

RICARDO SARAIVA DE CAMARGO

**SISTEMAS EIXO-RAIO DE MÚLTIPLA ATRIBUIÇÃO:  
MODELOS E ALGORITMOS**

Belo Horizonte  
13 de setembro de 2007

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
INSTITUTO DE CIÊNCIAS EXATAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**SISTEMAS EIXO-RAIO DE MÚLTIPLA ATRIBUIÇÃO:  
MODELOS E ALGORITMOS**

Tese apresentada ao Curso de Pós-Graduação  
em Ciência da Computação da Universidade  
Federal de Minas Gerais como requisito parcial  
para a obtenção do grau de Doutor em Ciência  
da Computação.

RICARDO SARAIVA DE CAMARGO

Belo Horizonte  
13 de setembro de 2007



UNIVERSIDADE FEDERAL DE MINAS GERAIS

FOLHA DE APROVAÇÃO

Sistemas Eixo-Raio de Múltipla Atribuição:  
Modelos e Algoritmos

RICARDO SARAIVA DE CAMARGO

Tese defendida e aprovada pela banca examinadora constituída por:

Ph. D. HENRIQUE PACCA LOUREIRO LUNA – Orientador  
Universidade Federal de Alagoas

Ph. D. ROBERTO DIÉGUEZ GALVÃO  
Universidade Federal do Rio de Janeiro

Ph. D. GERALDO ROBSON MATEUS  
Universidade Federal de Minas Gerais

Ph. D. REINALDO MORABITO NETO  
Universidade Federal de São Carlos

Ph. D. MAURÍCIO CARDOSO DE SOUZA  
Universidade Federal de Minas Gerais

Belo Horizonte, 13 de setembro de 2007

# Resumo

Em sistemas de telecomunicações e de transporte, o problema localização de concentradores com atribuição múltipla surge quando demandas entre muitos pares de origem-destino devem ser enviadas. Ao invés de servir cada demanda entre uma origem e um destino com uma conexão direta, os fluxos são reunidos a partir de concentradores (*hubs*) e enviados, através de uma rede de concentradores, para serem então entregues aos respectivos destinos. Sistemas com esse tipo de topologia são chamados de sistemas do tipo eixo-raio ou redes do tipo eixo-raio. Essas redes são projetadas para usufruírem da economia de escala obtida ao se concentrar o fluxo durante o transporte e/ou transmissão entre concentradores. Portanto, o problema consiste em encontrar a configuração de rede eixo-raio com o menor custo possível selecionando quais locais a se instalar um concentrador e atribuindo o tráfego a cada um, dados os custos de instalação, de transporte e a demanda entre os pares de origem e destino. No presente trabalho, três versões diferentes do problema são estudadas: o problema clássico, a economia de escala dependente do fluxo e o congestionado. Algoritmos eficientes baseados no método de Decomposição de Benders foram empregados na resolução dos problemas, resolvendo instâncias de grande porte, consideradas “fora do alcance” de métodos exatos, em um tempo aceitável.

**Palavras-chaves:** Redes eixo-raio, Decomposição de Benders; Otimização de Grande Porte

# Abstract

In telecommunication and transportation systems, the multiple allocation hub location problem arises when we must flow commodities or information between several origin-destination pairs. Instead of establishing a direct node to node connection from an origin to its destination, the flows are concentrated with others at facilities called hubs. These flows are transported on links established between hubs, being then splitted and delivered to its final destination. Systems with this sort of topology are named hub-and-spoke systems or hub-and-spoke networks. They are designed to exploit the scale economies attainable through the shared use of high capacity links between hubs. Therefore, the problem is to find the least expensive hub-and-spoke network, selecting hubs and assigning traffic to them, given the demands between each origin-destination pair and the respective transportation costs. In the present work, we study three different versions of the problem: the classical, the flow dependent scale economies and the congestioned. Efficient Benders decomposition algorithms have been deployed to tackle the problems. We have been able to solve some large instances, considered ‘out of reach’ of other exact methods in reasonable time.

**Keywords:** Hub-and-spoke networks; Benders decomposition; Large-scale optimization

*Aos meus pais, Irfêo e Léa Camargo, meus heróis.*

*“Não to mandei eu? Sê forte e corajoso; não temas, nem te espantes, porque o SENHOR, teu Deus, é contigo por onde quer que andares.” (Josué, 1:9)*

# Agradecimentos

Eu gostaria de agradecer primeiramente a **Deus**, meu grande parceiro.

Aos meus pais, Irfêo e Léa Camargo, pelo eterno amor e apoio, e por me possibilitarem seguir meus sonhos.

Ao Prof. Henrique Pacca Loureiro Luna, pela sua orientação paciente, pelas muitas sugestões e *insights* que permitiram este trabalho.

Ao Prof. Gilberto de Miranda Jr, meu amigo e parceiro de pesquisa, pela sua paciência infinita, por ter sempre motivado, ajudado e valorizado meu trabalho, indo muito além de seu papel de pesquisador.

