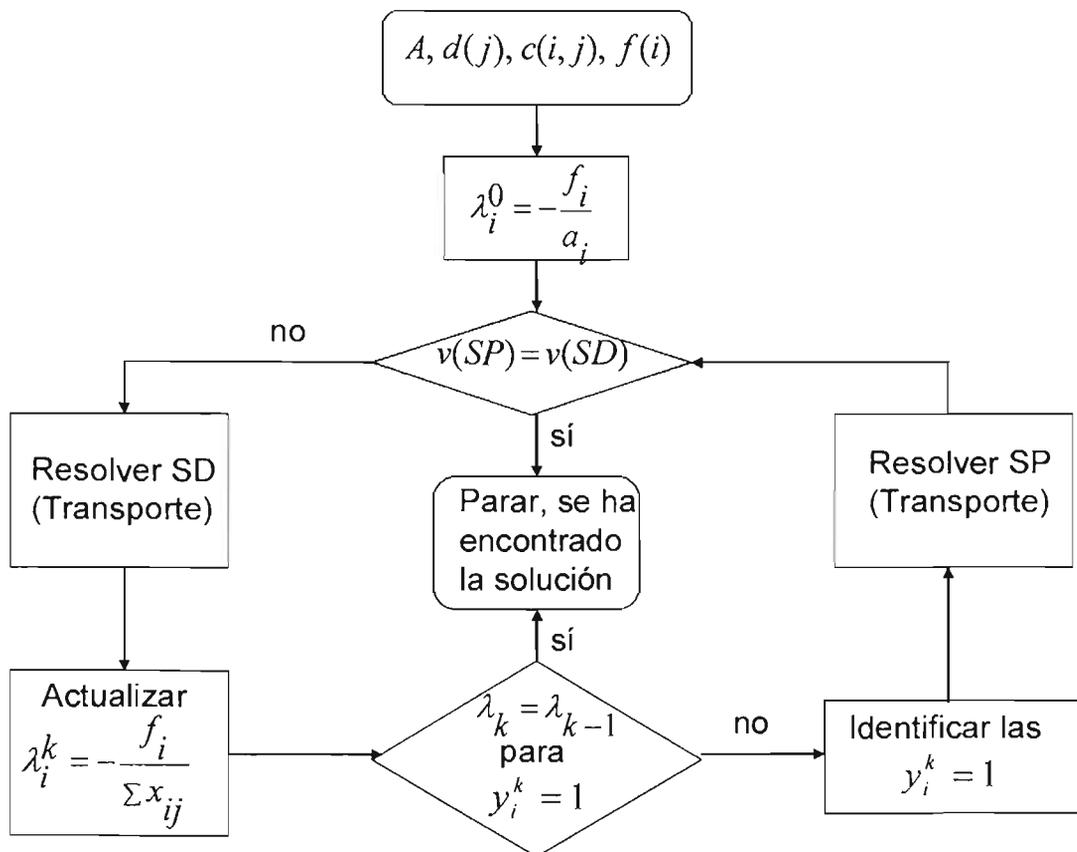


Fig. 15 Diagrama del algoritmo de Descomposición Cruzada Separable.

Convergencia del algoritmo

Un análisis de la convergencia del algoritmo de descomposición cruzada, desarrollado por Holmberg en 1990, indica que el algoritmo de descomposición cruzada tiene convergencia finita para problemas en los cuales el algoritmo de descomposición de Dantzig-Wolfe o el algoritmo de descomposición de Benders tengan convergencia finita. Este resultado es muy ventajoso, ya que la motivación para utilizar la descomposición cruzada es la eficiencia práctica, *i.e.*, es desarrollada para tener una rápida convergencia en la práctica, al reemplazar los problemas maestros por subproblemas más sencillos, tanto como sea posible.

Para el caso particular del problema AVD, se tiene que:

- i) El subproblema primal SP_y resulta ser un problema de transporte.
- ii) Al utilizar la relajación lagrangeana Separable, el subproblema dual SD_λ , resulta ser también un problema de transporte.

De tal forma es necesario observar si el algoritmo utilizado para solucionar los problemas de transporte tiene convergencia finita.

Como se mostró anteriormente, en iteraciones sucesivas, las rutas más económicas se van haciendo cada vez más baratas y en consecuencia se les asigna cada vez más flujo hasta que se saturan, ya sea agotando la capacidad del vehículo o satisfaciendo la demanda, es decir:

$$\sum_j x_{ij}^{k-1} \leq \sum_j x_{ij}^k \leq \sum_j x_{ij}^{k+1} \leq \dots \text{ con } \lambda_i^{k-1} \leq \lambda_i^k \leq \lambda_i^{k+1} \leq \dots \leq -\frac{f_i}{a_i}$$

De esta forma, se pueden utilizar los vehículos que satisfacen a la demanda obtenidos de la solución del subproblema dual Lagrangeano SD_λ como vehículos utilizados; es decir, los $y_i = 1$, con $i = 1, \dots, m$, que se fijarán en el subproblema primal SP_y .

Por otro lado, las etapas

- i) Fijar y_i es su valor actual y resolver el subproblema de Benders SP_y , que es un problema de transporte, para generar un nuevo valor para la cota superior de P ;
- ii) Fijar λ_i en su valor actual y resolver el subproblema Lagrangeano SD_λ , para generar un nuevo valor para la cota inferior de P ;

producen sucesivamente valores para las cotas inferior y superior, respectivamente, y es posible elegir cuidadosamente un valor muy pequeño ε , utilizado para definir una diferencia aceptable entre los valores óptimos de los subproblemas y no tener que resolver problemas maestros.

De lo anterior se afirma que el algoritmo de descomposición cruzada separable termina en un número finito de iteraciones, y obtiene la solución al problema AVD de la Fase II, ya que como se mencionó anteriormente, el principio fundamental de la Descomposición Cruzada, establece que los subproblemas primal SP_y y dual SD_λ , son problemas maestros relajados uno del otro y son pares duales.

3.6 Fase III: Definición de la sucesión de visitas.

Una vez que se sabe qué vehículos atenderán a cada cliente y qué cantidad de producto debe ser entregada por cada vehículo a cada cliente, es necesario conocer cuál es la sucesión de visitas que se harán a esos clientes, partiendo y regresando al almacén central y que hará menores los costos de transporte, los cuales dependen principalmente de las distancias recorridas. El problema mencionado es el típico Problema del Agente Viajero (TSP). En el Apéndice D se presenta un estudio comparativo de algoritmos para resolver el TSP de Castañeda (2000).

El TSP consiste en que “un viajero debe visitar cada ciudad exactamente una vez y debe regresar al punto de partida”. Se conoce el costo de viajar de cada ciudad a todas las demás. De tal forma, el viajero debe planear su itinerario de viaje, logrando que el costo total sea mínimo. Por lo que en términos generales el problema es encontrar el costo mínimo del circuito de longitud n .

De manera formal el problema se define como:

Instancia : Sea un conjunto C de m ciudades, la distancia $d(c_i, c_j) \in Z^+$ para cada par de ciudades $c_i, c_j \in C$, y un número entero positivo B .

Pregunta : ¿Hay una ruta de C que tenga una longitud B o menor, por ejemplo una permutación $\langle c_{\pi(1)}, c_{\pi(2)}, \dots, c_{\pi(K)}, c_{\pi(K+1)}, \dots, c_{\pi(m)} \rangle$ de C tal que

$$\left(\sum_{i=1}^{m-1} d(c_{\pi(i)}, c_{\pi(i+1)}) \right) + d(c_{\pi(m)}, c_{\pi(1)}) \leq B?$$

Considerando este problema como uno de teoría de grafos, se puede considerar que es un grafo a cuyas aristas entre los vértices, que representan las ciudades

que debe visitar el agente viajero, se les ha asociado un peso, que representa la distancia, costo, tiempo o alguna otra cantidad que se pretenda minimizar. Así, el objetivo es encontrar la ruta mínima que pase por cada ciudad exactamente una vez y regrese a la ciudad inicial. Nuevamente en términos de teoría de grafos, la meta es encontrar un circuito Hamiltoniano que minimice la suma de los pesos de las aristas.

Existe una gran cantidad de información referente al TSP, así como una gran cantidad de algoritmos para su solución. Mediante la consulta de trabajos recientes, se eligió un método que muestra buenos resultados en la solución de problemas con características similares a las del problema de la Fase III de la estrategia propuesta, tales como costos asociados, topología de la red, número de ciudades y facilidad de aplicación e implementación. De tal forma, el algoritmo propuesto es el método heurístico híbrido Dos Optimal-Prim, cuyos componentes son el Método Dos Optimal y el Método de Adaptación Prim. La información respecto de estos métodos, fue obtenida de Castañeda (2000), quien tiene un interesante trabajo de comparación de métodos heurísticos y exactos para el TSP.

3.6.1 Método Dos Optimal.

Esta técnica de solución recibe el nombre abreviado de "2-Opt" o bien "2-Operadores" u "Opción Doble". Según varios autores, esta es una de las heurísticas más exitosas para obtener una solución cercana al óptimo del TSP.

La heurística inicia con un ciclo Hamiltoniano llamado H , esta ruta inicial para el TSP puede ser arbitraria u ordenada $(1,2,\dots,n)$. Después se eliminan r aristas de H , obteniendo r caminos desconectados, algunos de los cuales pueden ser nodos aislados. A continuación se conectan estos r caminos de tal forma que producen otra ruta para el TSP llamada H' , en la que se usan aristas diferentes de aquellas que fueron eliminadas de H . Así H y H' difieren una de la otra en exactamente r aristas; las $(n-r)$ aristas que permanecen son las aristas en común. La siguiente figura ilustra el método 2-Opt.

Después se calcula el peso total, denominado $w(H')$, de la ruta H' , y si $w(H') < w(H)$, entonces se reemplaza H por H' y se repite el proceso; en otro caso, se toma otro conjunto de r aristas de H para intercambiarlas. Los intercambios (de conjuntos de r aristas) continúan hasta que ninguna otra mejora se pueda realizar por intercambio de r aristas. La solución final es llamada r -Optimal o r -Opt, y evidentemente en este caso se utiliza $r = 2$ obteniendo el método Dos-Optimal o 2-Opt. El procedimiento de intercambio de aristas terminará en un óptimo local, no necesariamente en un óptimo global, logrando una solución aproximada.

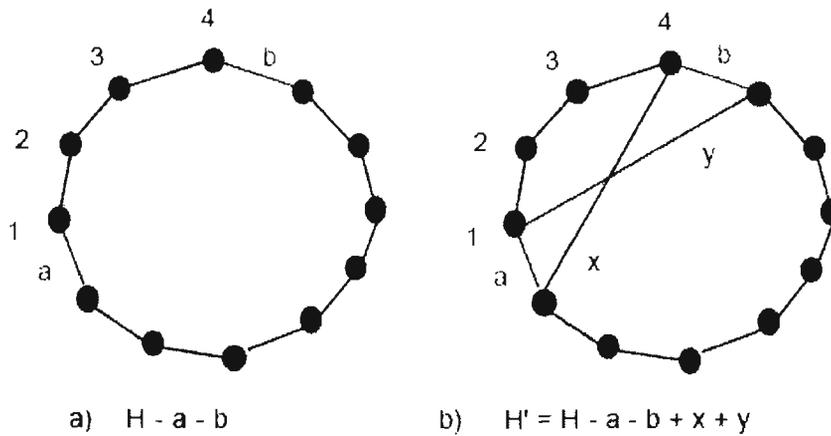


Fig. 16 Unión de vértices para formar un camino 2-Opt.

Este algoritmo para el TSP, ilustra una aproximación para resolver problemas de optimización, que se conocen como búsqueda local o por vecindades. Este método de mejoras sucesivas en la ruta del TSP, se usa para instancias simétricas, tanto como para asimétricas. Para los TSP simétricos con número de nodos $n \geq 5$, el valor de r puede variar desde 2 a n , sin embargo para los TSP asimétricos r no puede ser menor que 3.

De manera general, mientras mayor sea el valor de r en el procedimiento de intercambio, mejor será la solución, sin embargo, el esfuerzo computacional también se incrementa rápidamente. Hay $\binom{n}{r}$ subconjuntos de r aristas en un ciclo con n aristas. Intercambiar cada uno de estos pares de aristas una vez requiere tiempo $O(n^r)$, por lo que se debe decidir entre la exactitud de la solución y el esfuerzo computacional. Si $r = 2$, se tendrá un tiempo $O(n^2)$. Castañeda (2000), afirma que se propone el 2-Opt porque da muy buenas aproximaciones, particularmente si la ruta inicial es buena.

El procedimiento 2-Opt es una solución aproximada al problema simétrico del TSP con n nodos y una matriz de adyacencia o de pesos W . El ciclo inicial consiste del conjunto de aristas (de un TSP simétrico) $H = (x_1, x_2, \dots, x_n)$ en el orden x_1, x_2, \dots, x_n . Sea $X = \{x_i, x_j\}$ el conjunto de dos aristas en H , las cuales serán eliminadas y reemplazadas por las aristas $Y = \{y_p, y_q\}$, si hay una mejora. Esto es, $H' = (H - X) \cup Y$ es una ruta nueva y mejorada. Obsérvese que:

- a) las dos aristas x_i, x_j en X no pueden ser adyacentes y que
- b) una vez que el conjunto X ha sido seleccionado, el conjunto Y es determinado.

Así, hay

$$\frac{n(n-1)}{2} - n = \frac{n(n-3)}{2}$$

rutas posibles H' para una H dada. Para cualquiera de estas rutas se denotará la mejora con δ , obtenida por medio de:

$$\delta = w(H) - w(H') = w(x_i) + w(x_j) - w(y_p) - w(y_q)$$

Se examinarán todas las $n(n-3)/2$ rutas H' y se conservará una de ellas, aquella cuya δ sea máxima. Si $\delta_{\max} > 0$, se usa la correspondiente solución como la ruta inicial y se repite todo el proceso, *i.e.*, se continua sucesivamente la mejora de la ruta hasta que δ_{\max} sea un número negativo o cero.

El siguiente pseudocódigo describe al algoritmo 2-Opt como técnica de solución para el TSP, usando reemplazos de dos aristas en forma sucesiva:

```

{01} Procedure 2-Opt
{02}   begin
{03}     Sea  $H = (x_1, x_2, \dots, x_n)$  la ruta inicial
{04}     repeat
{05}        $\delta_{\max} \leftarrow 0$ ;
{06}       for  $i \leftarrow 1$  to  $(n-2)$  do
{07}         for  $j \leftarrow (i+2)$  to  $n$  or  $(n-1)$  do
{08}           begin
{09}             // el último caso es solamente cuando  $i = 1$ 
{10}             if  $(w(x_i) + w(x_j)) - (w(y_p) + w(y_q)) > \delta_{\max}$ 
{11}             then
{12}               // se ha encontrado un mejor par de aristas
{13}               begin
{14}                  $\delta_{\max} \leftarrow (w(x_i) + w(x_j)) - (w(y_p) + w(y_q))$ ;
{15}                 guardar tanto a  $i$  como a  $j$ 
{16}               end del if;
{17}             end del for j;
{18}           if  $\delta_{\max} > 0$ 
{19}           then
{20}             begin
{21}               // intercambia un par de aristas, porque la ruta es ahora más
{22}               corta
{23}                $H = H - \{x_i, x_j\} \cup \{y_p, y_q\}$ 
{24}             end del if;
{25}           until  $\delta_{\max} = 0$ ;
{26}     end del Procedure 2-Opt

```

En el Apéndice D se da un ejemplo del algoritmo anterior, realizando una descripción más detallada para un TSP simétrico de 5 ciudades.

3.6.2 Método Adaptación Prim.

El método Adaptación Prim pudo ser adaptado al TSP, debido a que produce un árbol de expansión mínima T (Minimum spanning tree), que es un grafo no dirigido conectado, con aristas ponderadas. Inicialmente se elige un vértice arbitrario y se permite que T sea un árbol que consiste de ese vértice seleccionado. Se repite la etapa de selección del siguiente vértice $n - 1$ veces, tomando en cuenta qué arista tiene el costo mínimo con exactamente un punto final en T , incluyendo ese vértice en T . Se dice que es un algoritmo voraz o glotón, porque siempre añade la arista de costo mínimo y suprime la mayor.

El algoritmo hace elecciones mínimas, manteniendo simultáneamente el subgrafo conexo y acíclico. No necesita que las aristas del grafo G estén ordenadas de antemano. La adaptación consiste en tomar los vértices del árbol abarcador mínimo y producir con ellos un ciclo hamiltoniano.

Sea $Best_S_so_Far$ el conjunto de vértices en el momento en el que el algoritmo se detiene. Sean $V(T)$ los vértices del subgrafo T (árbol de expansión mínima). Sea w el peso de la suma de las aristas ponderadas del subgrafo T y sea \mathbf{a} la matriz de adyacencia del grafo G . El procedimiento forma un árbol de expansión mínima. En cada etapa el algoritmo busca una arista de menor peso que conecte un vértice en T con un nuevo vértice fuera de T . Entonces añade esa arista y ese vértice a T y repite el proceso. El pseudocódigo del algoritmo se describe a continuación:

```

{01} Function Adaptación_Prim
{02}   begin
{03}     Se inicializa el vector col en ceros;
{04}     menor_en_fil  $\leftarrow$  •;
{05}     dir  $\leftarrow$  0;
{06}     i  $\leftarrow$  0;
{07}     Best_S_so_Far[0]  $\leftarrow$  0;
{08}     col[0]  $\leftarrow$  1;
{09}     for j  $\leftarrow$  1 to n_vertices - 1 do
{10}       begin
{11}         menor_en_fil  $\leftarrow$  •;
{12}         for k  $\leftarrow$  0 to n_vertices - 1 do
{13}           begin
{14}             if (col[k] = 0) and (menor_en_fil >  $\mathbf{a}$ [Best_S_so_Far[i]][k])
{15}               begin
{16}                 menor_en_fil  $\leftarrow$   $\mathbf{a}$ [Best_S_so_Far[i]][k];
{17}                 dir  $\leftarrow$  k;
{18}               end del if;
{19}             end del for k;

```

```

{20}     Best_S_so_Far[j] ← dir;
{21}     col[Best_S_so_Far[j]] ← 1;
{22}     i ← j;
{23}     w ← w + menor_en_fil;
{24}     end del for j
{25}     w ← w + a[Best_S_so_Far[n_vertices-1]][0];
{26}     return w;
{27}     end de la function Adaptación_Prim

```

El tiempo para encontrar la x más cercana en \mathbf{a} , y después actualizar registros es $O(n)$. Sin embargo, en el peor de los casos corre en un tiempo $O(n^2)$ debido a los dos ciclos anidados (líneas 09 y 12).

En el Apéndice D se muestra un ejemplo de método, de adaptación Prim para un TSP simétrico de 5 ciudades.

3.6.3 Método Híbrido Dos Optimal-Prim.

Según Castañeda (2000), la idea de la hibridación surgió debido a que la teoría de Dos-Optimal menciona que este método es mejor cuando se le inicia con una buena ruta. De tal forma que se decidió usar el Método de Adaptación de Prim para encontrar esa ruta inicial, y utilizarla para iniciar el método Dos-Optimal, así se obtiene el método Híbrido Dos optimal-Prim. Es este método híbrido es el que se propone para ser usado en la Fase III de la estrategia propuesta.