Ao Prof. Ricardo Poley Martins Ferreira, pelas caminhadas filosóficas.

Ao Prof. Geraldo Robson Mateus, pelo grande suporte durante minha passagem pelo Departamento de Ciência de Computação.

Ao Prof. Maurício Cardoso de Souza, pelas valiosas discussões, sempre precisas.

Ao Prof. Carlos Roberto Venâncio de Carvalho, por me apresentar o mundo da pesquisa operacional.

Aos professores, funcionários e colegas do Departamento de Ciência de Computação.



# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Exame da Literatura</b>	<b>5</b>
2.1	Exemplos de Sistemas Eixo-Raio . . . . .	5
2.1.1	Transporte Aéreo . . . . .	5
2.1.2	Transporte Público Urbano . . . . .	6
2.1.3	Serviços Postais . . . . .	7
2.1.4	Transporte de Carga Terrestre . . . . .	8
2.1.5	Telecomunicações . . . . .	8
2.2	Classificação e Taxonomia . . . . .	9
2.3	Formulações Matemáticas . . . . .	12
2.3.1	Formulação de Campbell . . . . .	13
2.3.2	Formulação de Klineciewicz . . . . .	14
2.3.3	Formulação de Skorin-Kapov et al. . . . .	14
2.3.4	Formulação de Aykin . . . . .	15
2.3.5	A Formulação de Ernst e Krishnamoorthy . . . . .	16
2.3.6	Formulação de Hamacher et al. . . . .	17
2.3.7	Formulações de Marín et al. . . . .	18
2.4	Métodos de Resolução . . . . .	21
2.5	O Método de Decomposição de Benders . . . . .	23
<b>3</b>	<b>Localização de Concentradores com Atribuição Múltipla</b>	<b>27</b>
3.1	Algoritmos de Decomposição de Benders . . . . .	28
3.1.1	Problema Mestre e Subproblema . . . . .	28
3.1.2	Algoritmo de Decomposição de Benders Clássico . . . . .	29
3.1.3	Algoritmo de Decomposição de Benders Multicorte . . . . .	31
3.1.4	Algoritmo de Decomposição de Benders $\varepsilon$ -Ótimo . . . . .	32
3.1.5	Subproblema . . . . .	33
3.2	Resultados Computacionais . . . . .	36
3.3	Comentários finais . . . . .	42
<b>4</b>	<b>Localização de Concentradores com Economia de Escala Dependente do Fluxo</b>	<b>45</b>

4.1	Definições e Formulações . . . . .	46
4.2	Algoritmo de Decomposição de Benders . . . . .	50
4.2.1	Subproblema . . . . .	50
4.2.2	Problema Mestre . . . . .	53
4.3	Resultados Computacionais . . . . .	55
4.4	Comentários Finais . . . . .	56
<b>5</b>	<b>Localização de Concentradores com Congestionamento</b>	<b>59</b>
5.1	Definições e Formulação . . . . .	60
5.2	Método de Decomposição de Benders Generalizado . . . . .	61
5.2.1	Problema Mestre . . . . .	62
5.2.2	Subproblema . . . . .	65
5.2.3	Algoritmo de Decomposição de Benders Generalizado . . . . .	70
5.3	Resultados Computacionais . . . . .	71
5.4	Comentários Finais . . . . .	73
<b>6</b>	<b>Conclusão</b>	<b>74</b>
	<b>Referências Bibliográficas</b>	<b>76</b>

# Lista de Figuras

1.1	(a) Conexões entre pares origem/destino. (b) Eixo-raio. . . . .	2
1.2	Exemplos de alocação das origens e dos destinos das demandas a concentradores em um sistema eixo-raio. . . . .	3
2.1	Exemplo de sistemas eixo-raio em transporte urbano. . . . .	7
2.2	Tipos de redes ligadas aos concentradores do <i>backbone</i> . . . . .	9
2.3	Exemplos de alocação de pontos de demanda a concentradores em um sistema eixo-raio. . . . .	10
2.4	Representação dos segmentos de custo . . . . .	12
2.5	Exemplo de uma solução com três concentradores em uma rota. . . . .	19
3.1	Rotas possíveis para o par $i - j$ . . . . .	35
3.2	Esforço computacional versus variação dos custos de instalação. . . . .	37
3.3	Número de concentradores instalados versus variação dos custos de instalação. . . . .	37
3.4	Esforço computacional versus mudança no coeficiente de variação dos custos fixos. . . . .	38
3.5	Esforço computacional ( $\alpha = 20\%$ ). . . . .	41
3.6	Esforço computacional ( $\alpha = 40\%$ ). . . . .	41
3.7	Esforço computacional ( $\alpha = 60\%$ ). . . . .	41
3.8	Esforço computacional ( $\alpha = 80\%$ ). . . . .	41
3.9	Proporção do tempo gasto pelo algoritmo BD1 executando o problema mestre e o subproblema para as instâncias AP20.4, AP30.4, AP40.4, AP50.4 e AP60.4 . . . .	42
3.10	Taxa de convergência do limite inferior dos algoritmos BD1 e BD2 para a instância AP30.6 . . . . .	42
3.11	Proporção do tempo gasto pelo algoritmo BD2 executando o problema mestre e o subproblema para as instâncias AP20.4, AP30.4, AP40.4, AP50.4 e AP60.4 . . . .	43
4.1	Função côncava linear por partes da conexão $k - m$ . Extraído de Klincewicz (2002) . . . .	46
4.2	Representação das variáveis $\tilde{x}_{ijk}^l$ . . . . .	48
4.3	Comparação da qualidade das relaxações lineares. . . . .	57
5.1	Função de congestionamento para diferentes valores de $b_k$ e $d_k = 1$ . . . . .	60
5.2	Modificação da representação da rede. . . . .	66

# Lista de Tabelas

2.1	Resumo dos protocolos . . . . .	11
3.1	Esforço computacional para o conjunto CAB. . . . .	39
3.2	Esforço computacional para o conjunto AP. . . . .	39
3.3	Resultandos computacionais para as instâncias do conjunto AP. . . . .	40
3.4	Resultandos computacionais para as instâncias do conjunto AP. . . . .	43
3.5	Resultandos computacionais para as instâncias do conjunto AP. . . . .	44
4.1	Parâmetros das funções côncavas lineares por partes. (Extraído de Klineciewicz (2002)) . . . . .	55
4.2	Tamanho das instâncias . . . . .	56
4.3	Comparação das relaxações lineares . . . . .	56
4.4	Resultados computacionais . . . . .	58
5.1	Resultados computacionais . . . . .	72
5.2	Distribuição de fluxo para uma rede com 10 pontos de demanda, $d = 1$ e $b = \{1, 1; 1, 5; 1, 7; 2, 0\}$ . . . . .	72

# Capítulo 1

## Introdução

*“Comece fazendo o possível; então faça o que é necessário; e, de repente, você estará fazendo o impossível.”*

São Francisco de Assis

Sistemas do tipo eixo-raio ou, em inglês, *“hub-and-spoke systems”*, tornaram-se uma importante área de pesquisa da teoria de localização nas últimas décadas. A relevância é em grande parte explicada pela sua utilização em sistemas logísticos, tanto de passageiros quanto de cargas, e em redes de telecomunicações.

Nesses tipos de sistemas, ao invés de servir cada demanda entre uma origem e um destino com uma conexão direta, os fluxos são reunidos a partir de concentradores (*hubs*)<sup>1</sup> e enviados, através de uma rede de concentradores, para serem então entregues aos respectivos destinos.

Dessa forma, as demandas entre muitos pares de origem-destino usufruem da economia de escala obtida ao se concentrar o fluxo durante o transporte e/ou transmissão. Essa economia de escala permite então uma significativa redução nos custos de transporte.

A figura 1.1 ilustra a diferença entre ligações entre pares de demanda origem-destino ligadas diretamente, figura (a), e via sistemas eixo-raio, figura (b).

As aplicações dos sistemas do tipo eixo-raio são muitas. Em sistemas logísticos, destacam-se: o transporte aéreo de passageiros e de cargas, serviços expressos para entregas de encomendas no dia seguinte (*overnight delivery*), sistemas de transporte de cargas rodoviárias e ferroviárias, e sistemas postais. Nesses casos, a demanda é especificada em termos de fluxos de passageiros ou produtos entre pares de cidades ou localidades, e é transportada por algum tipo de veículo: avião, carro, caminhão e trem. Os concentradores funcionam como terminais ou centros de triagem que agregam pequenas quantidades de fluxos em quantidades maiores. Essa agregação permite o compartilhamento do meio de transporte e, conseqüentemente, rateio dos custos de transporte; obtendo-se assim economia de escala.

No setor de telecomunicações, os modelos do tipo eixo-raio abordam desde o processamento distribuído de computadores até vídeo conferência, passando por redes de telefonia pública. Nesses casos, o fluxo é dado em função da transmissão de informação (voz, dados,

---

<sup>1</sup>O termo concentrador é utilizado de forma genérica, porém, dependendo da aplicação, é também conhecido como ponto de concentração, terminal de consolidação, centro de consolidação, terminal de triagem e centro de distribuição.