3.7 Ejemplo de aplicación de la estrategia propuesta.

Un distribuidor de grandes instrumentos, tiene un subsistema que involucra la distribución y el manejo de inventarios para un refrigerador específico, de 1m^3 , como una de sus líneas de producto. Se tienen un almacén central desde donde se realiza la distribución hacia 20 clientes, para lo cual se cuenta con 5 vehículos con diferentes capacidades.

Se requiere entonces encontrar las políticas de abastecimiento y distribución, para el día, que minimicen los costos relacionados con el inventario y la transportación, asegurando el abasto del producto.

El problema que se pretende resolver es un IRP con las características mencionadas en la Tabla 4. Para su solución se aplicará la estrategia propuesta en el presente trabajo. En la aplicación de cada fase se mostrarán los datos necesarios.

Fase I

Para aplicar el esquema de abastecimiento del ORS se tienen, para cada cliente, los costos y la demanda, y se calculan los valores de M, I, Q, q .

Con ellos se toma la decisión de atender o no, a cada cliente. La información utilizada y las decisiones tomadas se muestran a continuación.

Cliente	Costos						Decisión
	Ordenar	Almacenamiento	D	I	Q	q	
1	0.4	0.8	5	3	3	2	Ordenar q
2	0.6	0.8	5	3	3	2	Ordenar q
3	0.9	0.8	6	4	4	2	Ordenar q
4	0.9	0.8	2	1	3	1	Ordenar q
5	0.6	0.8	5	3	3	2	Ordenar q
6	0.6	0.8	6	5	3	1	No ordenar
7	0.6	1	3	3	2	0	No ordenar
8	0.2	0.6	9	3	3	6	Ordenar q
9	0.6	0.8	5	3	3	2	Ordenar q
10	0.4	1	5	2	2	3	Ordenar q
11	0.9	0.8	5	5	4	0	No ordenar
12	0.2	0.6	10	7	3	3	No ordenar
13	0.9	1	5	2	3	3	Ordenar q
14	0.4	0.8	7	2	3	5	Ordenar q
15	0.4	0.8	7	7	3	0	No ordenar
16	0.5	0.6	4	3	3	1	Ordenar q
17	0.9	0.6	1	0	2	1	Ordenar q
18	0.9	0.8	7	5	4	2	No ordenar
19	0.9	0.8	2	2	3	0	Ordenar q
20	0.4	0.8	8	3	3	5	Ordenar q

Como se puede observar, en el día deben atenderse sólo a 13 de los 20 clientes posibles. De tal forma que la pregunta: ¿a qué clientes se debe atender en el día?, queda respondida.

Fase II

En esta fase se aplica la Descomposición Cruzada Separable. Los datos necesario son:

$$m = \text{número de vehículos} = 5, \quad i = 1, 2, \dots, 5$$

$$n = \text{número de clientes} = 13, \quad j = 1, 2, \dots, 13$$

Vehículos	
Capacidad (m^3)	Costo fijo
$a_1 = 16$	$f_1 = 25$
$a_2 = 16$	$f_2 = 25$
$a_3 = 8$	$f_3 = 40$
$a_4 = 8$	$f_4 = 40$
$a_5 = 4$	$f_5 = 20$

La demanda de los 13 clientes que deben atenderse es:

$$\text{Demanda (unidades de producto)} = d = [2, 2, 2, 1, 2, 6, 2, 3, 3, 5, 1, 1, 5], \quad \sum_{j=1}^{13} d_j = 35$$

La matriz del problema de transporte que contiene los costos de viaje de cada vehículos hacia cada cliente, las capacidades de cada vehículos y las demandas de los clientes es:

	Clientes													Capacidad
	1	2	3	4	5	6	7	8	9	10	11	12	13	
Vehículo 1	327	138	630	712	555	792	786	1305	990	1347	1603	1456	1680	16
Vehículo 2	327	138	630	712	555	792	786	1305	990	1347	1603	1456	1680	16
Vehículo 3	240	101	462	522	407	581	576	957	726	988	1176	1068	1232	8
Vehículo 4	240	101	462	522	407	581	576	957	726	988	1176	1068	1232	8
Vehículo 5	153	64	294	332	259	370	367	609	462	629	748	680	784	4
Demanda	2	2	2	1	2	6	2	3	3	5	1	1	5	

Con esta información se comienza a aplicar la técnica de descomposición cuyo algoritmo se muestra en la sección 3.5.3.

Iteración 1

Paso 1 Sean $v_p(-\infty)$, $v_p(+\infty)$. Calcular los valores $\lambda_i^0 = -\frac{f_i}{a_i}$, $i = 1, \dots, 5$, con todos los vehículos en uso $Y_i^0 = 1$, $i = 1, \dots, m$.

$$\begin{array}{cccccc}
 y_1^0 = 1 & y_2^0 = 1 & y_3^0 = 1 & y_4^0 = 1 & y_5^0 = 1 & \\
 \lambda_1^0 = -1.6 & \lambda_2^0 = -1.6 & \lambda_3^0 = -5 & \lambda_4^0 = -5 & \lambda_5^0 = -5 &
 \end{array}$$

Paso 2 Resolver SD_{λ^0} con el incremento de λ_i^0 , tal que la matriz de costos con $c_{ij} - \lambda_{ij}$, es

	Clientes													Capacidad
	1	2	3	4	5	6	7	8	9	10	11	12	13	
Vehículo 1	328.6	139.6	631.6	713.6	556.6	793.6	787.6	1306.6	991.6	1348.6	1604.6	1457.6	1681.6	16
Vehículo 2	328.6	139.6	631.6	713.6	556.6	793.6	787.6	1306.6	991.6	1348.6	1604.6	1457.6	1681.6	16
Vehículo 3	245	106	467	527	412	586	581	962	731	993	1181	1073	1237	8
Vehículo 4	245	106	467	527	412	586	581	962	731	993	1181	1073	1237	8
Vehículo 5	158	69	299	337	264	375	372	614	467	634	753	685	789	4
Demanda	2	2	2	1	2	6	2	3	3	5	1	1	5	

Este problema de transporte se resuelve y su solución es

	Clientes													Capacidad
	1	2	3	4	5	6	7	8	9	10	11	12	13	
Vehículo 1														16
Vehículo 2	2	2	2	1	2	4	2							16
Vehículo 3								3	3	2				8
Vehículo 4										3	1	1	1	8
Vehículo 5						2								4
Demanda	2	2	2	1	2	6	2	3	3	5	1	1	5	

Costo total = 26639

Así que $v(SD) = 26639$.

Paso 2.1 Calcular nuevos valores $\lambda_i^1 = -\frac{f_i}{\sum_j x_{ij}}$ para los

servicios abiertos

$$\begin{array}{cccccc} y_1^1 = 0 & y_2^1 = 1 & y_3^1 = 1 & y_4^1 = 1 & y_5^1 = 1 \\ \lambda_1^1 = -1.6 & \lambda_2^1 = -1.7 & \lambda_3^1 = -5 & \lambda_4^1 = -5 & \lambda_5^1 = -5 \end{array}$$

Paso 3 Probar si $\lambda_i^0 = \lambda_i^1$ para los servicios abiertos,

$$y_2^1 = y_3^1 = y_4^1 = y_5^1 = 1$$

i	λ_i^0	λ_i^1
2	-1.6	-1.7
3	-5	-5
4	-5	-5
5	-5	-5

Como no todos son iguales, identificar los servicios abiertos, que resultan ser y_2, y_3, y_4 y y_5 .

Paso 4 Resolver el subproblema primal SP_{v_i} pero sólo para los servicios abiertos

	1	2	3	4	5	6	7	8	9	10	11	12	13	Capacidad
1														0
2	2	2	2	1	2	4	2							16
3								3	3	2				8
4						2				3	1	1	1	8
5													4	4
Demanda	2	2	2	1	2	6	2	3	3	5	1	1	5	

Costo total = 26515+25+40+40+20=26640

Como se ve $v(SP) = 26640$.

Paso 5 Probar si $v(SP) = v(SD)$. Dado que $v(SD) = 26639$ y $v(SP) = 26640$ y por lo tanto difieren en una unidad se decide observar qué pasa al hacer otra iteración y así, se regresa al paso 2 con λ_i^2 .

Iteración 2

Paso 2 Resolver el subproblema SD_{λ_i} con el incremento de las λ_i^1 , tal que la matriz de costos con $c_{ij} - \lambda_{ij}$, pero sólo con los vehículos en uso

		Clientes												Capacidad	
		1	2	3	4	5	6	7	8	9	10	11	12	13	
Vehiculo	1	328.6	139.6	631.6	713.6	556.6	793.6	787.6	1306.6	991.6	1348.6	1604.6	1457.6	1681.6	0
	2	328.7	139.7	631.7	713.7	556.7	793.7	787.7	1306.7	991.7	1348.7	1604.7	1457.7	1681.7	16
	3	245	106	467	527	412	586	581	962	731	993	1181	1073	1237	8
	4	245	106	467	527	412	586	581	962	731	993	1181	1073	1237	8
	5	158	69	299	337	264	375	372	614	467	634	753	885	789	4
Demanda		2	2	2	1	2	6	2	3	3	5	1	1	5	

Este problema de transporte se resuelve y su solución es

		Clientes												Capacidad	
		1	2	3	4	5	6	7	8	9	10	11	12	13	
Vehiculo	1														0
	2	2	2	2	1	2	4	2							16
	3								3	3	2				8
	4														8
	5						2				3	1	1	1	4
Demanda		2	2	2	1	2	6	2	3	3	5	1	1	5	

Costo total = 26639.8

Así que, $v(SD) = 26639.8$.

Paso 2.1 Calcular nuevos valores $\lambda_i^2 = -\frac{f_i}{\sum_j x_{ij}}$ para los servicios abiertos

$$\begin{array}{ccccc}
 y_1^2 = 0 & y_2^2 = 1 & y_3^2 = 1 & y_4^2 = 1 & y_5^2 = 1 \\
 \lambda_1^2 = -1.6 & \lambda_2^2 = -1.7 & \lambda_3^2 = -5 & \lambda_4^2 = -5 & \lambda_5^2 = -5
 \end{array}$$

Paso 3 Probar si $\lambda_i^1 = \lambda_i^2$ para los servicios abiertos, $y_2^1 = y_3^1 = y_4^1 = y_5^1 = 1$

i	λ_i^0	λ_i^1
2	-1.7	-1.7
3	-5	-5
4	-5	-5
5	-5	-5

Como todos son iguales el algoritmo termina.

De acuerdo a lo anterior, la solución de la Fase II es:

		Clientes												Capacidad	
		1	2	3	4	5	6	7	8	9	10	11	12	13	
Vehiculo	1														0
	2	2	2	2	1	2	4	2							16
	3								3	3	2				8
	4														8
	5						2				3	1	1	1	4
Demanda		2	2	2	1	2	6	2	3	3	5	1	1	5	

Costo total = 26639.8

Sólo se utilizan los vehículos 2,3,4 y 5. El vehículo 2 atenderá a los clientes: 1,2,3,4,5,6 y 7; el vehículo 3 atenderá a los clientes: 8,9 y 10; el vehículo 4 atenderá a los clientes: 6,10,11,12 y 13; el vehículo 5 atenderá al cliente: 13; y el patrón de entregas se puede ver en las celdas de la matriz.

Una vez que se ha determinado qué vehículos son los que se utilizarán, su asignación para la atención de los clientes y la cantidad de producto que será abastecido, se pasa a la última fase.

Fase III

En esta fase se realiza el ruteo para cada vehículo, resolviendo un TSP para cada uno. Debido a que para el ejemplo, el número de clientes es pequeño, se utilizó el método de Bifurcación y Acotación, obteniéndose los siguientes resultados.

Matriz de distancias de viaje para el vehículo 2

	Alm.	1	2	3	4	5	6	7
Alm.		218	92	420	475	370	528	524
1	218		310	202	464	588	310	606
2	92	310		512	383	278	620	432
3	420	202	512		262	400	108	404
4	475	464	383	262		138	370	196
5	370	588	278	400	138		450	154
6	528	310	620	108	370	450		296
7	524	606	432	404	196	154	296	

La ruta solución es la siguiente: Alm. \Rightarrow 2 \Rightarrow 5 \Rightarrow 4 \Rightarrow 7 \Rightarrow 6 \Rightarrow 3 \Rightarrow 1 \Rightarrow Alm. con una distancia de 1528.

Matriz de distancias de viaje para el vehículo 3

	Alm.	8	9	10
Alm.		870	660	898
8	870		368	85
9	660	368		238
10	898	85	238	

La ruta solución es la siguiente: Alm. \Rightarrow 9 \Rightarrow 10 \Rightarrow 8 \Rightarrow Alm. con una distancia de 1853.

Matriz de distancias de viaje para el vehículo 4

	Alm.	6	10	11	12	13
Alm.		528	898	1069	971	1120
6	528		427	652	649	649
10	898	427		225	323	222
11	1069	652	225		98	248
12	971	649	323	98		346
13	1120	649	222	248	346	

La ruta solución es la siguiente: Alm. \Rightarrow 12 \Rightarrow 11 \Rightarrow 13 \Rightarrow 10 \Rightarrow 6 \Rightarrow Alm. Con una distancia de 2494.

El vehículo 5 atenderá únicamente el cliente 13 viajando una distancia de 1568.

Capítulo 4

Eficiencia Computacional.

A pesar del interés que existe en el IRP dentro de la comunidad científica, es muy pobre la información que se obtiene de la comparación con soluciones obtenidas por otros autores y con las de bancos de problemas disponibles en línea, por ejemplo en www.tli.gatech.edu/. Esto se debe a que los resultados reportados implican diferentes instancias, o problemas diferentes que además suelen ser extremadamente particulares, incluyendo características tan diversas como: permitir o no la visita de un solo vehículo a cada cliente, permitir o no envíos directos, manejo de diferentes capacidades de vehículos, número de vehículos como variable de decisión, definición, simetría y densidad de las matrices de distancias o costos, etc. Adicionalmente, los recursos de programación y de equipos en que se realizan las corridas son también muy diferentes entre sí.

Por otro lado, al hablar de eficiencia de los algoritmos, los rápidos avances en la teoría de las computadoras hacen que las medidas físicas tales como tiempo de corrida o requerimientos de memoria, sean irrelevantes. Una medida más estandarizada es el número de operaciones elementales de cómputo que toma el resolver un problema en el peor caso. El desempeño promedio no es seguro. Puede haber casos particulares, técnicamente llamados instancias de un problema, que se comportan mucho peor que el promedio y nadie puede dar confianza en la suerte de intentar resolverlo.

La complejidad computacional estudia: la eficiencia de los algoritmos y la dificultad inherente de problemas de importancia práctica y/o teórica. En este contexto, el IRP es un problema NP-duro (ver apéndice A), cuyo término técnico es NP-completo, lo cual podría significar “abandonar toda esperanza de definir un algoritmo eficiente para la solución exacta de ese problema”. Sin embargo, conociendo esto, no se gasta el tiempo en proyectos imposibles y en lugar de ello se definen soluciones aproximadas para resolver casos especiales o se altera un poco el problema para transformarlo en un problema menor que sea tratable, como lo hacen las técnicas de descomposición.

El presente capítulo analiza y define la eficiencia de la estrategia propuesta. La primera sección analiza la estrategia, desde el enfoque de complejidad computacional (ver Apéndice A). Una de las razones es que a este enfoque se le considera el más formal; otra es que los pocos problemas reportados en literatura que coinciden con las características del IRP resuelto en este trabajo de investigación, no presentan aplicación ni datos contra los cuales comparar la eficiencia del algoritmo. Además, las condiciones en las que se desarrollan las pruebas son muy heterogéneas y dependen del equipo de cómputo utilizado, del tamaño y estructura de las matrices, de la topología de la red de distribución, etc.

Una segunda sección muestra los resultados del desempeño de la estrategia propuesta al aplicarla en experimentos computacionales con diferentes instancias del IRP tratado, diseñados aleatoriamente. Finalmente, el tercer apartado del capítulo compara a la estrategia con desarrollos similares reportados en literatura, especificando sus ventajas sobre ellos.

4.1 Complejidad de la estrategia propuesta.

En la idealización de la estrategia, se consideró en todo momento definir algoritmos que resolvieran problemas más tratables de manera sencilla e implementable. Como se citó, la parte medular de la estrategia es la aplicación de la Descomposición Cruzada Separable en la Fase II. Con ello se obtiene un algoritmo que, cuando se aplica al problema AVD, produce la solución sucesiva de dos subproblemas del tipo de transporte, incorporando un proceso de ping-pong entre ellos, evitando por completo el uso de problemas maestros a resolver. Puesto que es posible resolver el problema AVD empleando una serie de problemas de transporte. Es interesante hacer notar la enorme ventaja de tener este tipo de problemas, ya que sus algoritmos de solución son de bajo orden de complejidad, de hecho es $O(n^3)$ (Toth y Vigo, 2002). Otra ventaja es que existe una gran cantidad de software comercial con rutinas para resolverlos, e incluso se puede usar la programación con lenguajes de propósito general.

La estrategia propuesta está compuesta por las tres fases que fueron descritas. Los procesos de cada fase tienen el siguiente tiempo de ejecución:

Fase de la estrategia	Orden del algoritmo
Fase I	$O(n)$
Fase II	$O(n^3)$
Fase III	$O(n^2)$

Tabla 8 Complejidad de las fases de la estrategia propuesta.

De tal forma se considera que el orden de la estrategia propuesta es $O(n^3)$, ya que es el orden que acota superiormente a la estrategia completa. Esta situación la hace extremadamente eficiente y se espera que al ser comparada con los algoritmos que resuelven la misma instancia de IRP, sea siempre mucho más ventajosa.

4.2 Experiencia computacional.

Como se mencionó, la heterogeneidad de las instancias y problemas disponibles hacen imposible la comparación computacional de la eficiencia de los algoritmos (Toth y Vigo, 2002). Sin embargo, para conocer el desempeño del algoritmo propuesto, se resolvieron problemas que fueron generados aleatoriamente. Se definen a continuación las instancias y tipo de IRP que se resolvió con la estrategia propuesta.

Tomando en cuenta los resultados reportados en los trabajos revisados en el estado del arte, las instancias se pueden clasificar en “pequeñas” si tiene menos o hasta 10 nodos, “medianas” con más de 10 y menos de 50 nodos y “grandes” de 50 en adelante. Así, se generaron instancias entre 5 y 250 nodos. Además se utilizó una denominación de cada instancia propuesta por Toth y Vigo (2002) que permite determinar y unificar sus características. En particular, los nombres están conformados por tres campos que tienen la estructura $t\text{nnn-}kk$: el campo t es un carácter alfabético que indica el tipo del problema y que en este caso es igual a E , ya que todas las instancias tratadas son Euclidianas; el segundo campo llamado nnn consta de tres dígitos enteros que denotan el número de nodos, que incluye a los clientes y al almacén central; el tercer campo kk , consta de dos dígitos enteros que indican el número de vehículos disponibles.

4.2.1 Metodología.

Se evaluaron 9 problemas. Se manejaron 4 problemas pequeños con menos de 10 clientes, 2 medianos con 30 y 35 clientes respectivamente y 3 grandes, cada con 150, 200 y 250 clientes. Cabe mencionar que el último problema, era del tamaño más grande que permitieron los recursos de cómputo. Los problemas tienen las características mencionadas en la Tabla 7 (capítulo 3) y las configuraciones de valores para las capacidades, demandas y costos fijos, se tomaron de las instancias y ejemplos mostrados en Aceves (1996).