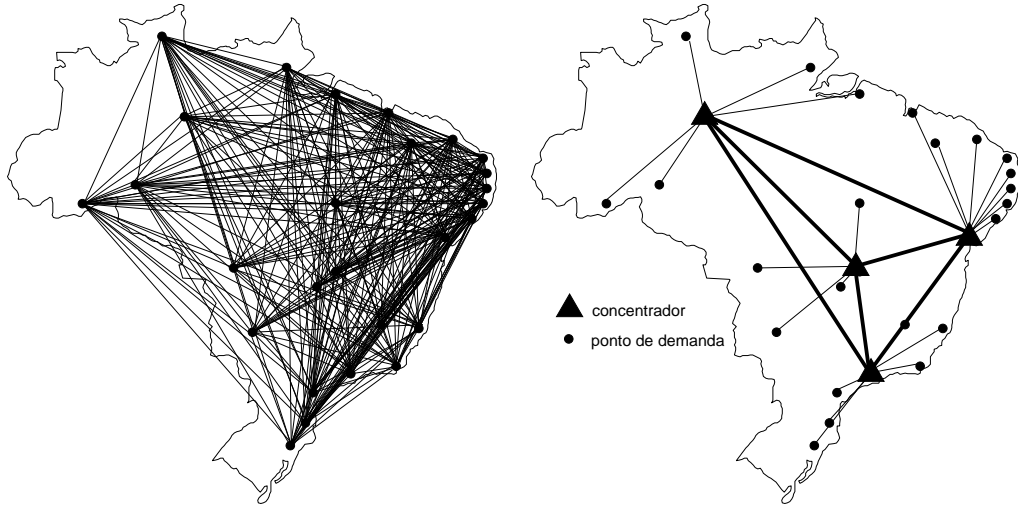


Figura 1.1: (a) Conexões entre pares origem/destino. (b) Eixo-raio.

vídeo) através de uma variedade de meios físicos de transporte (fibra ótica, par trançado, cabos coaxiais) ou pelo ar, como em enlaces de satélite e de microondas. O papel do concentrador, nessas redes, é o de funcionar como centros de multiplexação, comutação e roteamento, interligando desde pequenas redes corporativas até redes globais.

Nas aplicações acima, o desenho das redes consiste basicamente na resolução de dois tipos de problemas: o primeiro é um problema de localização que envolve a instalação de facilidades que atuarão como concentradores e a ligação entre elas, isto é, os *eixos*; enquanto que o segundo é um problema de atribuição, ou seja, como as origens e os destinos das demandas são alocados (ligados) aos concentradores instalados, isto é, a formação dos *raios*.

Apesar das semelhanças, os problemas do tipo eixo-raio distinguem-se dos problemas clássicos de localização. Nestes, geralmente, os pontos de demandas são atendidos, total ou parcialmente, por facilidades a serem localizadas. Enquanto, que naqueles, as demandas ocorrem entre pares de origem-destino. Nestes casos, a função dos concentradores não é o atendimento das demandas, mas o de funcionar como centros de comutação ou concentração de fluxo. Os concentradores permitem assim a consolidação, seguida de um redirecionamento e, posteriormente, de uma separação dos fluxos; sendo portanto facilidades intermediárias no caminho entre os diversos pares de origem-destino.

Enquanto os concentradores são considerados totalmente conectados entre si, formando uma *clique*; as origens e os destinos das demandas podem ser alocados aos concentradores instalados de duas formas: alocação simples, ou seja, as origens e os destinos das demandas podem ser alocados à apenas um único concentrador (veja figura 1.2 A); e alocação múltipla, implicando que se podem conectá-los a mais de um concentrador da rede (veja figura 1.2 B).

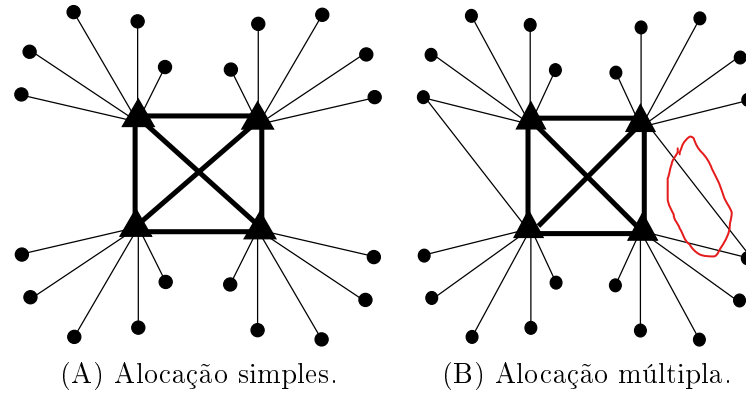


Figura 1.2: Exemplos de alocação das origens e dos destinos das demandas a concentradores em um sistema eixo-raio.

Apesar das diversas variantes dos sistemas eixo-raio, o foco principal deste trabalho é propor formulações e abordagens eficientes para resolver problemas de localização de concentradores com múltipla atribuição. Algumas das principais contribuições desta tese são:

- o desenvolvimento de um algoritmo exato e competitivo para a resolução de instâncias do problema de localização de concentradores com atribuição múltipla;
- uma formulação matemática para o problema de localização de concentradores com atribuição múltipla e tratando a economia de escala como uma função de custo côncava linear por partes;
- o desenvolvimento de um algoritmo exato e competitivo para a resolução de instâncias do problema de localização de concentradores com atribuição múltipla, tratando a economia de escala como uma função de custo côncava linear por partes;
- a extensão das formulações existentes para o problema de localização de concentradores com atribuição múltipla e considerando os efeitos de congestionamento;
- o desenvolvimento de um algoritmo para a resolução de instâncias do problema de localização de concentradores com atribuição múltipla e considerando os efeitos de congestionamento.

Este trabalho está organizado da seguinte forma:

No capítulo 2, é apresentado um exame da literatura, mostrando a taxinomia da área, as principais formulações matemáticas e alguns dos métodos de resolução do problema de localização de concentradores com atribuição múltipla em redes do tipo eixo-raio.

No capítulo 3, um algoritmo especializado, baseado no método de Decomposição de Benders, é aplicado na resolução de instâncias padrão do problema de localização de concentradores com atribuição múltipla. Resultados computacionais são apresentados para verificar a eficiência do algoritmo e a qualidade das respostas obtidas.

No capítulo 4, a economia de escala é representada por uma função côncava linear por partes dependente do fluxo entre concentradores. Uma formulação matemática, com o limite de programação linear mais justo do que o presente na literatura, é proposta. Mais uma vez, um algoritmo especializado, baseado no método de Decomposição de Benders, é aplicado na resolução de instâncias consideradas, até então, “fora do alcance” dos métodos exatos em computadores pessoais.

No capítulo 5, além da localização dos concentradores e da atribuição múltipla dos clientes, o nível de congestionamento dos concentradores é considerado. Esse congestionamento é representado por uma função convexa não-linear típica das áreas de transporte. Um algoritmo especializado, baseado no método de Decomposição de Benders Generalizado, é usado na resolução de instâncias do problema. Resultados computacionais são apresentados para verificar a eficiência do algoritmo e a qualidade das respostas obtidas.

O capítulo 6 encerra o trabalho, apresentando as conclusões, os comentários finais e as sugestões para os trabalhos futuros.



## Capítulo 2

# Exame da Literatura

*“Se estes e estas podem, porque não eu?”*

Santo Agostinho

Neste capítulo, são apresentados alguns exemplos de sistemas eixo-raio, uma taxonomia da área, as principais formulações matemáticas para o problema de localização de concentradores com múltipla atribuição, assim como uma breve discussão de alguns algoritmos de resolução existentes na literatura.

### 2.1 Exemplos de Sistemas Eixo-Raio

As aplicações de sistemas do tipo eixo-raio são as mais variadas possíveis, destacando-se: transporte aéreo, transporte público urbano, serviços postais, sistemas logísticos, e, finalmente, telecomunicações.

#### 2.1.1 Transporte Aéreo

No transporte aéreo, além do custo operacional e do desempenho do sistema, o conforto dos passageiros é uma preocupação constante. Além do impacto do custo nos preços das tarifas, os efeitos da decolagem e do pouso dos aviões são considerados no desenho dessas redes (Bryan e O’Kelly, 1999).

Uma forma de reduzir os custos é empregando sistemas do tipo eixo-raio. Ao invés de se interligar diretamente todos os pontos de demanda ou criar trechos extremamente longos, segmentam-se as rotas, criando assim o conceito de “pernas”. Nos trechos compartilhados por conexões, aeronaves maiores podem ser usadas para se concentrar o fluxo de passageiros. Essa concentração permite, conseqüentemente, um possível rateio dos custos.

Drezner e Drezner (2001) e Sasaki et al. (1999), preocupados com o desconforto dos passageiros, consideram casos especiais de desenho de redes do tipo eixo-raio para o transporte aéreo nos quais somente um concentrador é usado, qualquer que seja o caminho entre a origem e o destino. Ao fazerem isso, o projeto da rede pode ser transformado em um problema das  $p$ -medianas. Por sua vez, Jaillet et al. (1996) apresentam um estudo comparativo de modelos

que permitem o uso de um número qualquer de concentradores no caminho da origem até o destino da demanda.

Apesar de poucos trabalhos considerarem as influências dos custos operacionais do transporte aéreo na demanda, alguns autores se destacam nesse sentido: Skorin-Kapov (1998) propõe um modelo que distribui, de forma justa, os custos de prestação dos serviços aos usuários da rede; enquanto Marianov et al. (1999) propõem uma abordagem alternativa que relaciona a demanda entre origem e destino ao custo de se usar a rede de concentradores.

Os efeitos de congestionamento no desempenho geral da rede são outra grande preocupação das empresas aéreas e dos gerenciadores de aeroportos (concentradores). No transporte aéreo, o congestionamento pode ser abordado tanto nos concentradores (aeroportos) quanto nas conexões (aeronaves).

No primeiro caso, Marianov e Serra (2003) se destacam ao apresentarem um modelo que considera o congestionamento nos concentradores através de um sistema de filas  $M/D/c$ , onde  $c$  é o número pistas de pouso e de decolagem. O modelo não apenas assegura, através de um conjunto de restrições, um nível mínimo de serviço, como propõe possíveis expansões de capacidade, aumentando-se ou diminuindo-se o número de pistas de pouso e decolagem dos aeroportos. Por sua vez, Elhedhli e Hu (2005) usam uma formulação não-linear para considerar os custos de congestionamento na função objetivo. Uma linearização, por um conjunto finito de retas, é então proposta para simplificar o próprio modelo.