Los datos generados fueron:

- ⇒ Demandas de clientes como v.a. con distribución Uniforme (1,10).
- ⇒ Posiciones geográficas de los clientes: se generaron coordenadas X y Y con distribución Uniforme, para X : $U(1,44.6)$ y para Y : $U(1,60)$. Con

estos valores se emulaba el área correspondiente al Distrito Federal y, por consiguiente, actividades circunscritas al mismo.

- ⇒ Las matrices de distancias obtenidas a partir de las posiciones geográficas fueron matrices densas (de hecho, completamente llenas).
- ⇒ Las capacidades de los vehículos (m^3) tomaron los valores 4,8,16,y 24.
- ⇒ Los costos fijos variaron en 20, 25, 30,35,40,50.

Las variables aleatorias fueron generadas con el paquete estadístico STATGRAPHICS.

Para los problemas E005-03, E006-04, E008-02, E009-04, E031-09 y E036-11, fue posible obtener el óptimo, utilizando el algoritmo de Bifurcación-Acotación o Branch and Bound (B-B), por medio del paquete LINGO. Para los problemas grandes no es posible aún la obtención de óptimo con los métodos exactos con los que se cuenta.

Después todos los problemas se resolvieron con el algoritmo propuesto, por medio de un programa de cómputo desarrollado en Pascal. El sistema utilizado fue Pentium 4 2.00Ghz-256 MB RAM.

4.2.2 Resultados.

La tabla 9 muestra los resultados de las soluciones de los problemas trabajados.

La meta era evaluar la estrategia propuesta bajo condiciones de la vida real, poniendo énfasis en el balance entre la calidad de la solución obtenida y el correspondiente tiempo de cómputo.

Esta última consideración es de gran importancia para la generación de un plan de manejo de inventarios y distribución de productos, en situaciones reales, en las que se deben tomar decisiones rápidas continuamente.

Entre los resultados de la estrategia propuesta se encuentran:

- Tiempo de solución extremadamente corto, ya que en la solución de los problemas se realizaron como máximo tres iteraciones y en todos los casos se obtuvieron los resultados en 3 segundos o menos.
- La calidad de las soluciones fue excelente en los problemas pequeños y medianos: E005-03, E006-04, E008-02, E009-04, E031-09 y E036-11 ya que se obtuvo el óptimo.

Problema	Número de clientes	Vehículos	Solución con la estrategia propuesta	Tiempo	Solución óptima B-B	Tiempo	% de desviación de la solución óptima
E005-03	4	3	9819	2 seg	9819	3 seg	0
E006-04	5	4	16724	2 se	16724	3 seg	0
E008-02	7	2	21256	2 seg	21256	3 seg	0
E009-04	8	4	13114	2 seg	13114	3 seg	0
E031-09	30	9	4332	2 seg	4332	28 días	0
E036-11	35	11	3120	2 seg	3120	35 días	0
E151-08	150	8	14770	2 seg	No se conoce	-	-
E201-16	200	16	44586	3 seg	No se conoce	-	-
E251-20	250	20	72123	3 seg	No se conoce	-	-

Tabla 9 Comparaciones con la solución óptima.

Es importante considerar el enorme esfuerzo computacional que representó la solución de los problemas E031-09 y E036-11 con B-B, con 28 y 35 días respectivamente. Por lo tanto, para un sistema logístico en el que se toman decisiones diariamente, la obtención de la solución óptima con métodos exactos como B-B resulta totalmente ineficiente. Sin embargo, en este punto es necesario recordar que, la forma más adecuada de comparar la eficiencia de un algoritmo es con base en su complejidad computacional y no en el tiempo de obtención de la solución o en el número de iteraciones.

Entre las ventajas de la estrategia propuesta se encuentran:

- ❖ rapidez
- ❖ calidad de la solución
- ❖ posibilidad de tratar con gran cantidad de clientes
- ❖ implementación sencilla y de bajo costo.

El que sea fácil de programar hace que pueda ser construido software de acuerdo con las necesidades específicas de los usuarios, en cuanto a los formatos de entrada y reportes de salida. Gracias a la existencia de un programa de aplicación del algoritmo, el análisis del IRP del tipo “¿qué pasaría si?” se hace de manera sencilla, rápida y grandes posibilidades de retroalimentación en la toma de decisiones.

Desventajas:

- ❖ puede presentar brecha de optimalidad
- ❖ la solución de agente viajero es, finalmente, por medio de heurísticas
- ❖ no es posible comparar la calidad de la solución para problemas grandes.

Otra desventaja podría ser el problema inherente que se presenta al resolver problemas de transporte, ya que es necesario considerar que al aplicar el algoritmo de transporte, si un vector básico degenerado aparece, es matemáticamente posible entrar en una secuencia sin fin de pasos de pivoteo degenerados. Si esto ocurre, las soluciones básicas factibles y el valor de la función objetivo no cambiarán en la secuencia completa, y el criterio de optimalidad no se satisfará. En el algoritmo simplex, este fenómeno es llamado ciclaje por degeneración. Sin embargo, esta desventaja es relativa, ya que si bien el ciclaje por degeneración es matemáticamente posible, no se ha visto que ocurra cuando se resuelven problemas reales. Además, se han desarrollado técnicas simples especiales, para garantizar que el ciclaje por degeneración no ocurrirá. Estas técnicas agregan solamente un paso más de trabajo, pero han mostrado ser innecesarias al resolver problemas de la vida real (Murty, 1995).

4.3 Comparación con otras estrategias.

Se seleccionaron de entre todos los trabajos analizados en la Tabla 3 del Capítulo 1, aquellos que manejaban la misma instancia del IRP tratada. Se definieron tres características básicas que determinaban a un modelo reportado en literatura como similar al tratado en esta investigación y por tanto susceptible de comparación.

Las tres características fueron:

- el dominio de decisión: **tiempo**
- la topología de la red: **uno a muchos**
- tipo de la demanda: **determinística**

De tal forma, los problemas seleccionados y sus métodos de solución se muestran en la siguiente tabla 6

A continuación se menciona brevemente el trabajo tratado y se hace la comparación, en la medida de lo posible, con la estrategia propuesta.

Autores	Método de solución		
	Descomposición	Agrupación	Otros modelos
Chien, Balakrishnan y Wong (1989)	No	No	Heurístico y MIP (Rel. Lagrangeana)
Achutan (2000)	No	No	Bifurcación y Corte (MIP)
Campbell, Clarke y Savelsbergh (2002)	Agrupar I – Rutear II	Distancia	Heurístico e IP

Tabla 10 Métodos de solución para instancias similares al IRP tratado.

Chien et al (1989) Los autores sugieren la descomposición de un problema de periodo múltiple en una serie de problemas de un solo periodo. Se crea un programa entero que manipula i) el abastecimiento de un inventario limitado disponible en la planta para los clientes, ii) la asignación de clientes a cada vehículo y iii) la secuencia de la ruta. Su método consiste en relajar el problema, descomponiéndolo en dos subproblemas, uno de abastecimiento de inventario y otro subproblema que asigna vehículos a clientes.

Debido a que los costos de inventario son asumidos como una estructura lineal, los problemas de abastecimiento se reducen a un simple problema mochila continuo, el cual se resuelve eficientemente por un procedimiento glotón. La solución de estos dos problemas proporciona una cota inferior para el costo total, el cual puede ser mejorado por la actualización de los multiplicadores de Lagrange utilizando un método de búsqueda de subgradiente. Un procedimiento heurístico basado en la solución Lgrangeana es también desarrollado.

La experiencia computacional muestra que el procedimiento de solución es capaz de producir soluciones de buena calidad para varios problemas prueba generados aleatoriamente, variando las estructuras de los beneficios y los costos; y para combinaciones del número de vehículos disponibles y su capacidad. Se manejaron experimentos con 12 redes de 20, 25 y 30 nodos. Los nodos fueron generados aleatoriamente en una cuadrícula de 100x100 y los costos fijos y variables fueron múltiplos de la distancia Euclideana entre nodos. Se eligieron costos fijos de 10 veces el costo variable de viaje para cada viaje de i a j .

Las demandas fueron generadas como una v.a. uniformemente distribuida con media de 8 y 5 veces el promedio de la variable de costo, respectivamente. La cantidad de inventario disponible en el almacén fue de 80% de la demanda total esperada. Se evaluaron 4 combinaciones del número de vehículos disponibles (V) y su capacidad (b): ($V = 2, b = 16$), ($V = 2, b = 64$), ($V = 8, b = 16$), ($V = 8, b = 64$). Para cada combinación se generaron 3 instancias del problema.

El procedimiento se adecuó para producir soluciones entre 1-2% de optimalidad y no hubo una diferencia significativa entre diferentes combinaciones de (V, b) para cada tamaño de problema. En el análisis de segundos de CPU se observó que los

requerimientos computacionales para problemas de 30 nodos fueron, en promedio, de 3.6 veces con respecto a los problemas de 20 nodos. Este incremento en tiempo de computadora se consideró razonable y por esa razón implementable para tamaños más grandes de problemas.

La principal diferencia con este trabajo es que existe una capacidad de producto limitada en el almacén central. De hecho esta es la principal diferencia con respecto al IRP manejado, ya que esta situación determina restricciones que modifican de forma significativa las decisiones tomadas en cuanto a la asignación de cantidades de producto por entregar a los clientes. Crea además el problema de desabasto y se hace necesario manejar penalizaciones relacionadas con esa falta de producto para satisfacer la demanda.

Por otro lado, en la solución del MIP se utiliza un método de relajación Lagrangeana cuyo desempeño es menor al de la Descomposición Cruzada Separable empleada en esta investigación.

Achutan (2000) Trabaja con un algoritmo de Bifurcación y Corte para resolver un MIP de un solo producto para un período de planeación múltiple. En cada período de tiempo se debe i) decidir la cantidad de producto entregada a cada punto de venta y ii) resolver un VRP para la distribución. El modelo se construyó para un número máximo de 4 períodos de tiempo (días). El algoritmo fue desarrollado en lenguaje C utilizando el paquete CPLEX para resolver las respectivas relajaciones de LP y ejecutado en una computadora Silicon Graphics R5000.

Se evaluaron dos grupos de problemas, un grupo de 4 nodos y el otro de 5. El primer grupo fue adoptado de un ejemplo de 4 nodos de un trabajo previo de Vismanathan y Mathur (1977), para diferentes combinaciones de parámetros de costo. Todos los problemas tienen las mismas coordenadas de localización y tasas de demanda. Los costos de ordenación se eligieron de entre un rango de 40-100, los costos de almacenamiento se obtuvieron de un rango de 0.02-0.05.

En el segundo grupo, los problemas de prueba fueron generados aleatoriamente con 5 nodos. Las coordenadas de los puntos de venta fueron generadas con una distribución uniforme [0,100], las tasas de demanda se generaron con una distribución uniforme [0,1000]. El rango de valores para los parámetros de costo fue el mismo que para el primer grupo de problemas. Para todos los problemas se eligió un número máximo de período = 4. El número de variables binarias para el primer grupo fue 100 y 144 para el segundo.

El tiempo límite de corrida fue de 3 horas y para el primer conjunto se obtuvo el óptimo en todos los casos (10) y para el segundo sólo en 3 (de 10). A pesar de los resultados obtenidos no es posible todavía, resolver problemas grandes. Debido a lo anterior, la comparación con esta estrategia es improcedente, considerando que una de las motivaciones principales de esta investigación es la solución de instancias grandes del IRP.

Campbell et al. (2002) Desarrolla un caso real muy interesante, en la empresa norteamericana distribuidora de gas industrial PRAXAIR, que cuenta con 60 plantas productoras y atiende a 10,000 consumidores. La estrategia de solución utilizada consta de dos fases. En la primera, por medio de un IP, determina los conjuntos de clientes que deben ser atendidos, asegurando que no haya desabasto, cumpliendo con las cotas inferiores y superiores de los inventarios y tomando en consideración la localización geográfica, capacidades de inventario y tasas de uso compatibles. En la segunda fase utiliza un heurístico de inserción compuesto de un procedimiento glotón aleatorio de búsqueda adaptativa (GRASP), para construir las rutas de los conglomerados de clientes.

Soluciona problemas para 50 y 87 clientes obteniendo buenos resultados en cuanto a utilización de vehículos, volúmenes entregados, volúmenes entregados por milla recorrida, entre otros indicadores. Sin embargo, los autores mencionan el grave problema al que se enfrentan al resolver el modelo de IP de la primera fase, debido a la inmensa cantidad de combinaciones de grupos de clientes que se pueden presentar, a pesar del "pequeño" número que representan.

Es importante mencionar entonces, que esta dificultad sería superada mediante el uso de técnicas de descomposición, como el de la estrategia propuesta en este trabajo. De hecho, la inclusión de alguna de estas técnicas fortalecería enormemente el desempeño de la estrategia utilizada por Campbell. De tal forma, se considera que la estrategia propuesta en la presente investigación es más eficiente.

Conclusiones y Extensiones

1. Conclusiones.

El Problema de Inventario-Distribución es un importante problema logístico que ha recibido creciente atención en los años recientes. El IRP característico considera la actividad de distribución de una organización que opera en un almacén central y abastece a una gran conjunto de clientes diseminados geográficamente. El almacén central es el responsable del mantenimiento del adecuado nivel de inventario de los clientes. De tal forma, el IRP tiene la presencia simultánea de dos aspectos logísticos básicos: ruteo y control inventarios; estos dos combinados en un ambiente dinámico. En este trabajo de investigación se ha desarrollado y aplicado un estrategia eficiente para resolver instancias no triviales del problema mencionado, utilizando principalmente la técnica de Descomposición Cruzada Separable y una técnica heurística híbrida.

El desarrollo del estado del arte del IRP, permitió conocer que su modelación es relativamente nueva, y que no es posible decir aún, que existe una formulación general estándar. Este desarrollo también dio paso a la posibilidad de clasificación del IRP, con base en elementos tales como topología de la red, tipo de demanda y dominio de decisión, entre otros. Esta clasificación permite un mayor conocimiento del problema.

La extensa revisión de información permitió conocer las líneas de acción a seguir para el mejoramiento de las técnicas de solución utilizadas. Las dos clasificaciones principales con respecto a los métodos de solución son los métodos bajo Dominio de Frecuencia y los métodos bajo Dominio de Tiempo. Dentro de los métodos del primer tipo, el método de agregación, a pesar de ser de los primeros que surgieron, es el más robusto y utilizado por diversos autores, debido a la facilidad de comprensión y utilización, y la flexibilidad de adaptación a diferentes circunstancias. Por otro lado, en los métodos del segundo tipo, los modelos de flujo dinámico con demanda desconocida, representan un campo abierto de exploración y mejora.

Otro resultado del estado del arte muestra que, como en el caso del IRP, la integración de funciones logísticas, que originalmente estaban aisladas, tales como el control de inventarios y la distribución de productos, tiene el potencial de proporcionar beneficios significativos a las organizaciones, en la forma de ahorros en los costos y mejora de la eficiencia y servicio al cliente.

Sin embargo, es importante mencionar que para el éxito en la implementación de un modelo de IRP es necesario incluir consideraciones de tecnología EDI entre los clientes y el almacén central, así como los avances que existan en tecnologías semejantes, tales como el código electrónico EPC (Electronic Product Code o Tag) que ahora permitirá rastrear cada pieza de producto, desde que se produce en las líneas de ensamble hasta que se consume y desecha, y que mejorará radicalmente la precisión de los inventarios, el manejo de categorías en los anaqueles, reducción de faltantes y los costos en la cadena de suministro. Tecnologías como el EPC permitirán el surgimiento de nuevas líneas de investigación y la mejora de las existentes, en la investigación y el desarrollo de modelos de optimización de procesos logísticos.

Un punto importante de la conclusión es advertir la necesidad de promover y continuar la investigación para caracterizar la metodología de solución óptima de instancias grandes del IRP y reducir su complejidad computacional.

La estrategia de solución propuesta consta de tres fases. La primera fase indica a qué clientes debe visitarse en el día, empleando el método de control de inventarios de Reabastecimiento Opcional, por medio del cual el universo de clientes se reduce a sólo aquellos que “requieren” ser abastecidos.

Sin embargo, la parte medular de la estrategia propuesta recae en el uso de la Descomposición Cruzada Separable de la segunda fase, que al ser aplicada a un problema de asignación de vehículos y distribución, denominado AVD, asigna clientes a vehículos cumpliendo con la demanda de cada cliente y la capacidad máxima de los vehículos, al mínimo costo de viaje. Gracias a la aplicación de esa técnica de descomposición, se obtuvo un algoritmo que itera únicamente entre dos subproblemas de transporte, el subproblema dual (SD), con la siguiente estructura:

$$\begin{aligned}
 & \underset{x}{\text{Minimizar}} \quad \sum_i \sum_j (C_{ij} - \lambda_i) x_{ij} \\
 & \text{sujeto a} \\
 & \quad \sum_i x_{ij} \geq d_j, \quad \forall j \\
 & \quad \sum_j x_{ij} \leq a_i, \quad \forall i \\
 & \quad x_{ij} \geq 0, \lambda_i < 0, \forall i, j
 \end{aligned}$$

que se resuelve en el algoritmo, con valores de λ_i , para $i = 1, \dots, m$ de

$$\lambda_i^0 = -\frac{f_i}{a_i}, \quad \forall i \text{ como valor inicial, y}$$

$$\lambda_i^k = -\frac{f_i}{\sum_j x_{ij}}, \quad \forall i \text{ para los valores subsecuentes;}$$

y el subproblema primal (SP) que también es un problema de transporte, cuya estructura es:

$$\text{Minimizar} \quad \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} + \sum_{i=1}^m f_i y_i,$$

sujeto a

$$\sum_{i=1}^m x_{ij} = 1, \quad \forall j$$

$$\sum_{j=1}^n d_j x_{ij} \leq a_i y_i, \quad \forall i$$

$$0 \leq x_{ij} \leq y_i, \quad \forall i, j.$$

Las etapas del algoritmo de Descomposición Cruzada Separable son:

- i) fijar y_i es su valor actual y resolver el subproblema de Benders SP_{y_i} , que es un problema de transporte, para generar un nuevo valor para la cota superior de P ;
- ii) fijar λ_i en su valor actual y resolver el subproblema Lagrangeano SD_{λ_i} , para generar un nuevo valor para la cota inferior de P ;

las cuales producen sucesivamente valores de las cotas inferior y superior, respectivamente, para la solución del problema AVD de la segunda fase.

El tener que resolver únicamente problemas de transporte, es una situación en extremo ventajosa, tomando en cuenta la baja complejidad de los algoritmos de solución de este tipo de problemas, $O(n^3)$, y la existencia de numeroso software comercial para resolverlos.

La tercera fase de la estrategia propuesta completa el ruteo de vehículos, indicando la secuencia en que cada vehículo debe visitar a los clientes que le han sido asignados, lo cual implica la solución de problemas de agente viajero. El algoritmo aplicado en esta fase es la heurística híbrida Dos Optimal-Prim, que mostró excelentes aproximaciones al óptimo, para el tamaño de problemas manejado.

De tal forma que, se logró obtener una estrategia fácil de implementar, que da resultados de buena calidad, y que es eficiente en esfuerzo computacional aun en problemas grandes, para obtener la solución de un problema que en esencia es NP-duro, objetivo general del presente trabajo de investigación.

Para evaluar la eficiencia de la estrategia propuesta, se obtuvo su orden de complejidad, que resultó ser $O(n^3)$. Esto hace que la eficiencia de la estrategia esté muy por encima de los métodos utilizados por el IRP, que se revisaron en el estado del arte. También se comparó con los desarrollos que resolvían instancias del IRP que coincidían con el tratado en esta investigación; esta comparación dio como resultado la superioridad de la estrategia propuesta, al manejar problemas de mayor tamaño, y con menor esfuerzo computacional.