No segundo caso, Jaillet et al. (1996) propõem um modelo que aborda o congestionamento nas conexões, através de um conjunto de restrições de capacidade, permitindo a alocação de aeronaves de qualquer tamanho a trechos do caminho desde a origem até o destino da demanda.

### 2.1.2 Transporte Público Urbano

Nesse tipo de aplicação, os concentradores são conhecidos como estações de embarque e desembarque, podendo não estar diretamente conectados entre si. Além disso, nem todos os custos são proporcionais às distâncias entre pares de origem e de destino. A idéia principal é criar corredores de descolamento entre concentradores (terminais) e, a partir desses, atender os pontos de demanda usando linhas locais de ônibus. A figura 2.1 ilustra a ligação entre concentradores e entre clientes a concentradores.

Apesar das pesquisas dessa área serem ainda limitadas, os trabalhos de Nickel et al. (2000) e de Kimms (2006) se destacam. Nickel et al. (2000) apresentam um modelo que faz a localização de interconexões entre concentradores, no lugar de apenas concentradores. O modelo busca aumentar o máximo possível o fluxo entre concentradores através do aumento do desconto na conexão entre dois concentradores. Isso assegura que o passageiro não escolherá a rota de menor caminho ou a de menor tempo, mas a de menor custo (tarifa).

Kimms (2006), através de uma manipulação de índices das variáveis, permite adaptar algumas formulações matemáticas do transporte aéreo na elaboração dos corredores de deslocamento. A manipulação permite ainda o uso de diferentes níveis de desconto entre cada concentrador.

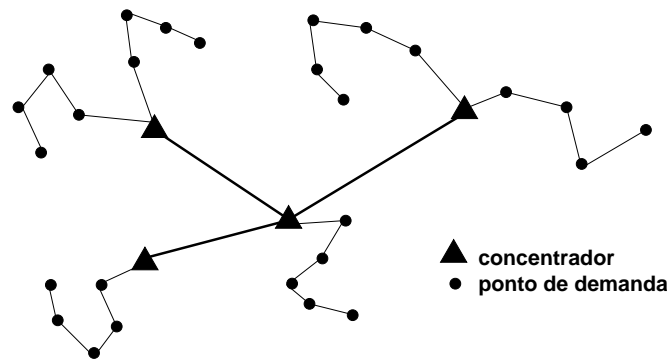


Figura 2.1: Exemplo de sistemas eixo-raio em transporte urbano.

### 2.1.3 Serviços Postais

Em serviços postais, os pontos de demanda representam distritos de coleta de correio, o fluxo corresponde ao volume de cartas e/ou pacotes e os concentradores funcionam como facilidades de consolidação e processamento (separação) de correspondência, mais conhecidos como centros de triagem.

Em redes desse tipo, os custos de coleta e de distribuição não são iguais, uma vez que envolvem processos e meios de transporte diferentes. No primeiro, o cliente se dirige à uma unidade dos correios; enquanto no último, um veículo se encaminha às casas dos clientes. Além disso, uma carta pode sair de um ponto de origem de demanda ir até um concentrador e retornar ao mesmo distrito, agora como destino, para ser entregue, em outras palavras, o distrito de coleta pode ter demanda para ele mesmo. Isto pode acontecer, uma vez que o processamento e a separação das cartas acontecem nos concentradores (Ernst e Krishnamoorthy, 1996). Normalmente, o número de pontos de demanda dessas redes é relativamente maior do que os problemas de transporte aéreo e público.

Os trabalhos realizados para os correios australianos e estadunidenses se sobressaem no emprego de sistemas eixo-raio para serviços postais. No primeiro, Ernst e Krishnamoorthy (1996, 1999) propõem uma formulação matemática na qual o custo de transporte é segmentado em três parcelas: coleta, processamento, distribuição. Essa segmentação permite que o problema seja visto como uma rede de três camadas: uma camada para os pontos de origem, uma outra para os concentradores e, finalmente, uma para os pontos de destino do fluxo. O problema pode ser resolvido balanceando-se o fluxo entre as camadas.

No segundo caso, Donaldson et al. (1999) propõem um modelo próprio de localização de centros de separação do tipo *cross-docking*<sup>1</sup> para os correios dos Estados Unidos da América. Além de determinar a localização dos centros, a formulação matemática proposta determina, através de restrições de tempo de entrega, o número de caminhões a serem usados nas rotas de cada concentrador. O modelo utilizado é muito próximo dos sistemas logísticos convencionais de transporte de carga.

<sup>1</sup> *Cross-docking* é um sistema de distribuição no qual as mercadorias recebidas em um centro não são colocadas no estoque do armazém, mas são o mais rapidamente preparadas para o carregamento e a distribuição para os clientes ("Chegou e já sai!").

### 2.1.4 Transporte de Carga Terrestre

Sistemas eixo-raio para carga terrestre são muito similares às redes eixo-raio aéreas e postais, fornecendo transporte entre vários pares de origem e de destino. O papel do concentrador é o mesmo, servir como ponto de transferência e consolidação.