Por otro lado, para evaluar el desempeño de la estrategia, se resolvieron nueve problemas de diferentes tamaños: 4,5,7,8,30,35,150,200 y 250 clientes, generados aleatoriamente; obteniendo como resultado que el tiempo de solución es extremadamente pequeño, de hecho, aun para el problema de 250 clientes el tiempo requerido fue de tres segundos. En las primeras seis instancias se obtuvo el óptimo.

Considerando los argumentos anteriores, podemos afirmar que, tanto el objetivo general como los objetivos específicos de este trabajo de investigación fueron alcanzados.

El enfoque de la optimización de sistemas, como una posibilidad para la competitividad de las organizaciones, debe ser una directriz en la cual es necesario que se actúe, tomando en cuenta los retos económicos, sociales, políticos, tecnológicos, a los cuales se enfrentan los tomadores de decisiones. La competitividad es un arma con la cual todo tipo de organizaciones y empresas pueden continuar vigentes y mantenerse exitosas en sus ámbitos de acción, logrando su desarrollo económico.

Es posible utilizar paquetes de optimización para hacer interfaces con sistemas específicos, que implementen el desarrollo de y/o mejora de métodos de solución de problemas. Esta situación hace que los usuarios finales aprovechen el potencial de dichos métodos, sin requerir conocer su fundamento teórico. En otras palabras, es posible facilitar el uso de las técnicas y teorías matemáticas, trabajándolas como cajas negras para el usuario, siendo éste el principal beneficiario.

Existen en el mercado varios sistemas para el manejo de inventarios y ruteo, sin embargo, resultan extremadamente costosos, además de excesivos para las necesidades que las micro y pequeñas empresas puedan tener. De tal forma, la construcción de soluciones a la medida de este tipo de empresas, representa un nicho de mercado muy importante que es factible de explotar, pudiendo presentar situaciones redituables muy convenientes pues además, la estrategia propuesta es fácilmente implementable en equipo de cómputo de bajo costo y la información requerida para su funcionamiento es de fácil obtención.

2. Extensiones.

- ☞ Como continuación del presente trabajo se encuentra el análisis del desempeño de la estrategia propuesta con otras instancias del problema.
- ☞ Alguna investigación futura del IRP podría tomar en consideración el análisis de redes más complejas, por ejemplo, redes en cuya selección de rutas esté considerado que el costo de transportación varía a medida que las entregas son realizadas. Algoritmos que consideren explícitamente la localización de varios clientes y varios almacenes, así como el manejo de escenarios dinámicos más completos.
- ☞ Un interesante tópico es la real evaluación en el uso de información en tiempo real, mediante el uso de dispositivos EDI, EPC, Localización Vía Satélite, etc., ya que juegan un papel importante en la configuración de los sistemas.
- ☞ Es posible analizar la inclusión de puntos de transbordo, su número y localizaciones óptimas, considerar la restricción de distancia total recorrida por el vehículo, que está relacionada con el número de horas que le está permitido trabajar a un chofer.
- ☞ En otras circunstancias, restricciones de ventanas de tiempo deben ser tomadas en cuenta, *i.e.*, restricciones que imponen cotas inferiores y superiores en el tiempo de arribo de los vehículos al lugar en que se encuentran los clientes.
- ☞ También es posible lograr avances en métodos de validación que podrían ser requeridos al desarrollar técnicas heurísticas, dada la dificultad de encontrar soluciones óptimas.
- ☞ Mayor investigación es necesaria para la consideración explícita de múltiples productos en el sistema, línea de trabajo de Speranza y Ukovich (1994b), que combina la localización de productos dentro de camiones en el IRP, y que podría ser una contribución significativa al campo.

- ☞ Es importante el manejo de escenarios en los que existan envíos de emergencia, considerando sistemas de transporte restringidos.
- ☞ En muchas aplicaciones prácticas, existe un componente dinámico significativo. Por ejemplo, el caso en que un distribuidor entregue un producto desde un almacén central a clientes, cuya localización geográfica varía en el tiempo.
- ☞ El análisis de diferentes instancias del IRP, bajo consideraciones de demanda estocástica, es un campo de trabajo abierto y muy poco explorado.

Apéndice A

Complejidad Computacional.

La complejidad computacional estudia:

- la eficiencia de los algoritmos y
- la "dificultad" inherente de problemas de importancia práctica y/o teórica.

El principal descubrimiento en esta área fue que los problemas computacionales pueden variar tremendamente en cuanto al esfuerzo requerido para resolverlos de forma precisa. Como se indicó, el IRP es un problema NP-duro. El término técnico para un problema duro es "NP-completo", lo cual esencialmente significa: "abandonar toda la esperanza de definir un algoritmo eficiente para la solución exacta de ese problema". Sin embargo, conociendo que un problema es NP-completo y así sus limitaciones, los expertos no gastan el tiempo en proyectos imposibles y en lugar de ello aceptan aproximaciones menos ambiciosas, por ejemplo, definir soluciones aproximadas, para resolver casos especiales o alterar un poco el problema para transformarlo en un problema menor que sea tratable.

Al hablar de eficiencia de los algoritmos, los rápidos avances en la teoría de las computadoras hacen medidas físicas (tiempo de corrida, requerimientos de memoria) irrelevantes. Una medida más estandarizada es el número de operaciones elementales de cómputo que toma el resolver un problema en el peor caso. El desempeño promedio no es seguro. Puede haber casos particulares (técnicamente llamados instancias de un problema) que se comportan mucho peor que el promedio y nadie puede dar confianza en la suerte de intentar resolverlo.

El número de operaciones elementales obviamente depende del tamaño de los datos de problema. Clasificar 1 billón de números es completamente diferente de clasificar 10, de tal forma, expresamos el número de operaciones elementales (suponiendo el escenario del peor caso) como una función de algún o algunos números característicos suficientes, para calcular el volumen del trabajo que será realizado. Para un algoritmo de clasificación, este número es simplemente el número n de números que serán clasificados.

Ahora, supongamos un algoritmo que resuelve un problema de tamaño n en un máximo de $7n^3 + 5n^2 + 27$ operaciones.

Para tal función, estamos primeramente interesados en su tasa de crecimiento conforme n se incrementa. Queremos distinguir entre tasas de crecimiento suaves y "explosivas", por lo tanto, diferencias tales como $7n^3$ y n^3 no son realmente importantes (además de que grandes diferencias en las constantes no surgen en la práctica). Podemos también descartar los términos de orden bajo, debido a que es el mayor grado quien determina la tasa de crecimiento.

El resultado final es que la complejidad de este algoritmo puede ser suficientemente descrita por la función $g(n) = n^3$. Formalmente, decimos que este algoritmo es "de orden $O(n^3)$ ". Es también usual decir que este algoritmo "toma un tiempo $O(n^3)$ ". Este simbolismo es un recordatorio de que esta función representa el comportamiento en el peor caso para tamaños suficientemente grandes.

Notemos que para n suficientemente grande: $\log n < n < n^2 < n^3 < 2^n$.

Esto a veces se establece como :

$$O(\log n) < O(n) < O(n \log n) < O(n^2) < O(n^3) < O(2^n).$$

El IRP generaliza al TSP, el cual es un bien conocido problema de Optimización Combinatoria, y presenta una notoria dificultad al tratar de resolverse de manera exacta. Sin embargo su formulación es muy simple: "Dadas las coordenadas de n ciudades, defina el tour cerrado más corto que visita a cada ciudad exactamente una vez". Este problema ha servido como un punto de referencia para casi cada nueva idea algorítmica y fue uno de los primeros problemas de optimización establecidos, que después mostró ser NP-completo. De hecho, representa el problema en el que nuevas ideas algorítmicas para la solución de los Problemas de Ruteo de Vehículos (VRP), son experimentados.

Algoritmos de tiempo polinomial y exponencial.

Aquellos algoritmos que tienen una complejidad de tiempo polinomial o subpolinomial (esto es, aquellos que toman un tiempo $O(g(n))$ donde $g(n)$ es un polinomio o una función acotada por un polinomio), son prácticos.

Los algoritmos que tienen tiempos de corrida de orden $O(\log n)$, $O(n)$, $O(n \log n)$, $O(n^2)$, $O(n^3)$, etc., son llamados algoritmos de tiempo polinomial.

Por otro lado, algoritmos de complejidades que no se pueden acotar por funciones polinomiales son llamados algoritmos de tiempo exponencial. Estos incluyen órdenes de crecimiento explosivo los cuales contienen también factores no exponenciales como $n!$.

Hay varios argumentos que soportan la tesis de que “polinomial es un sinónimo para práctico”: primero, los algoritmos de tiempo polinomial están en una mucho mejor posición para explotar mejoras tecnológicas de velocidad para computadoras. Esto puede mostrarse mejor en la siguiente tabla:

Función de complejidad de tiempo	Tamaño de la mayor instancia del problema solucionable en 1 hora		
	Con una computadora actual	Con una computadora 100 veces más rápida	Con una computadora 1000 veces más rápida
n	$N1$	$100N1$	$1000N1$
n^2	$N2$	$10N2$	$31.6N2$
n^3	$N3$	$4.64N3$	$10N3$
2^n	$N4$	$N4 + 6.64$	$N4 + 9.97$
3^n	$N5$	$N5 + 4.19$	$N5 + 6.29$

Por ejemplo, en el caso de $O(n^2)$:

una computadora actual corre para an^2 segundos, $1hr. = aN_2^2$,

una computadora 100 veces más rápida corre para $an^2 / 100$,

$1hr. = aN_x^2 / 100$, por lo tanto $100N_2^2 = N_x^2$, $N_x = 10N_2$.

Y en el caso de $O(2^n)$:

una computadora actual corre para $a2^n$ segundos, $1hr. = a2^{N_4}$,

una computadora 100 veces más rápida corre para $a2^n / 100$ segundos,

$1hr. = a2^{N_x} / 100$, por lo tanto

$100(2^{N_4}) = 2^{N_x}$, $N_x = N_4 + (\log 100 / \log 2) = N_4 + 6.64$.

Como puede verse, la mejora en tecnología es multiplicativa en algoritmos de tiempo polinomial y únicamente aditiva en algoritmos de tiempo exponencial. La situación es mucho peor que la mostrada en la tabla si las complejidades involucran factoriales.

Un segundo argumento es que, en algoritmos de tiempo polinomial, una vez que son descubiertos, pasan por una serie de mejoras, esto es, reducciones en las constantes en sus funciones de complejidad y en sus propios grados. Este es un resultado empírico y, en parte, explica por qué complejidades como $O(n^{80})$ u $O(10^{100}n)$ no aparecen en la práctica. Los algoritmos de tiempo polinomial típicamente tienen un grado de entre 2 o 3, y no incluyen coeficientes muy grandes. La brecha real para la solución de un problema, está en la definición del primer algoritmo de tiempo polinomial. Para problemas combinatorios, el “salto” a la clase polinomial requiere comúnmente un análisis profundo de la naturaleza del problema (si es que el salto puede hacerse).

Por supuesto hay excepciones a estas reglas. El caso más famoso es el excelente desempeño del Método Simplex, el cual siempre ha trabajado muy bien en la práctica, si bien es probablemente exponencial (hay casos artificiales "patológicos" donde el Simplex muestra su comportamiento exponencial oculto). Podría suceder que el método captura alguna propiedad significativa del problema que aún debe ser descubierta.

La conclusión general es que un problema puede considerarse "eficientemente resuelto" cuando se ha encontrado un algoritmo de tiempo polinomial para su solución.

Algunos ejemplos de problemas de optimización que son considerados "bien resueltos" en este sentido son:

- Programación Lineal (existen algoritmos polinomiales)
- Algunos problemas de Ruta más Corta
- Algunos problemas de Ruta más Larga (por ejemplo los cálculos de PERT/CPM simple)
- El problema de Árbol de Mínima Expansión
- Algunos problema de flujo de redes
- Problemas de Pareamiento
- Problemas de Transporte, Asignación y Transbordo
- Varios casos especiales de problemas duros

Esta lista, a pesar de todos los problemas que incluye, es muy corta comparada con la de problemas para los cuales, los algoritmos de tiempo polinomial no son conocidos.

Problemas de Decisión y el Problema de Factibilidad.

Un marco contextual apropiado para el análisis de problemas es el de los problemas de optimización, esto es, minimizar o maximizar una función (la función objetivo), sujeta a una o más restricciones.

Una clase especial de problemas de optimización requiere únicamente de una respuesta "sí" o "no". Estos son llamados problemas de decisión. La teoría de Complejidad Computacional implica estricta atención en estos problemas por su uniformidad. Los resultados pueden ser fácilmente generalizados ya que:

1. para cada problema de optimización, existe una versión de decisión,
2. cualquier resultado de complejidad para la versión de decisión se mantiene también para el problema original, esto es, el problema original no es "mucho más duro" que la versión de decisión.

Se sabe que dado un problema de optimización, se define un problema de decisión asociado a él, esto es, una pregunta que puede ser contestada sí o no. Por otro lado, varios problemas computacionales bien conocidos son problemas de decisión. Además, se ha puntualizado que un problema de optimización original no es más difícil que el problema de decisión, cualquier

resultado probado sobre la complejidad del problema de decisión será aplicable también al problema de optimización.

Llamemos a $\pi(*)$, "la versión de decisión del problema * ". La definición de problema de decisión a partir del problema de optimización, permite estudiar ambos tipos de problemas de manera uniforme.

De central importancia en el contexto son los problemas de factibilidad. Si para algún problema de decisión existe un conjunto de valores "sí" - "no" que asignados a las variables componen un resultado "sí", entonces se tiene un problema factible (problema de factibilidad).

Algoritmos no-determinísticos.

Un algoritmo no-determinístico es una herramienta teórica que no existe en la realidad. Es como un algoritmo ordinario, con la excepción de que tenemos permitido usar la siguiente instrucción imposible:

goto **both** label 1, label 2

Esto divide el cómputo de dos procesos paralelos en un solo paso en el mismo procesador (no en procesamiento paralelo).

Para tener una idea del poder de cómputo no-determinístico, consideremos la conocida versión del problema de Programación Entera 0-1 que es un problema "duro" (no se conoce un algoritmo determinístico para él):

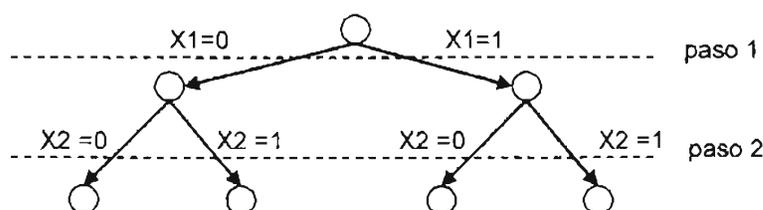
Dada una matriz A de $m \times n$ y un m -vector entero b , ¿existe un n -vector x , con elementos 0-1, tal que $Ax=b$?

Esta poderosa instrucción permite resolver el problema en tiempo polinomial, precisamente por medio de la exploración de todas las posibles combinaciones de valores de los elementos del vector x :

```

begin
  for j=1,...,n
    goto both A,B
    A:  $x_j = 0$ 
    goto again
    B:  $x_j = 1$ 
    : again
  next j
  if  $x=(x_1, \dots, x_n)$  satisfies  $Ax=b$  then output "yes" else output "no"
end

```



De hecho, si tuviésemos computadoras capaces de ejecutar esta instrucción extra, no sería necesaria la teoría de Complejidad Computacional.

Clases P y NP.

Estamos ahora en posición de definir formalmente dos clases importantes de problemas de decisión, las clases P y NP.

La clase P consiste de todos aquellos problemas de decisión para los cuales existe un algoritmo de tiempo polinomial.

Para la clase NP, simplemente requerimos que alguna respuesta "sí" sea "fácilmente" verificable. Tanto la codificación de la respuesta, como el tiempo que toma verificar su validez deben ser "cortos", esto es, tiempo polinomial o polinomialmente acotado. Formalmente decimos que cualquier instancia "sí" del problema tiene la propiedad de "pertenencia".

NP se establece para "Polinomial No-determinístico", debida a una alternativa (y equivalente) definición basada en la noción de algoritmos no-determinísticos. Una observación de la definición anterior es que: con la finalidad de mostrar que un problema pertenece a la clase NP, no es necesario mostrar de qué manera esta respuesta, fácilmente verificable, puede ser producida, en otras palabras, la solución del problema, es suficiente con mostrar que tal estado de pertenencia existe.

Para clarificar lo anterior, consideremos el Problema del Agente Viajero (TSP), que dice: "dado un número n de ciudades, $n \geq 3$ y entero, una matriz $n \times n$ de distancias enteras no negativa $C = [c_{ij}]$, y un entero no negativo L : "¿existe un tour cerrado que pasa exactamente una vez por cada ciudad, con una longitud total $\leq L$?".

Si imaginamos que tenemos una respuesta "sí", ésta es un orden de ciudades, digamos $c_1, c_2, c_3, \dots, c_n$, tal que la suma de las ligas $(c_1, c_2), (c_2, c_3), \dots, (c_n, c_1)$ es menor o igual a L . La forma en que se obtuvo esta solución es irrelevante. Es claro que podemos verificar fácilmente la distancia del tour, así que sabemos que $\pi(\text{TSP}) \in \text{NP}$. Por otro lado, no conocemos si el problema está también en la clase P, debido a que ningún algoritmo polinomial ha sido aún encontrado para él.

Por lo tanto, la clase NP consiste de todos los problemas "razonables" de importancia teórica y/o práctica. Para problemas que no están en la clase NP, la propia verificación de que una solución es válida puede ser extremadamente difícil.

Evidentemente, los problemas "fáciles" de la clase P son también "razonables", esto es, ellos también pertenecen a NP. Esto es P es un subconjunto de NP ($P \subseteq \text{NP}$).

La cuestión importante es ahora si P es un subconjunto propio de NP , esto es, $P \subset NP$. Esta es una conjetura muy importante en Complejidad Computacional, en otras palabras dice: "no todos los problemas pueden resolverse eficientemente". Aunque no existe la prueba formal aún, existe una fuerte evidencia para la validez de esta conjetura.

Reducciones y transformaciones polinomiales.

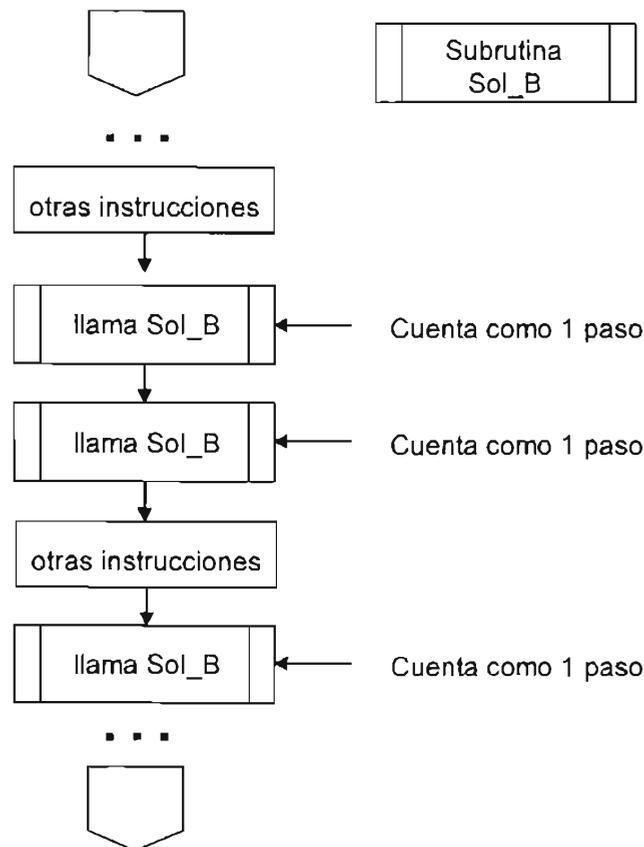
Las herramientas básicas para relacionar las complejidades de varios problemas son las reducciones y transformaciones.

Decimos que un problema A se reduce en tiempo polinomial a otro problema B , si y sólo si:

1. hay un algoritmo para A que usa una subrutina para B ,
2. cada llamado a la subrutina para B cuenta como un único paso, y
3. el algoritmo para A corre en tiempo polinomial.

Lo anterior implica que la subrutina para B puede ser llamada a lo más un número polinomialmente acotado de veces.

Algoritmo para resolver A



El algoritmo para resolver A es polinomial $\Rightarrow A$ se reduce polinomialmente a B

La implicación práctica viene de la siguiente proposición:

Si A reduce polinomialmente a B y existe un algoritmo de tiempo polinomial para B, entonces hay un algoritmo de tiempo polinomial para A.

La prueba está basada en el hecho de que la composición de dos polinomiales es polinomial.

Los tres casos básicos son los siguientes:

- (A reduce a B) y (B es "fácil") \Rightarrow A es fácil
- (A reduce a B) y (A es "duro") \Rightarrow B es duro (debido a que si B fuera fácil A sería duro y eso es una contradicción)
- (A reduce a B) y (B es "duro") \Rightarrow (no hay conclusión para A, un caso muy común)

Ahora, supongamos que representamos la "dureza" de un problema con una variable binaria, esto es, $d=1$ si el problema es "duro" y $d=0$ si el problema es "fácil".

Entonces, "A reduce polinomialmente a B" puede expresarse con $d_B \geq d_A$. Desde luego:

- ($d_B \geq d_A$) y ($d_B = 0$) $\Rightarrow d_A = 0$
- ($d_B \geq d_A$) y ($d_A = 1$) $\Rightarrow d_B = 1$
- ($d_B \geq d_A$) y ($d_B = 1$) $\Rightarrow d_A$ puede tomar cualquier valor

De aquí, podemos justificar el decir que "B es al menos tan duro como A":

Si A reduce polinomialmente a B, entonces es posible implicar que A puede verse como un caso especial de B y en consecuencia, "B es al menos tan duro como A".

Por ejemplo, el problema de TSP reduce a $\pi(\text{TSP})$, así, esta versión es al menos tan dura como el problema original TSP.

Un caso especial de reducciones polinomiales es la siguiente: el algoritmo para A, primero construye las entradas para la subrutina hipotética B, entonces llama a ésta última exactamente una vez, y finalmente regresa la respuesta al algoritmo para A.

Las reducciones de este tipo son llamadas transformaciones polinomiales (decimos que un problema A se transforma en tiempo polinomial al problema B, y se usa la notación $A \propto B$). Son especialmente útiles debido a que la relación de transformabilidad polinomial es transitiva y ésto es fácilmente probado usando de nuevo el hecho de que la composición de dos polinomiales es polinomial.

Problemas NP-completos.

Si $A \leq B$ para todo $A \in NP$, entonces algún problema perteneciente a la clase NP puede ser visto como un caso especial de B y llamarlo NP-Duro. Dicho de otra forma, otro C es una problema NP-duro, si pertenece a NP y puede ser visto como un caso especial de B.

Decimos que un problema es NP-completo si:

- i. pertenece a la clase NP y
- ii. todos los demás problemas en NP se transforman polinomialmente a él.

Un problema NP-completo tiene la siguiente propiedad importante:

Si existe un algoritmo eficiente (esto es, polinomial) para algún problema NP-completo, entonces existe un algoritmo eficiente para cada problema en NP.

Lo anterior debido a que, si decimos que P^* es este problema:

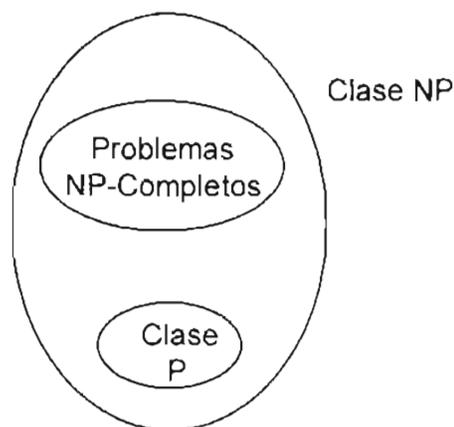
todos los problemas en NP se transforman polinomialmente a P^* y dado que la transformación polinomial es un caso especial de reducibilidad polinomial, se aplica el caso 1:

(todos los problemas en NP se reducen a P^*) y (P^* es "fácil") \Rightarrow todos los problemas en NP son fáciles.

Por esta razón, los problemas NP-completos son el principal objetivo en la búsqueda de algoritmos eficientes.

Además, dado que $A \leq B$, $B \in NP$ y A es NP-completo, entonces B es también NP-completo.

El siguiente es un mapa del ámbito NP:



Esta es una sobre-simplificación del conocimiento actual en el campo.

Complejidad del Problema de Inventario-Distribución (IRP).

Actualmente no se conoce si el IRP pertenece o no a la clase P. Sin embargo, es posible considerar que pertenece a la clase NP.

Se afirma lo anterior, ya que es posible asignar valores 0 ó 1 a las variables x_{ijk} y y_{ik} en la formulación del problema entero del IRP, de manera que todas las restricciones sean satisfechas y el valor de la función objetivo no exceda de un cierto valor g .

Al asociar un problema de decisión al IRP, se tiene un vector 0-1 con $n^2k + nk$ elementos, que representan los valores de las $n^2k + nk$ variables x_{ijk} y y_{ik} . Entonces la conversión binaria del IRP es de longitud $n^2k + nk$ que tiene el mismo orden de complejidad que la versión de decisión. Así, para afirmar la factibilidad de una combinación de valores 0-1, esto es, que se cumplen las restricciones y que el valor de la función objetivo no excede a g , se requieren cálculos del orden de $n^2k + nk$. Dado que la longitud de la versión de decisión tiene un orden de complejidad polinomial y el tiempo necesario para verificar la factibilidad también, se confirma que la $\pi(\text{IRP}) \in \text{NP}$.

Es necesario determinar que el IRP es NP-Duro. Para ello consideremos primeramente que el componente de ruteo de vehículos del problema, VRP, es NP-duro (Cousineau-Ouimet, 2002 y Garey y Johnson, 1979); ahora este problema además incluye restricciones de inventario. Entonces decimos que VRP reduce a IRP, ya que al resolver el IRP se resuelve el VRP e IRP es al menos tan difícil como VRP.

Entonces, si VRP reduce a IRP y VRP es duro, entonces IRP es duro, *i.e.*, por el Caso 1 :

$$(d_{\text{IRP}} \geq d_{\text{VRP}}) \text{ y } (d_{\text{VRP}}=1) \Rightarrow d_{\text{IRP}}=1.$$

Dado lo anterior, se puede afirmar que el IRP es NP-duro. De hecho, todos los problemas no triviales de ruteo de vehículos son NP-duros (Bramel y Simchi-Levi, 1995).

Apéndice B

Principios de Descomposición de Benders y Descomposición Dual Lagrangeana.

Descomposición de Benders.

Sea el siguiente problema de MIP:

$$\begin{array}{ll} \text{Minimizar} & Cx \\ & x \in S \\ \text{s.a:} & \\ & Ax \geq b \\ & x \geq 0 \end{array} \quad (P)$$

donde:

x es un vector n -dimensional de valores reales,
 A es una matriz de orden $m \times n$,
 C y b son vectores con dimensiones conformables.

Sea S un subconjunto de R^n restringiendo algunos elementos de x a ser enteros.

Cuando la matriz A es particionada con $n = n_1 + n_2$ y $m = m_1 + m_2$, podemos escribir al problema como:

$$\begin{array}{ll} \text{Minimizar} & C^1 x_1 + C^2 x_2 \\ & x \in S \\ \text{s.a:} & \\ & A_1^1 x_1 + A_1^2 x_2 \geq b_1 \\ & A_2^1 x_1 + A_2^2 x_2 \geq b_2 \\ & x_1 \geq 0, x_2 \geq 0 \end{array} \quad (P)$$

donde

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = [A^1, A^2] = \begin{bmatrix} A_1^1 & A_1^2 \\ A_2^1 & A_2^2 \end{bmatrix}$$

con

A_1 matriz de orden $m_1 \times n$, A_2 matriz de orden $m_2 \times n$,

A^1 matriz de orden $m \times n_1$, A^2 matriz de orden $m \times n_2$,

A_j^i matriz de orden $m_i \times n_j$ con $i = 1, 2$ y $j = 1, 2$.

Si se particiona A de forma que x_2 y $A_2 x \geq b_2$ sean las variables y restricciones complicadas (las que requieren ser enteras), entonces si fijamos x_2 (a valores enteros), o relajamos $A_2 x \geq b_2$, se obtienen como resultado problemas más fáciles de resolver. Si definimos a S como

$$S = \{x = (x_1, x_2) \mid x_1 \geq 0, x_2 \geq 0, x_2 \in Z\}$$

y suponemos que el conjunto

$$\{x \mid x \in S, A_1 x \geq b_1\}$$

es no vacío y acotado; entonces fijando $x_2 \in Z$ vamos a tener que lo que se debe hacer es

$$\text{Min}_{x_2 \in Z} \left\{ \begin{array}{l} \text{Min}_{x_1 \geq 0} C^1 x_1 + C^2 x_2 \\ \text{s.a:} \\ A_1^1 x_1 \geq b_1 - A_1^2 x_2 \\ A_2^1 x_2 \geq b_2 - A_2^2 x_2 \end{array} \right\}$$

Obteniendo el dual del problema interior se tiene:

$$\text{Min}_{x_2 \in Z} \left\{ \begin{array}{l} \text{Max } u_1(b_1 - A_1^2 x_2) + u_2(b_2 - A_2^2 x_2) + C^2 x_2 \\ \text{s.a:} \\ u_1 A_1^1 + u_2 A_2^1 \leq C^1 \\ u_1 \geq 0, u_2 \geq 0 \end{array} \right\}$$

o si definimos a T_{PA} como la región factible del dual del problema interior y a l como sus puntos extremos, entonces se puede escribir

$$\text{Min}_{x_2 \in Z} \text{Sup} \{u_{1l}(b_1 - A_1^2 x_2) + u_{2l}(b_2 - A_2^2 x_2) + C^2 x_2; l \in T_{PA}\}$$

lo que indica que se obtiene el mínimo (para las $x_2 \in Z$, i.e., fijas en enteros), del supremo de la función objetivo, pero sólo en los puntos extremos de la región factible del dual, evidentemente, eso hará que se reduzcan los puntos posibles y así el esfuerzo al resolver (P).

Ahora, el problema anterior se puede escribir como

$$\left. \begin{array}{l} \text{Min } x_0 \\ x_2 \in Z, x_0 \\ \text{s.a:} \\ u_{1t}(b_1 - A_1^2 x_2) + u_{2t}(b_2 - A_2^2 x_2) + C^2 x_2 \leq x_0; t \in T_{PA} \end{array} \right\} (MP)$$

que es llamado problema maestro primal, es un problema de MIP con una sola variable continua que es x_0 , cuyas restricciones son llamados cortes primales o de Benders.

Asociado hay un subproblema primal o de Benders, (SP_x) que es

$$\begin{array}{l} \text{Min } C^1 x_1 + C^2 x_2 \\ x_1 \geq 0 \\ \text{s.a:} \\ A_1^1 x_1 \geq b_1 - A_1^2 x_2 \\ A_2^1 x_1 \geq b_2 - A_2^2 x_2 \end{array}$$

que mantiene un valor fijo de x_2 y donde $u_t = (u_{1t}, u_{2t})$ para $t \in T_{PA}$ son los puntos extremos de la región factible del dual de (SP_x) .

El algoritmo de descomposición de Benders itera entre el denominado subproblema y el problema maestro. El problema maestro proporciona los valores de la variable fija, la cual es utilizada en el subproblema. Cada vez que el subproblema es resuelto, se genera una nueva solución extrema, la cual se utiliza para construir un corte de Benders que se incluye en el problema maestro. El número de cortes se incrementa en cada iteración y es posible mostrar que el subproblema, da el corte que es maximal en el valor que tiene la variable fija; es decir, dará el corte más violado en cada iteración. De esta forma, se generan los cortes necesarios y el algoritmo se detiene en un número finito de iteraciones, después de generar, en el peor de los casos, todos los cortes. Con lo cual la solución óptima se encuentra.

Descomposición Dual-Lagrangeana (Dantzig-Wolfe).

Sea nuevamente el problema (P) planteado originalmente como:

$$\begin{aligned}
& \underset{x \in S}{\text{Minimizar}} && C^1 x_1 + C^2 x_2 \\
& \text{s.a:} && \\
& && A_1^1 x_1 + A_1^2 x_2 \geq b_1 \\
& && A_2^1 x_1 + A_2^2 x_2 \geq b_2 \\
& && x_1 \geq 0, x_2 \geq 0
\end{aligned} \tag{P}$$

Paralelamente a la aproximación de Benders, es posible obtener el proceso de descomposición dual por el método de Lagrange. De esta forma, el problema dual Lagrangeano de (P) con respecto a la restricción $A_2^1 x_1 + A_2^2 x_2 \geq b_2$ puede establecerse como

$$\underset{u_2 \geq 0}{\text{Max}} \left\{ \begin{array}{l} \underset{x \in S}{\text{Min}} C^1 x_1 + C^2 x_2 + u_2 (b_2 - A_2^1 x_1 - A_2^2 x_2) \\ \text{s.a:} \\ A_1^1 x_1 + A_1^2 x_2 \geq b_1 \end{array} \right\}$$

el cual puede simplificarse como

$$\underset{u_2 \geq 0}{\text{Max}} \left\{ \begin{array}{l} \underset{x \in S}{\text{Min}} Cx + u_2 (b_2 - A_2 x) \\ \text{s.a:} \\ A_1 x \geq b_1 \end{array} \right\}$$

y escribirse como

$$\underset{u_2 \geq 0}{\text{Max}} \inf \{ C^2 x_t + u_2 (b_2 - A_2 x_t); t \in T_D \}$$

donde t para $t \in T_D$, son los puntos extremos del contorno convexo del subproblema dual. Entonces el problema dual de (P) se puede establecer como

$$\underset{u_2 \geq 0, u_0}{\text{Max}} \left. \begin{array}{l} u_0 \\ \text{s.a:} \\ Cx_t + u_2 (b_2 - A_2 x_t) \geq u_0, t \in T_D \end{array} \right\} \tag{MD}$$

que resulta ser el problema maestro dual Lagrangeano, (MD). Y el correspondiente subproblema dual queda como

$$\left. \begin{array}{l} \text{Min}_{x \in S} Cx + u_2(b_2 - A_2x) \\ \text{s.a:} \\ A_1x \geq b_1 \end{array} \right\} (SD u_2)$$

El subproblema dual $(SD u_2)$ es comúnmente llamado el problema Lagrangeano, las restricciones del problema maestro dual (MD) son los cortes duales.

Teorema 1. Considere al problema original (P) escrito como

$$\begin{array}{ll} \text{Minimizar} & Cx \\ \text{s.a:} & \\ & A_1x_1 \geq b_1 \\ & A_2x_2 \geq b_2 \\ & x_1 \geq 0, x_2 \geq 0 \text{ y entera} \end{array} \quad (P)$$

Sea λ un vector no negativo y conformable denominado "multiplicador de Lagrange", entonces, podemos escribir el problema como:

$$\begin{array}{ll} \text{Min} & Cx + \lambda(b_2 - A_2x_2) \\ \text{s.a:} & \\ & A_1x_1 \geq b_1 \\ & x_1 \geq 0, \lambda \geq 0 \end{array} \quad (LR_\lambda)$$

Por conveniencia, supongamos que (P) es factible y que el conjunto $S = \{x_1 | A_1x_1 \geq b_1, x_1 \geq 0\}$ de soluciones factibles es finito y acotado. Entonces si existe un $\lambda \neq 0$, tal que, la solución óptima con (x_1^*, x_2^*) hace que $b_2 = A_2x_2^*$, el problema (P) ha sido resuelto.

Prueba. Sean (x_1^*, x_2^*) la solución mínima para el problema (LR_λ) , entonces se tiene que

$$Cx^* + \lambda(b_2 - A_2x_2^*) \leq Cx + \lambda(b_2 - A_2x_2)$$

Por lo tanto, para toda $x_1 \geq 0, x_2 \geq 0$ y entera, se cumple que:

$$Cx^* \leq Cx + \lambda(A_2x_2^* - A_2x_2)$$

Entonces para todas las soluciones no negativas se tiene que, $A_2x_2^* \leq A_2x_2$ y la desigualdad anterior se cumple. Pero para estas soluciones $A_2x_2^* - A_2x_2 \leq 0$, con lo que se tiene que, $Cx^* \leq Cx$ puesto que $\lambda \geq 0$. Entonces, (x_1^*, x_2^*) resuelven al problema (P). *q.q.d.*

El teorema indica que si los multiplicadores $\lambda \geq 0$ son seleccionados de tal forma que $b_2 = A_2x_2^*$, es posible solucionar el problema (P) sin la restricción de desigualdad, con lo que el problema resulta más fácil de resolver. Sin embargo, la dificultad con la relajación Lagrangeana consiste en encontrar los multiplicadores $\lambda \geq 0$ tales que, $A_2x_2^* = b_2$.

Por otro lado, un aspecto interesante es el hecho de que el dual del problema (MD) tiene la forma familiar del problema Dantzig-Wolfe. Considere el problema maestro dual (MD) escrito como:

$$\left. \begin{array}{l} \text{Max } u_0 \\ u_2 \geq 0, u_0 \\ \text{s.a:} \\ Cx_t + u_2(b_2 - A_2x_t) \geq u_0, t \in T_D \end{array} \right\} \text{(MD)}$$

cuyo dual puede establecerse como:

$$\begin{array}{l} \text{Min } Cx \\ x \in S \\ \text{s.a:} \\ A_2x \geq b_2 \end{array}$$

sean $x = \sum_j \lambda_j x_j$, $\sum_j \lambda_j = 1$, $\lambda_j \geq 0$, entonces se tiene que

$$\begin{array}{l} \text{Min } C \sum_j \lambda_j x_j \\ \text{s.a:} \\ A_2 \sum_j \lambda_j x_j \geq b_2 \\ \sum_j \lambda_j = 1 \\ \lambda_j \geq 0 \end{array}$$

y definiendo a

$$p_j = A_2 x_j \quad \text{y} \quad a \quad f_j = C x_j$$

se tiene la forma familiar del problema maestro de Dantzig-Wolfe

$$\begin{aligned} \text{Min} \quad & \sum_j f_j \lambda_j \\ \text{s.a.} \quad & \sum_j p_j \lambda_j = b \\ & \sum_j \lambda_j = 1 \\ & \lambda_j \geq 0. \end{aligned}$$

Apéndice C

Relajación Lagrangeana Separable.

Dado un problema de MIP con dos o más conjuntos de restricciones, tal que los problemas con alguno de esos conjuntos de restricciones relajados son relativamente más fáciles de resolver, es posible definir una relajación Lagrangeana que lo descomponga en dos o más subproblemas, cada uno sobre un conjunto de restricciones. La técnica usada consiste en introducir una o más copias del vector de variables de decisión, usar una de esas copias en cada conjunto de restricciones y dualizar la condición de identidad.