Entretanto, o desenho dessas redes pode possuir preocupações diferentes dependendo da estratégia a ser adotada na utilização do veículo no transporte da carga (O’Kelly, 1998), isto é: (i) se a carga transportada usa toda a capacidade do meio de transporte, *i.e.* carga completa, também chamada de *truckload* (TL); ou (ii) se a carga transportada não usa toda a capacidade do meio de transporte, também chamada de *less-than-truckload* (LTL) ou carga fracionada ou parcelada.

No primeiro caso (i), a preocupação maior no desenho do sistema eixo-raio não é com a consolidação de carga, mas sim com a diminuição da duração da viagem realizada pelo condutor transportando a carga. Nesses casos, os concentradores têm o papel de servirem de pontos de troca de condutores, permitindo assim que os mesmos voltem às respectivas casas com maior frequência (Taylor et al., 1995).

No segundo caso (ii), a preocupação maior no desenho do sistema eixo-raio é a consolidação da carga em volumes maiores para que a economia de escala no transporte seja obtida. Fatores como a distância e o tempo máximo de duração do transporte, assim como a capacidade e o tamanho da frota de veículos, são elementos relevantes a serem considerados no modelo. Como as atividades de coleta e de distribuição podem ou não ser feitas pelo mesmo veículo, problemas tais como roteamento de veículos e de atribuição de rotas a concentradores podem surgir, aumentando a complexidade do desenho desse tipo de rede (Nagy e Salhi, 2005).

### 2.1.5 Telecomunicações

Normalmente, em sistemas eixo-raio de telecomunicações, os custos de instalação e de interconexão dos concentradores e a ligação desses aos pontos de demanda são ordens de grandeza maior do que o custo de transporte. Isto acontece devido ao baixo custo operacional da rede uma vez instalada.

Além do custo de transporte ser geralmente menor do que em redes logísticas do tipo eixo-raio, há diversas formas de se conectar os pontos de origem e de destino da demanda aos concentradores. Klinecicz (1998) apresenta uma extensa revisão sobre as diferentes formas de interconexão desses pontos, mais precisamente em redes do tipo tributária e de *backbone*. A figura 2.2 ilustra os principais tipos de redes tributárias ligadas aos concentradores que, por sua vez, formam um dos tipos de rede *backbone*, no caso constituindo um grafo completo.

Na figura 2.2, uma rede em árvore está conectada ao concentrador *A*, enquanto uma em forma de caminho, ligando os pontos de demanda, está ligada ao concentrador *B*. Esses dois tipos são freqüentemente referenciados como linhas do tipo multi-ponto. Uma rede em estrela está associada ao concentrador *C*. Nesse tipo, cada ponto está ligado diretamente a um concentrador, como na telefonia fixa. Finalmente, uma rede em anel está conectada ao concentrador *D*. Esse tipo é muito comum em aplicações que fazem uso de fibra ótica. Apesar

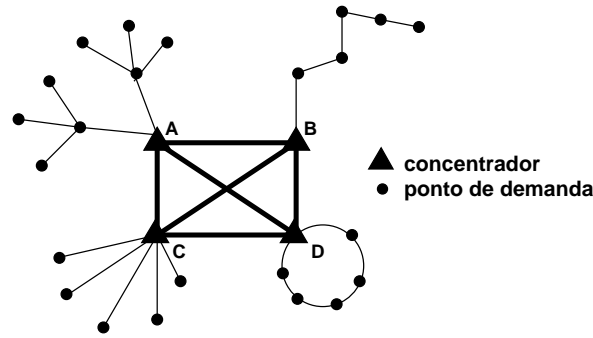


Figura 2.2: Tipos de redes ligadas aos concentradores do *backbone*

de não ilustrado, existem outros tipos de interconexão de concentradores, como por exemplo, em forma de anel, de árvores e de estrelas.

Um outro tipo de aplicação, também na área de telecomunicações, é o proposto por Gavish (1987a). Gavish é um dos primeiros autores a usar o conceito de eixo-raio em sistemas de computação distribuída, no qual cada ponto produz uma transação que é armazenada em um banco de dados. Um número de computadores devem ser instalados em locais diferentes para processar essas transações e gerar relatórios a partir delas. Gavish modela os concentradores como computadores, os pontos de origem como fontes de transações e os pontos de destino como os relatórios. A transferência de dados entre os concentradores ocorre quando um concentrador precisa de dados de outros concentradores para gerar um relatório. Gavish (1987b) ainda acrescenta custos de congestionamento durante as transações para observar o desempenho do sistema.

Na próxima seção 2.2, uma classificação e uma taxonomia são apresentadas para os problemas da área.

## 2.2 Classificação e Taxonomia

Basicamente, os sistemas eixo-raio possuem dois tipos de problemas: os com alocação simples e os com alocação múltipla<sup>2</sup>. Nos problemas de alocação simples, cada ponto de demanda só pode ser alocado a um único concentrador; enquanto que nos problemas de alocação múltipla, um ponto de demanda pode interagir com mais de um concentrador, veja o exemplo da figura 2.3.