Este nuevo esquema es interesante ya que los subproblemas Lagrangeanos contienen a casi todas las restricciones originales, mientras que la relajación convencional inevitablemente pierde al menos a uno de los conjuntos de restricciones especialmente estructurados.

La Relajación Lagrangeana Separable es un caso especial de Relajación Lagrangeana. Antes de mostrarla se considerarán los siguientes conceptos fundamentales de la relajación para problemas de optimización.

Relajación para problemas de optimización.

Sean los problemas

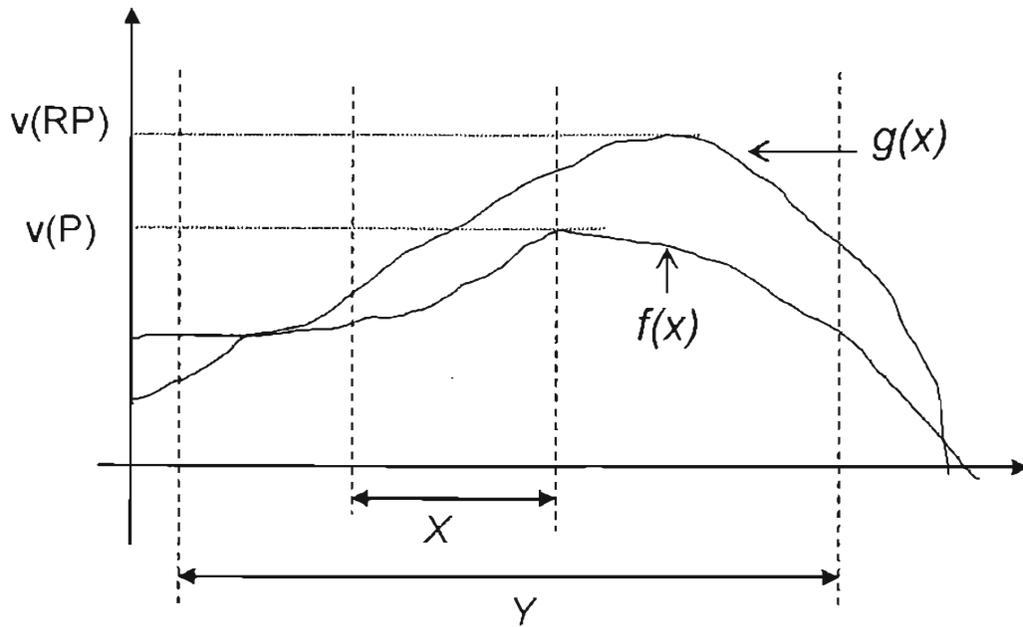
$$(P) \quad \text{Max} \{f(x) | x \in X\} \quad \text{y}$$
$$(RP) \quad \text{Max} \{g(x) | x \in Y\}$$

se dice que (RP) es una relajación del problema (P) si cumple:

- i) $Y \supseteq X$
- ii) $\forall x \in X, \quad g(x) \geq f(x)$

por lo tanto, $v(RP) \geq v(P)$.

Gráficamente el concepto de relajación se puede representar como en la siguiente figura.



Relajación Lagrangeana para Problemas de Optimización.

Sea el problema

$$(P) \quad \text{Max}_x \{fx | Ax \leq b, Cx \leq d, x \in X\}$$

Si se conoce cómo resolver el siguiente problema

$$\text{Max}_x \{fx | Cx \leq d, x \in X\}$$

es posible construir una Relajación Lagrangeana de (P) tal que

$$(LR_\lambda) \quad \text{Max}_x \{fx + \lambda(b - Ax) | Cx \leq d, x \in X\}$$

donde (LR_λ) es una relajación de (P) si cumple con

- i) $FS(LR_\lambda) \supseteq FS(P)$
- ii) $\forall x \in FS(P), \quad fx + \lambda(b - Ax) \geq fx$

donde $FS(\bullet)$ = conjunto factible del problema (\bullet) , por lo tanto,

$$v(LR_\lambda) \geq v(P), \quad \text{para } \forall \lambda \geq 0.$$

Donde (LR) es llamado el dual Lagrangeano tal que

$$v(LR) = \underset{\lambda \geq 0}{\text{Min}} v(LR_\lambda)$$

Relajación Lagrangeana Separable.

Sea el siguiente problema

$$\begin{aligned} & \text{Max } fx \\ & \text{s.a:} \\ & \quad Ax \leq b \quad (P) \\ & \quad Cx \leq d \\ & \quad x \in X \end{aligned}$$

el cual también puede ser expresado como

$$\text{Max } \{fx | Ax \leq b, Cx \leq d, x \in X\} = \text{Max } \{fx | Ax \leq b, Cx \leq d, uAx \leq ub\}$$

para algún $u \geq 0$. Y para algún $Y \supseteq X$, el problema (P) es equivalente a

$$\begin{aligned} & \text{Max } \{fx | Ax \leq b, uAx \leq ub, Cx \leq d, x \in X, y = x, y \in Y\} \\ & \leq \text{Max } \{fx + \lambda(y - x) | Cx \leq d, uAx \leq ub, x \in X, Ay \leq b, y \in Y\} \\ & = \text{Max } \{(f - \lambda)x | Cx \leq d, uAx \leq ub, x \in X\} + \text{Max } \{\lambda y | Ay \leq b, y \in Y\} \end{aligned}$$

Sean P_x, P_y los problemas en x y y respectivamente, y sean también $v(\cdot)$ el valor óptimo del problema (\cdot) , $FS(\cdot)$ su conjunto factible y $OS(\cdot)$ su conjunto óptimo. Entonces

$$v(P) \leq v(\hat{D}) = \underset{u \geq 0, \lambda}{\text{Min}} \{v(P_x) + v(P_y)\} = \underset{u \geq 0, \lambda}{\text{Min}} (LP_u)$$

donde (\hat{D}) es el dual Lagrangeano Separable.

Es posible establecer que la Relajación Lagrangeana Separable es muy ventajosa al tratar con problemas en los cuales no existen las llamadas restricciones complicadas. Debido a que con este esquema ninguna de las restricciones originales desaparece, no es necesario elegir entre la calidad de la cota que se obtiene y el grado de dificultad del problema que queda. (Aceves, 1996).

Por otro lado, también es posible reforzar a los subproblemas que se obtienen con este tipo de descomposición, al adicionar restricciones válidas, tanto al subproblema en x como al subproblema en y , bajo el principio de que una simple restricción no necesariamente complica más al sistema.

Otra de las ventajas de utilizar la estrategia de descomposición Lagrangeana Separable, es la de poder identificar estructuras ocultas especiales, debido a la modelación flexible que tiene, la cual permite obtener una forma de bloque angular en la matriz de restricciones. Situación que no es posible con la relajación Lagrangeana Convencional.

Apéndice D

Comparación de Algoritmos para el Problema del Agente Viajero.

Como se mencionó en el Capítulo 3, en la tesis de maestría elaborada en el año 2000 por Carolina Castañeda Roldán (Universidad de las Américas, Puebla), se realizó un estudio comparativo del Problema del Agente Viajero o TSP (Traveling Salesman Problem) con distintos métodos de solución, tanto exactos como aproximados.

En el trabajo mencionado se implementaron en Java siete técnicas de solución reunidas en el software llamado "Técnicas de Solución al TSP". Ante esto, menciona que puede elegirse entre tres enfoques:

- i) Si el grafo es pequeño se obtiene una solución óptima a través de una búsqueda exhaustiva, evaluando todas las posibilidades. Sin embargo, para grafos grandes este enfoque resulta prohibitivo por su costo, a pesar del grado de eficiencia que se intente dar al software que lo resuelva.
- ii) Buscar información adicional sobre el problema, ya que podría suceder que el grafo tuviera estructuras especiales y que de esta forma no fuese necesario probar todas las posibilidades para hallar una solución óptima.
- iii) El tercer enfoque consiste en modificar un poco el problema y encontrar una solución buena aunque no necesariamente óptima. De hecho, podría ser suficiente con encontrar una solución que determinara un costo cercano al mínimo y la ruta respectiva en grafos pequeños o grandes. Un algoritmo que produce con rapidez soluciones buenas, pero no necesariamente óptimas, se denomina heurístico.

En el trabajo se implementó una técnica de solución exacta, la "Búsqueda Exhaustiva Ingenua". También se exploró el segundo enfoque, aprovechando ciertas propiedades que permitían "podar" las ramas del árbol de búsqueda, al

obtener la solución exacta sin explorar en forma completa el espacio de soluciones, por lo que se seleccionó la técnica de búsqueda "Una Ramificación y Acotamiento Ingenuo" y la técnica "Una Mejor Ramificación y Acotamiento".

Por otro lado, debido a que las técnicas mencionadas tenían un costo alto en tiempo computacional, se implementaron técnicas con el tercer enfoque en las que, sacrificando la exactitud, se pudiera obtener un buen valor aproximado en tiempos relativamente cortos. De tal forma, se seleccionaron las técnicas aproximadas de búsqueda "Dos Optimal", "Adaptación de Prim", "Híbrido Dos Optimal - Prim" y la técnica de Redes Neuronales usando la "Red Neuronal de Hopfield".

La comparación de los algoritmos y el análisis de los resultados del trabajo, es meramente a nivel algorítmico y computacional, pero se realiza un análisis estadístico del comportamiento del software. Las instancias evaluadas del TSP consideraron los siguientes aspectos:

- 1) El Número de ciudades: 3, 4, 50, 100, etc.
- 2) Topología del TSP: simétrico, asimétrico o geométrico. Para la representación del problema TSP se hizo uso de matrices de adyacencia como estructura computacional.
- 3) Límite inferior y superior de los números aleatorios generados: se construyó un generador aleatorio de instancias del TSP para el caso simétrico, asimétrico y geométrico, por lo que se requirió de un límite inferior y superior para los números aleatorios, que representaron las distancias entre las ciudades y que se almacenaron en la respectiva matriz de adyacencia.
- 4) Tipo de solución: las técnicas implementadas fueron "Métodos Exactos", "Métodos Aproximados" o "Todos los Métodos". Dentro de los exactos se seleccionaron "Búsqueda Exhaustiva Ingenua" (Naive Exhaustive-Search), "Ramificación y Acotamiento Ingenuo" (Naive Branch and Bound) y "Una mejor Ramificación y Acotamiento" (A Better Branch and Bound). De los métodos aproximados, los implementados fueron "Dos Optimal", "Adaptación de Prim", "Híbrido Dos Optimal-Prim", y "Red Neuronal de Hopfield".

Las técnicas exactas seleccionadas fueron implementadas para TSP simétrico, asimétrico y geométrico, mientras que las aproximadas sólo lo fueron para TSP simétrico y geométrico debido a la naturaleza de los algoritmos.

Para los fines de este trabajo de investigación, en este apéndice se incluye únicamente la comparación que se hizo entre un algoritmo exacto y algoritmos aproximados, utilizados en la estrategia propuesta para la solución del IRP. La heurística mencionada es el Método Dos Optimal-Prim, que se utiliza en la Fase III de la estrategia.

Comparación de los resultados de un Algoritmo Exacto contra Algoritmos Aproximados.

En el trabajo de Castañeda (2000) se detallan las pruebas realizadas para comparar el comportamiento de algoritmos exactos con aproximados. En este anexo se mostrarán sólo algunos de los resultados. En conclusión, se observará que la diferencia en exactitud de algunos métodos aproximados es pequeño, por lo que en la práctica se sugiere trabajar con cualquiera de estos métodos aproximados, ya que el costo en tiempo no justificó la diferencia.

Se graficaron en diagrama de barras las diferencias en exactitud, ver Fig. 1 a Fig. 4.

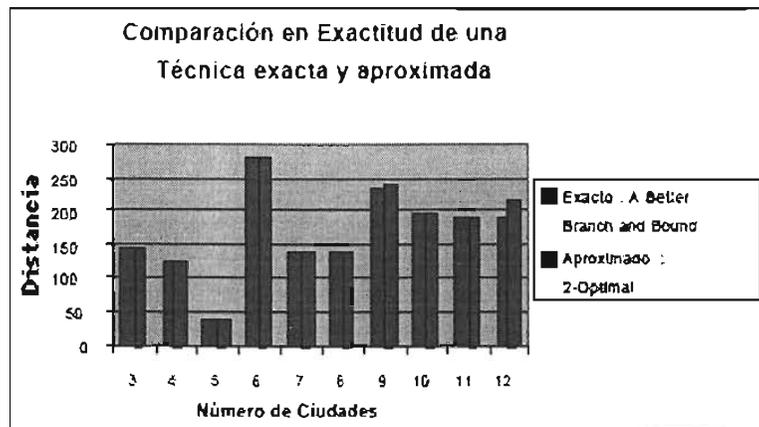


Fig.1 Gráfica de Barras de 2-Optimal.

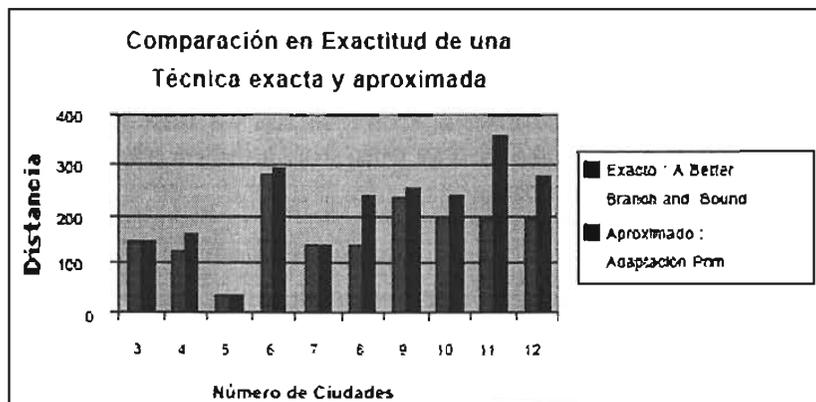


Fig. 2 Gráfica de Barras de Adaptación Prim.

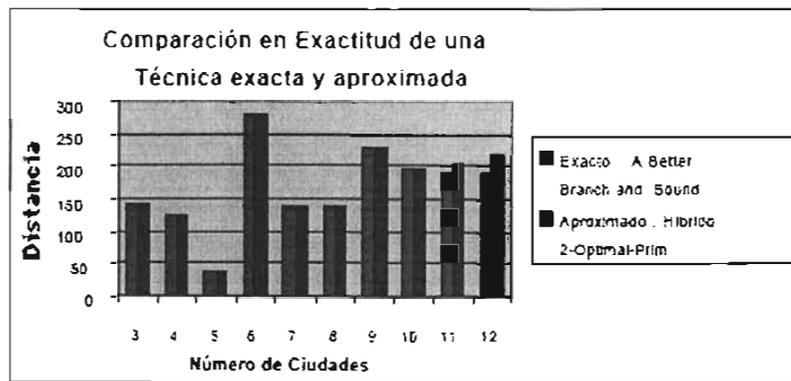


Fig. 3 Gráfica de Barras de Híbrido Dos Optimal-Prim.

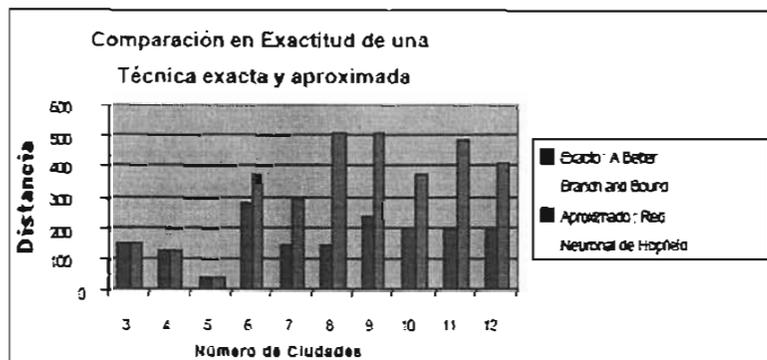


Fig. 4 Gráfica de Barras de Red Neuronal de Hopfield.

En las figuras 1 a 4 se observa que, para la muestra de 10 problemas, tanto para el método Dos-Optimal como el Híbrido Dos Optimal-Prim, 8 problemas TSP dan resultados exactos y dos de ellos dan aproximados. Lo anterior en contraste con el método Adaptación de Prim y Redes Neuronales de Hopfield que de 10 problemas, sólo 3 dan resultados exactos y 7 de ellos dan resultados aproximados. Esto sugiere que los dos primeros tienden a obtener una mejor aproximación que los dos últimos.

Análisis del Tiempo de Ejecución con respecto al número de ciudades.

Se generó una prueba con 10 matrices de adyacencia, de 10 problemas diferentes del TSP simétrico de 3 a 12 ciudades. Se corrieron experimentos, tanto con las técnicas exactas como con las aproximadas para hacer una mejor comparación. El conjunto de instancias de la prueba de los métodos aproximados se resume en la siguiente tabla y los resultados se graficaron en 3D para lograr una mejor apreciación de su comportamiento.

Resumen de Matrices Simétricas de 3 a 12 Ciudades de Métodos Aproximados											
TipoPAV/Tiempo (miliseg)	3	4	5	5	7	8	9	10	11	12	
2-Optimal	1	0	1	1	1	1	2	3	4	6	
Adaptación Prim	2	0	1	1	1	0	0	0	1	1	
Híbrido 2-Optimal-Prim	0	0	0	1	1	1	1	1	4	2	
Red Neuronal de Hopfield	67	152	90	997	7451	721	15227	20541	37617	34816	

Tabla 1 Muestra del TSP Simétrico de 3 a 12 Ciudades usando Métodos Aproximados.

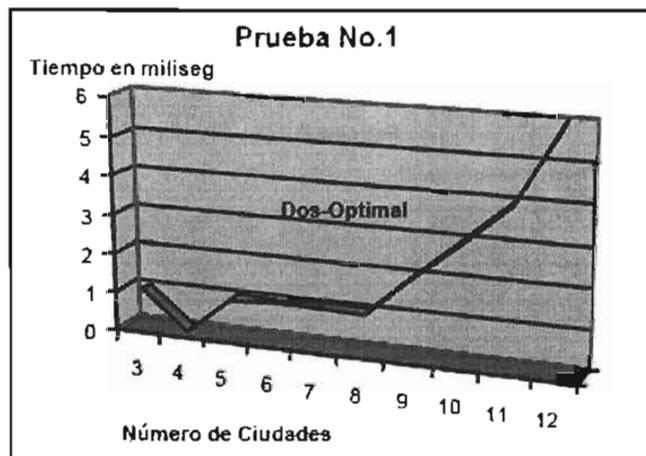


Fig. 5 Búsqueda 2-Opt en 3D.

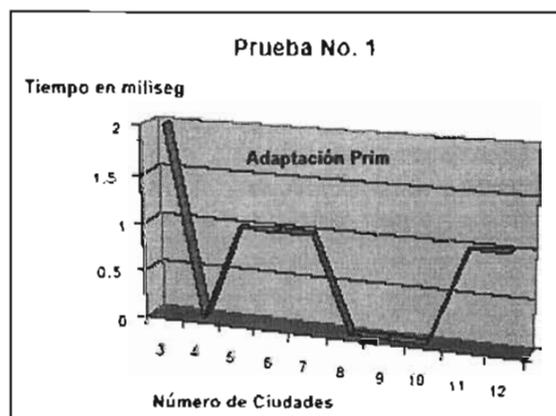


Fig. 6 Búsqueda Adaptación de Prim en 3D.

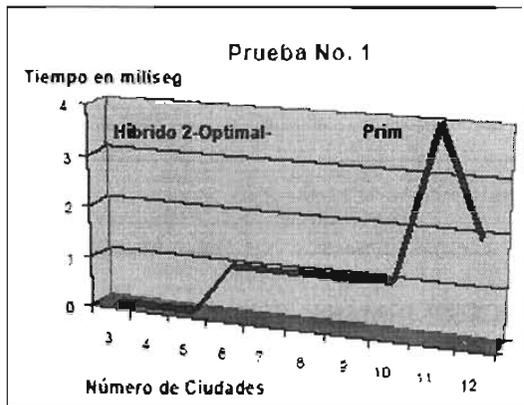


Fig.7 Búsqueda Híbrido Dos Optimal-Prim en 3D.

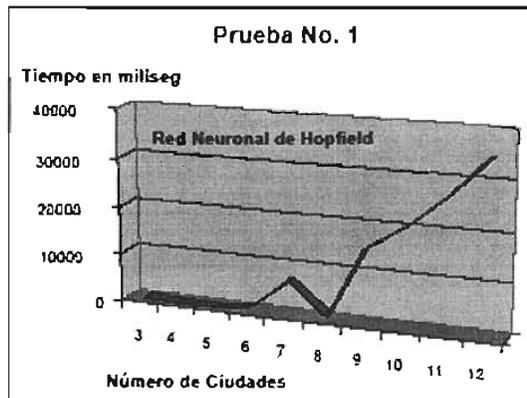


Fig. 8 Búsqueda Red Neuronal de Hopfield en 3D.

Al comparar el tiempo de ejecución entre una técnica exacta y una aproximada se obtuvieron los siguientes resultados.

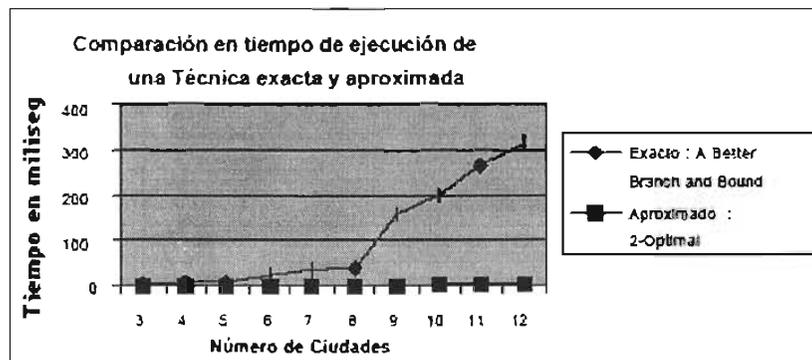


Fig. 9. 2-Optimal contra Exacto en Tiempo de Ejecución.

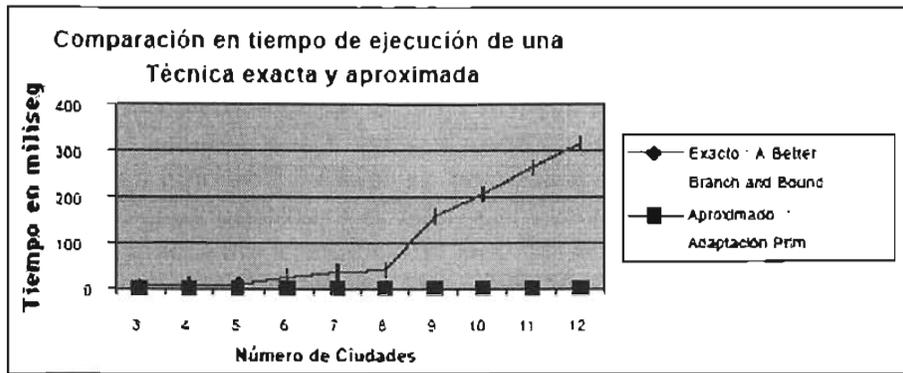


Fig.10 Adaptación de Prim contra Exacto en Tiempo de Ejecución.

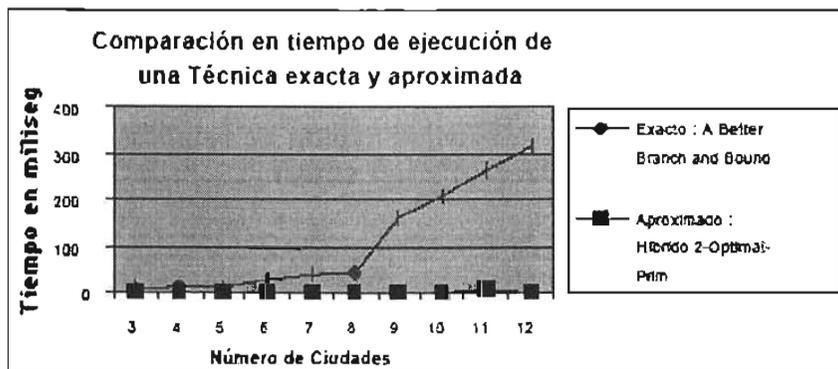


Fig. 11 Híbrido Dos Optimal-Prim contra Exacto en Tiempo de Ejecución.

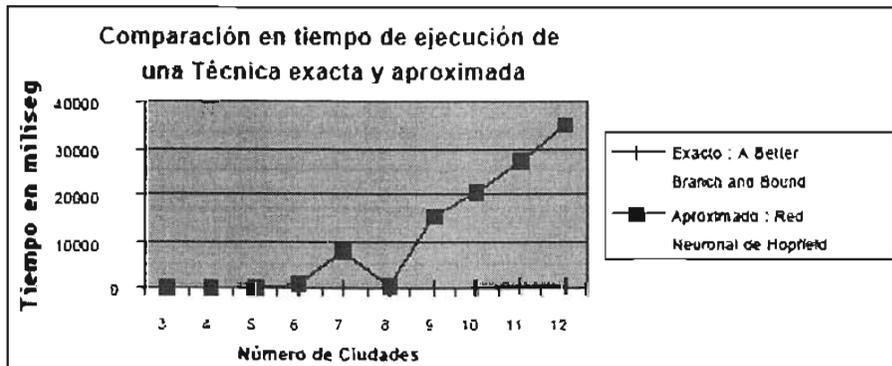


Fig. 12 Red Neuronal de Hopfield contra Exacto en Tiempo de Ejecución.

Como se puede ver, de la figura 9 a la 12 se sugiere la utilización de diferentes tipos de técnicas de solución, dependiendo del número de ciudades. Para pocas ciudades (hasta antes de 8) es posible (ya que se trata de miliseg) utilizar métodos exactos, ya que el tiempo de respuesta es relativamente

pequeño y tiene como ventaja sobre un aproximado, que la solución es la exacta.

Con un número de ciudades mayor de 10, la diferencia en tiempo aumenta, por lo que son mejores los métodos aproximados. Cabe señalar que siempre la curva de los aproximados (a excepción de la Red Neuronal de Hopfield), se prolonga por abajo de la curva de un método exacto, lo cual significa que el tiempo de respuesta será menor que la respuesta de una técnica exacta, aunque en algún momento la respuesta será aproximada.

La Red Neuronal de Hopfield requiere de tiempos mayores del resto de los aproximados; pero, en este punto cabe aclarar que la red Neuronal de Hopfield se corrió con los parámetros default de Hopfield, por facilidad en el muestreo. Pueden mejorarse los resultados en exactitud y tiempo, si se juega con los parámetros, es decir, se pueden variar los parámetros iniciales de la red, a otros diferentes a los que Hopfield establece por default.

Debe tomarse en cuenta que la bondad del método de la Red de Hopfield sobre el resto de los métodos aproximados, es que éstos usan siempre una ruta inicial que van mejorando. La Red Neuronal de Hopfield por el contrario parte de cero, es decir, sin ruta inicial. Sin embargo, encuentra siempre una ruta válida.

Ejemplo del Método 2-Opt.

Usando el Método 2-Opt descrito en el Capítulo 3, se realiza una descripción más detallada para un TSP simétrico de 5 ciudades (Castañeda, 2000). A continuación se muestra el TSP con su respectiva matriz de adyacencia.

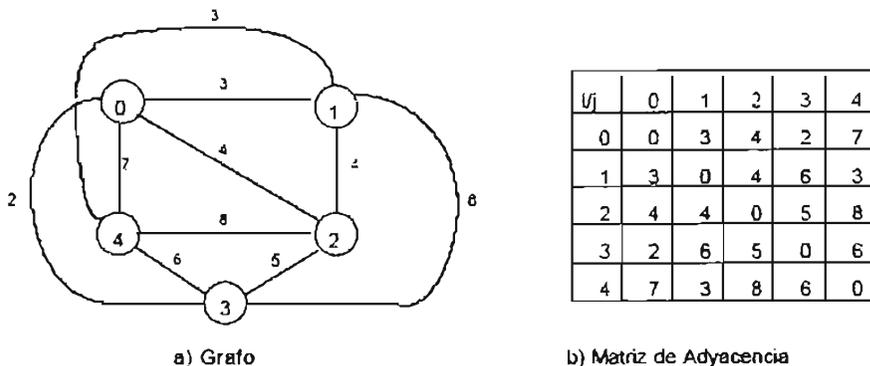


Fig. 1 Un problema TSP simétrico de 5 ciudades.

Con los datos de la matriz de adyacencia se calcula el circuito Hamiltoniano inicial (ruta inicial) a mejorar y que es 0-1-2-3-4-0 con un peso de $3+4+5+6+7 = 25$.

Ahora, reconsidérese la Fig. 1, para comenzar con el recorrido de la Fig. 2a. El proceso se inicia con $\delta_{\max} = 0$ y con la selección de 2 aristas; la formada por los vértices (0, 1) y una arista vecina (2, 3) (figura 2a). Se proponen dos aristas nuevas, las formadas por el cruce de (i_1, j_1) , y la otra por el cruce de (i_2, j_2) , cuyas distancias son :

$$\begin{aligned} \text{Distancia } (i_1, i_2) &= \text{Distancia } (0, 1) = 3 \\ \text{Distancia } (j_1, j_2) &= \text{Distancia } (2, 3) = 5 \\ \text{Distancia } (i_1, j_1) &= \text{Distancia } (0, 2) = 4 \\ \text{Distancia } (i_2, j_2) &= \text{Distancia } (1, 3) = 6 \end{aligned}$$

Se calcula su peso :

$$\max 1 = (a[i_1][i_2] + a[j_1][j_2]) - (a[i_1][j_1] + a[i_2][j_2]) = -2$$

Pero $\max 1$ no es mayor que δ_{\max} así que se buscan otras 2 parejas.

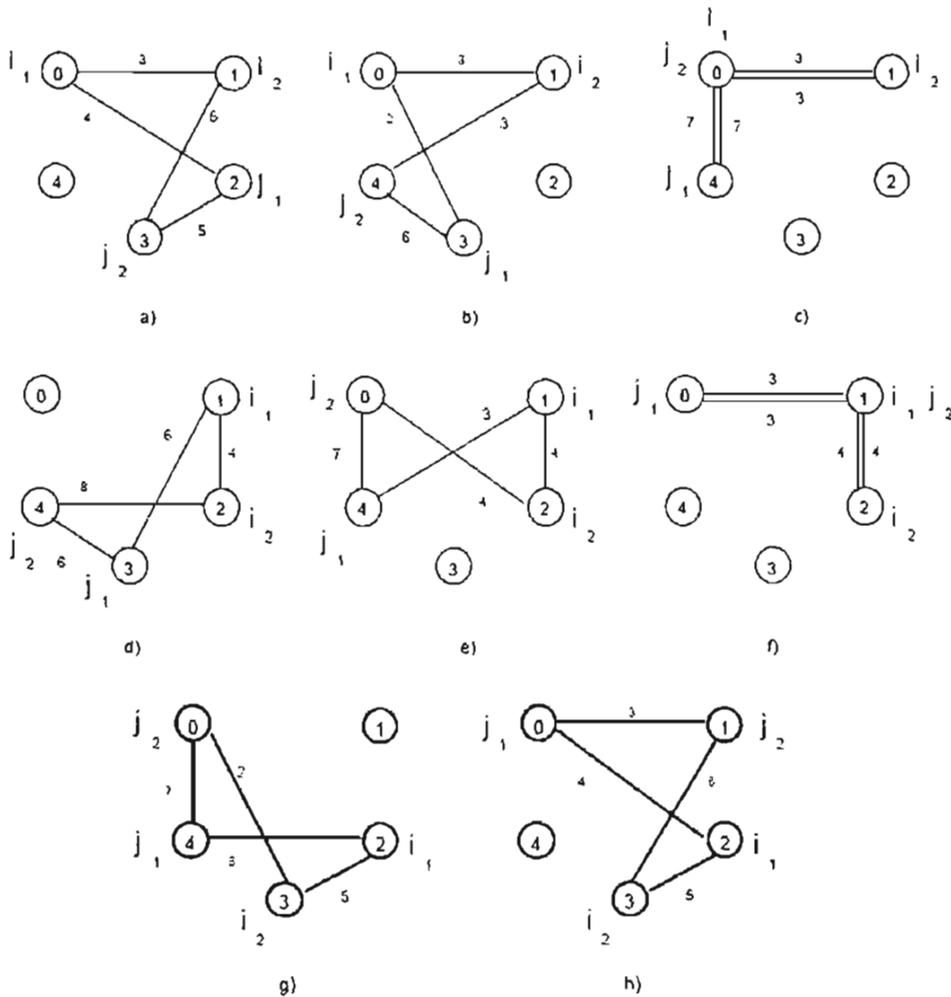


Fig. 2 Aristas mejoradas de un problema TSP simétrico de 5 ciudades.

Se reemplazan (i_1, i_2) y (j_1, j_2) , cuyo costo total es 9, por (i_1, j_1) y (i_2, j_2) , con un costo total de 5; y una diferencia $(3+6) - (2+3) = 4$, como se muestra en la Fig. 2b, dando un peso actual de $25(\text{peso anterior}) - 4(\text{mejora}) = 21$. A pesar de que algunos autores dicen que las aristas vecinas no deben ser adyacentes, hay al final de cada iteración una adyacente como se muestra en la figura 2c y 2f.

Se continua el proceso de búsqueda de mejores aristas, encontrando otra mejora como se muestra en la siguiente figura.

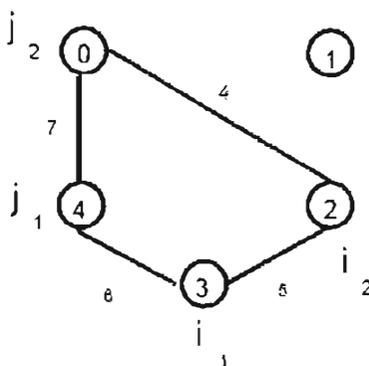


Fig. 3 Otro par de aristas mejoradas.

Esta vez se reemplazan (i_1, i_2) y (j_1, j_2) , cuyo costo total es 12, por (i_1, j_1) y (i_2, j_2) , con un costo total de 10; y una diferencia de 2, como se muestra en la Fig. 3. Dando un peso actual de $21(\text{peso anterior}) - 2(\text{mejora}) = 19$. Se puede comprobar que no hay ningún par de aristas que se pueda eliminar de esta figura y ser reemplazado con ventaja por aristas cruzadas con los mismos extremos. Por lo que el peso, es mínimo, dando el recorrido óptimo mostrado en la Fig. 4a, esta ruta se redibuja en la Fig. 4b, siguiendo los nodos para que sea más entendible. La ruta, en la que la ciudad inicial es 0 está dada 0-3-4-1-2-0 con un peso óptimo de 19.

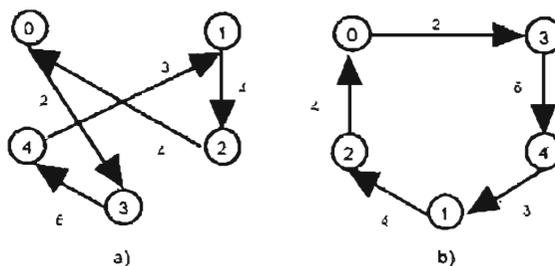


Fig. 4 Ruta óptima después de 2-Opt.

Es fácil verificar que, para una k fija, el número de diferentes transformaciones con opción k que se necesita considerar, si existen n vértices, es $O(n^k)$. Por ejemplo, el número exacto es $n(n-3)/2$ para $k = 2$. Sin embargo, el tiempo requerido para obtener un recorrido localmente óptimo puede ser mucho mayor que esto, ya que se pudieron haber realizado muchas transformaciones

locales antes de alcanzar un recorrido localmente óptimo, y cada transformación de mejora introduce aristas nuevas que pueden participar en transformaciones posteriores que mejoran aún más el recorrido fijado.

Ejemplo del Método Adaptación Prim.

Se realiza una descripción más detallada del Método Adaptación Prim descrito en el Capítulo 3, para el TSP Simétrico de 5 ciudades que se muestra en la Fig. 1.

Se parte de la Fig. 5a; puede iniciarse en cualquier nodo, pero en este caso será el nodo 0. Se busca la arista con el menor peso que salga del nodo 0, siendo el nodo 3 con un peso de 2. Y se repite la operación, recordando que un nodo que ya fue seleccionado no puede participar nuevamente. Como se muestra en la Fig. 5a a la Fig. 5g.

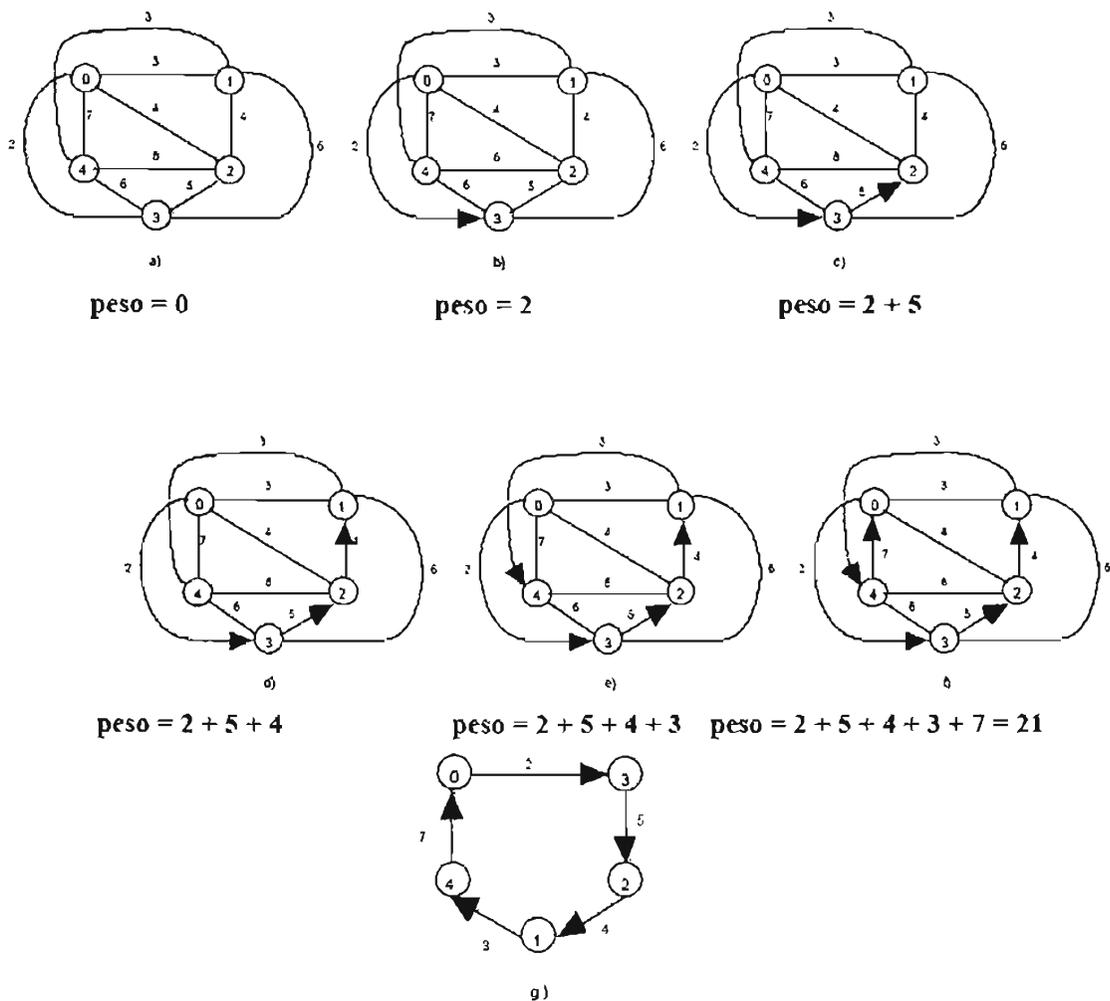


Fig. 5 Ruta óptima después de Adaptación_Prim.

En la Fig. 5d hay dos números de peso mínimo con valor 4, pero el nodo 0 no puede seleccionarse porque primero formaría un ciclo, lo cual no se permite en el árbol abarcador de Prim; y segundo, se repetiría el nodo 0 algo no permitido en el TSP. Por lo que sólo queda el mínimo 4 que va al vértice 1.

En la Fig. 5f se encuentra el peso mínimo y la ruta óptima, ésta se reacomoda para que se distinga mejor en la Fig. 5g. Además puede notarse que el resultado de la ruta óptima es 21, que es una aproximación al valor real que es de 19.

Referencias

Aceves García, R. (1996) Un Algoritmo para Resolver el Problema de Localización de Servicios con Restricciones de Demanda y Adicionales. *Tesis Doctoral (Investigación de Operaciones)*, México: DEP-FI, UNAM.

Achutan, N.R. (2000) A Branch and Cut Algorithm for Single Commodity, *Trabajo presentado en la 5th Conference of the Association of Asian-Pacific Operations Research Societies, APORS2000, Singapore.*

Adelman, D. (2003) Price-Directed Replenishment of Subsets: Methodology and Its Application to Inventory Routing. *Manufacturing & Service Operations Management* 5, No. 4, 348-371.

Anily, S. (1974) The general multi-retailer EOQ problem with vehicle routing costs. *European Journal of Operational Research* 79, 451-473.

Anily, S. and Federgruen, A. (1990a) One warehouse multiple retailer systems with vehicle routing costs. *Management Science* 36, 92-114.

Anily, S. and Federgruen, A. (1990b) A class of euclidean routing problems with general route cost functions. *Mathematics of Operations Research* 15, 268-285.

Anily, S. and Federgruen, A. (1991a) Rejoinder to comments on one warehouse multiple retailer systems with vehicle routing costs. *Management Science* 37, 1497-1499.

Anily, S. and Federgruen, A. (1991b) Structured partitioning problems. *Operations Research* 39, 130-149.

Anily, S. and Federgruen, A. (1993) Two-echelon distribution systems with vehicle routing costs and central inventories. *Operations Research* 41, 37-47.

Anily, S. and Bramel, J. (2004) An asymptotic 98.5%-effective lower bound on fixed partition policies for the inventory-routing problem. *Discrete Applied Mathematics* 145, 22-39.

Assad, A. and Golden, B. L. (1995) Arc routing methods and applications. *Handbooks in Operations Research and Management Science, Vol. 8: Network Routing*. Eds. M. O. Ball, T. L. Magnanti, C. L. Monma, and G. L. Nemhauser, 375-483. Amsterdam: Elsevier.

Baita, F., Ukovich, W., Pesenti, R. and Favaretto, D. (1998) Dynamic Routing-And-Inventory Problems: A Review, *Transportation Research* 32, No. 8, 585-598.

Bard, J., Huang, L., Jaillet, P. and Dror, M. (1998) A Decomposition Approach to the Inventory Routing Problem with Satellite Facilities, *Transportation Science* 32, No. 2, 189-203.

Benjamin, J. (1989) An analysis of inventory and transportation costs in a constrained network. *Transportation Science* 23, 177-183.

Berman, O. and Larson R.C. (2001) Deliveries in an Inventory/Routing Problem Using Stochastic Dynamic Programming. *Transportation Science* 35, 192-213.

Bertazzi, L., Paletta, G., and Speranza, M.G. (2002) Deterministic Order-Up-To Level Policies in an Inventory Routing Problem. *Transportation Science* 36, No. 1, 119-132.

Bertazzi, L., Speranza, M. G. and Ukovich, W. (1997) Minimization of logistic costs with given frequencies. *Transportation Research* 31, 327-340.

Blanchini, F., Queyranne, M., Rinaldi, F. and Ukovich, W. (1996a) A feedback strategy for periodic network flows. 27, 25

Blanchini, F., Rinaldi, F. and Ukovich, W. (1996 b) A dynamic game model for distribution problems with non-stochastic uncertainty. *International Journal of Production Economics* 45, 479-487.

Blanchini, F., Rinaldi, F. and Ukovich, W. (1997) A network design problem for distribution system with uncertain demands. *SIAM journal on Optimization* 7, 560-578.

Blanchini, F. and Ukovich, W. (1993) A linear programming approach to the control of discrete-time periodic systems with uncertain inputs. *Journal of Optimization Theory and Applications* 78,523-539.

Blumenfeld, D. E., Burns, L. D., Diltz, J. D. and Daganzo, C. F. (1985) Analyzing trade-offs between transportation inventory and production costs on freight networks. *Transportation Research* 19, 361-380.

Blumenfeld, D. E., Burns, L. D., Daganzo, C. F., Frick, M. C. and Hall, R. W. (1987) Reducing logistics costs at General Motors. *Interfaces* 17, 26-47.

- Bramel, J. and Simchil Levi, D. (1995) A location based heuristic for general routing problems. *Operations Research* 43, 649-660.
- Burns, L. D., Hall, R. W., Blumenfeld, D. E. and Daganzo, C. F. (1985) Distribution strategies that minimize transportation inventory costs. *Operations Research* 33, 469-490.
- Campbell, A., Clarke L., and Savelsbergh M. (2002) Inventory Routing in Practice. *The Vehicle Routing Problem*. Eds. P. Toth and D. Vigo., 309-330. Philadelphia: Society for Industrial and Applied Mathematics.
- Campbell, A., Clarke L., Kleywegt, A. and Savelsbergh M. (1998) The Inventory Routing Problem. *Fleet Management and Logistics*. Eds. T.G. Crainic, G. Laporte. Kluwer, 95-112. Massachusetts: Academic Publishers.
- Campbell, A., and Hardin, J. (2005) Vehicle minimization for periodic deliveries. *European Journal of Operational Research* 165, 668-684.
- Campbell, J. F. (1993) One-to many distribution with transshipments: an analytic model. *Transportation Science* 27, 330-340.
- Castañeda Roldán, C. (2000) Estudio comparativo de diversos métodos de solución del problema del agente viajero (PAV). *Tesis de Maestría en Ingeniería en Sistemas Computacionales*, Puebla: Universidad de las Américas.
- Chandra, P. and Fisher, M.L. (1994) Coordination of production and distribution planning. *European Journal of Operational Research* 72, 503-517.
- Chan, L.M., Federgruen A. and Simchi-Levi D. (1998) Probabilistic Analyses and Practical Algorithms for Inventory-Routing Models. *Operations Research* 46, 96-106.
- Chan, Y., Carter, W. and Burnes, M. (2001) A multiple-depot, multiple-vehicle, location-routing problem with stochastically processed demands. *Computers & Operations Research* 28, 803-826.
- Chase, R., Aquilano, N. and Jacobs, R. (2001), *Operation Management for Competitive Advantage*. New York: MacGraw-Hill.
- Chen, D. F., Egbelu, P. J. and Wu, C. T. (1994) Production planning models for a central factory with multiple satellite factories. *International Journal of Production Research* 32, 1431-1450.
- Cheng L. and Durán, M. (2004) Logistics for world-wide crude oil transportation using discrete event simulation and optimal control. *Computers & Chemical Engineering* 28, 897-911.

Chien, W., Balakrishnan, A. and Wong, R. (1989) An integrated inventory allocation and vehicle routing problem. *Transportation Science* 23,67.

Chikan, A. (ed.) (1990) *Inventory Models*. Kluwer, Dordrecht.

Christiansen, M. (1999) Decomposition of a Combined Inventory and Time Constrained Ship Routing Problem. *Transportation Science* 33, 3-16.

Christofides, N. and Eilon, S. (1969) An algorithm for the vehicle dispatching problems. *Operations Research* 20, 309-318.

Clarens, G. C. and Hurdle, V. F. (1975) An operating strategy for a commuter bus system. *Transportation Science* 9, 1-20.

Cousineau-Ouimet, K. (2002) A Tabu Search Heuristic for the Inventory Routing Problem. *Artículo no editado. Department of Quantitative Methods, École des Hautes Études Commerciales*. Montreal, Canada.

Daganzo, C. Y. (1985) Supplying a single location from heterogeneous sources. *Transportation Research* 19, 409-419.

Daganzo, C. F. (1987) The break-bulk role of terminals in many-to-many logistic networks. *Operations Research* 35, 543-555.

Daganzo, C. F. (1988) A comparison of in-vehicle and out-of vehicle freight consolidation strategies. *Transportation Research* 22, 173-180.

Daganzo, C. (1996) *Logistics System Analysis*. New York: Springer.

Daganzo, C. F. and Newell, G. F. (1985) Physical distribution from a warehouse: Vehicle coverage and inventory levels. *Transportation Research* 19B, 397-407.

Daganzo, C. F. and Newell G. F., (1986) Configuration of physical distribution networks. *Networks* 16, 113-132.

Daganzo, C. F. and Newell, G. F. (1993) Handling operations and the lot size trade-off. *Transportation Research* 27, 167-183.

Dror, M. and Ball, M (1987) Inventory/routing: reduction from an annual to a short-period problem. *Naval Research Logistics* 34, 891-905.

Dror, M. and Levy, L. (1986) A vehicle routing improvement algorithm comparison of a greedy and a matching implementation for inventory routing. *Computers and Operations Research* 13, 33-45.

Dror, M. and Trudeau, P. (1988) Inventory routing: operational design. *Journal of Business Logistics* 9, 165-183.

- Dror, M. and Trudeau, P. (1996) Cash flow optimization in delivery scheduling. *European Journal of Operational Research* 88, 504-515.
- Ernst, R. and Pyke, D. F. (1993) Optimal base stock policies and truck capacity in a two-echelon system. *Naval Research Logistics* 40, 879-903.
- Farvolden, J. M., Laporte, G. and Xu, J. (1993) solving an inventory allocation and routing problem arising in grocery distribution, Internal Report 886 Université de Montréal, Centre de recherché sur les transports.
- Federgruen, A., Prastacos, G. and Zipkin, P. (1986) An allocation and distribution model for perishable products. *Operations Research* 34, 75-82.
- Federgruen A. and Simchi-Levi D. (1995) Analysis of Vehicle Routing and Inventory-Routing Problems. *Handbooks in Operations Research and Management Science. Vol. 8. Network Models*. Eds. M. O. Ball, T.L. Magnanti, C.L. Monma, and G. L. Nemhauser, 225-330. Amsterdam: Elsevier.
- Federgruen, A. and Tzur, M. (1996) Detection of minimal forecast horizons in dynamic programs with multiple indicators of the future. *Naval Research Logistics* 43, 169-189.
- Federgruen, A. and Zipkin. P. (1984) A comprised vehicle routing and inventory allocation problem. *Operations Research* 32 1019-1037.
- Fisher, M. (1995) Vehicle Routing. *Handbooks in Operations Research and Management Science, Vol. 8: Network Routing*. eds. M. O. Ball, T. L. Magnanti, C. L. Monma, and G. L. Nemhauser, pp.1-33 Elsevier, Amsterdam.
- Frank, A. and Tardos, E. (1988) Generalized polymatroids and submodular flows. *Mathematical programming* 42, 489-563.
- Fujishige, S. (1991) Submodular functions and optimization. *Annals of Discrete mathematics* 471, North Holland.
- Gallego, G. and Simchi-Levi, D. (1990) On the effectiveness of direct shipping strategy for the one-warehouse multi-retailer R-systems. *Management Science* 36, 240-243.
- Garey, M. and Johnson, D. (1979) *Computers and intractability: A guide to the theory of np-completeness*. San Francisco, Calif.: Freeman.
- Geoffrion, A.M. (1974) Multicommodity distribution system design by Benders decomposition. *Management Science* Vol.20, No.5, 822-844.
- Ghiani, G., Laporte, G. and Musmanno R. (2004) *Introduction to logistics Systems Planning and Control*. John Wiley & Sons, England.

- Golden, B., Assad, A., and Dahl, R. (1984) Analysis of a large scale vehicle routing problem with an inventory component. *Large Scale Systems* 7, 181-190.
- Hahm, J. and Yano, C. A. (1995) The economic lot and delivery scheduling problem: powers of two policies. *Transportation Science* 29, 22-241.
- Hall, R. W. (1985) Determining vehicle dispatch frequency when shipping frequency differs among suppliers. *Transportation Research* 19B, 421-431.
- Hall, R. W. (1991) Comment on One warehouse multiple retailer systems with vehicle routing costs. *Management Science* 37, 1496-1497.
- Hall, R. W. (1992) A note on bounds for direct shipping cost. *Management Science* 38, 1212.
- Herer, Y. and Levy, R. (1997) The Metered Inventory Routing Problem, a Integrative Heuristic Algorithm. *Production Economics* 51, 69-81.
- Holmberg, K. (1990) On the Convergence of Cross Decomposition. *Mathematical Programming* 47, 269-296.
- Hwang, H. (1999) A Food Distribution Model for Famine Relief. *Computers & Industrial Engineering* 37, 335-338.
- Jackson, P. L., Maxwell, W. L. and Muckstadt, J. A. (1988) Determining optimal reorder intervals in capacitated production-distribution systems. *Management Science* 34, 938-958.
- Jaillet, P., Bard, J., Huang, L. and Dror, M. (2002) Delivery Cost Approximations for Inventory Routing Problems in a Rolling Horizon Framework. *Transportation Science* 36, No. 3, 292-300.
- Jünger, M., Reinelt, G. and Rinaldi, G. (1995) The Traveling Salesman Problem. *Handbooks in Operations Research and Management Science, Vol. 7. Network Models*. Eds. M. O. Ball, T. L. Magnanti, C. L. Monma, and G. L. Nemhauser, pp. 225-330 Elsevier, Amsterdam.
- Kleywegt, A., Nori V., and Savelsbergh, M. (2002) Dynamic Programming Approximations for a Stochastic Inventory Routing Problem. *Transportation Science* 38, No. 1, 42-70.
- Kleywegt, A., Nori V., and Savelsbergh, M. (2002) The Stochastic Inventory Routing Problem with Direct Deliveries. *Transportation Science* 36, 94-118.
- Klincewicz, J. G. (1990) Solving a freight transport problem using facility location techniques. *Operations Research* 38, 99-109.

- Langeving, A., Mbaraga, P. and Campbell, J. F. (1991) Continuous approximation models in freight distribution: an overview. *Transportation research* 30B, 163-188.
- Lasdon, L.S. (1970) *Optimization Theory for Large Systems*. USA: McMillan Series in Operations Research.
- Maxwell, W. L. and Muckstadt, J. A. (1985) Establishing consistent and realistic reorder intervals in production-distribution systems. *Operations Research* 33, 1316-1341.
- Miller, D. M. (1987) Interactive, computer-aided ship scheduling system. *European Journal of Operational Research* 32, 363-379.
- Muckstadt, J. A. and Roundy, R. O. (1993) Analysis of multistage production systems. *Handbooks in Operations Research and Management Science. Vol. 4: Logistics of Production and Inventory*. eds. S. C. Graves, A. G. G. Rinnooy Kan, and P. P. Zipkin, Elsevier, pp. 59-131. Elsevier, Amsterdam.
- Murty, K. G., (1995) *Operations Research: Deterministic Optimization Models*. New Jersey: Prentice-Hall.
- Newel, G. P. (1971) Dispatching policies for a transportation route. *Transportation Science* 5, 91-105.
- Papadimitriou, C. M. (1994) *Computational complexity*. Massachusetts: Addison-Wesley.
- Popken, D. A. (1994) Algorithm for the multiattribute multicommodity flow problem with freight consolidation and inventory costs. *Operations Research* 42, 274-286.
- Potvin, J.Y., Kervahut, T. and Rosseau, J-M. (1992) A tabu-search heuristic for the vehicle routing problem with time windows, Working paper 885, Université, de Montréal, Centre de recherche sur les transports.
- Qu, W., Bookbinder, J. and Iyogun P. (1999) An integrated inventory-transportation system with modified periodic policy for multiple products. *European Journal of Operational Research* 115, 254-269.
- Rusdiansyah, A., Tsao, D. (2004) An integrated model of the periodic delivery problems for vending-machine supply chains. *Journal of Food Engineering* Article in press.
- Reiman, S., Rubio, R. and Wein, L.M. (1999) Heavy Traffic Analysis of the Dynamic Stochastic Inventory-Routing Problem. *Transportation. Science* 33, 361-380.

Roundy, R. (1985) Ninety eight percent effective integer ratio lot sizing for one warehouse multi-retailer systems. *Management Science* 31, 1416.

Salomon, M. (1991) *Deterministic Lot Sizing Models for Production planning*, Springer, Berlin.

Sarmiento, A. M. and Nagi R. (2003) A Review of Integrated Analysis of Production-Distribution Systems. *Artículo no editado. Department of Industrial Engineering, University of New York at Buffalo*. Buffalo, N.Y.

Solomon, M. (1987) Vehicle routing and scheduling problems with time windows constraints. *Operations Research* 35, 254-265.

Speranza, M. G. and Ukovich, W. (1994a) Analysis and integration of optimization models for logistic systems. *International Journal of Production Economics* 35, 183-190.

Speranza, M. G. and Ukovich, W (1994b) Minimizing transportation and inventory costs for several products on a single link. *Operations Research* 42, 879-874.

Speranza, M. G. and Ukovich, W. (1996) An algorithm for optimal shipments with given frequencies. *Naval Research Logistics* 43, 655-671.

Syslo, Maciej M. (1983) *Discrete optimization algorithms with pascal programs*. Englewood cliffs, New Jersey : Prentice Hall.

Toth, P. and Vigo, D. (2002) *The Vehicle Routing Problem. SIAM Monographs on Discrete Mathematics and Applications*. Philadelphia: Society for Industrial and Applied Mathematics.

Trudeau, P. and Dror, M. (1992) Stochastic inventory routing: route design with stockouts and route failures. *Transportation Science* 26, 171-184.

Van Roy, T. J. (1983) Cross decomposition for mixed integer programming. *Mathematical Programming* 25, 46-63.

Van Roy, T. J. (1984) A cross decomposition algorithm for capacitated facility location. *Operation Research Society of America* 34, 145-163.