No exemplo, no caso da alocação simples (ilustração da esquerda), as demandas de 1 para os demais pontos são roteadas, inicialmente, através do concentrador localizado em 7. Ou seja, independentemente dos destinos das demandas com origem em 1, elas obrigatoriamente serão enviadas primeiramente para o concentrador localizado em 7.

No caso da alocação múltipla (ilustração da direita), os pontos de demanda podem ser alocadas a mais de um concentrador. No exemplo, as demandas de 1 para os demais pontos podem ser roteadas tanto pelo concentrador em 7 ou pelo em 8.

<sup>2</sup>Durante o texto, os termos alocação e atribuição são usados como sinônimos, sem erro de interpretação.

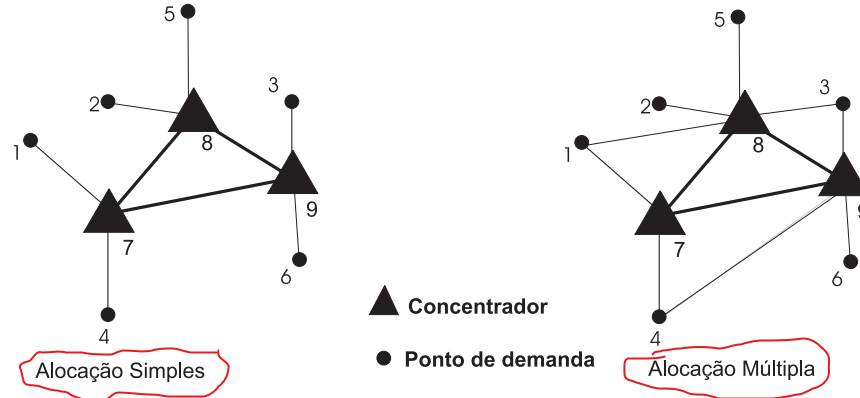


Figura 2.3: Exemplos de alocação de pontos de demanda a concentradores em um sistema eixo-raio.

Baseando-se nesses dois tipos de problemas, O’Kelly e Miller (1994) propõem uma taxinomia para os modelos do tipo eixo-raio, classificando os problemas de acordo com: o tipo de atribuição; o tipo de conexão entre concentradores, se esses estão todos interconectados ou não; o tipo de conexão entre os pontos de demanda, se esses podem se comunicar diretamente, sem o uso de algum concentrador ou não.

O’Kelly e Miller (1994) sugerem 8 protocolos resumidamente apresentados na tabela 2.1. Os protocolos A-D estão relacionados a problemas de alocação simples, enquanto os protocolos E-G estão relacionados com os problemas de alocação múltipla. Os problemas de alocação múltipla podem ser reduzidos aos problemas de alocação simples, desde que só se permita aos pontos de origem ou de destino da demanda se conectarem a apenas um concentrador.

Além dos protocolos criados por O’Kelly e Miller (1994), Hamacher e Nickel (1998) propõem um esquema de classificação mais geral que pode ser aplicado a outros modelos de localização. Esse esquema é composto de 5 parâmetros:

$$\alpha/\beta/\gamma/\delta/\eta$$

Onde:

- $\alpha$  : é o número de concentradores e/ou pontos de demanda.
- $\beta$  : é o tipo de modelo de localização com relação ao espaço de decisão: “D”, localização discreta; “C”, localização contínua.
- $\gamma$  : são restrições adicionais e/ou variações do problema (MA - alocação múltipla, SA - alocação simples, CAP - capacitados, CONG - congestionados).
- $\delta$  : é o tipo de interação entre os concentradores e os pontos de demanda, normalmente expressa em termos de distância ou algum sistema específico de custos.
- $\eta$  : é a função objetivo, normalmente a minimização dos custos de transporte ( $\sum_{\text{fluxo}}$ ) e dos custos de instalação dos concentradores ( $\sum_{\text{concentradores}}$ ).

Tabela 2.1: Resumo dos protocolos

Protocolo	Descrição	Exemplo
A	Envolve problemas de localização de concentradores com atribuição simples e grafo completo entre concentradores.	
B	Envolve problemas de localização de concentradores com atribuição simples permitindo a escolha das conexões que compõe a rede entre concentradores, isto é, não é necessário ter um grafo completo entre concentradores.	
C	Envolve problemas de localização de concentradores com atribuição simples e grafo completo entre concentradores, permitindo conexões diretas entre os pontos de demanda, isto é, sem passar por nenhum concentrador.	
D	Envolve problemas de localização de concentradores com atribuição simples, permitindo conexões diretas entre os pontos de demanda, isto é, sem passar por nenhum concentrador. Permite ainda a escolha das conexões entre concentradores.	
E	Envolve problemas de localização de concentradores com atribuição múltipla e grafo completo entre concentradores.	
F	Envolve problemas de localização de concentradores com atribuição múltipla permitindo a escolha das conexões entre concentradores, isto é, não é necessário ter um grafo completo entre concentradores.	
G	Envolve problemas de problemas de localização de concentradores com atribuição múltipla e grafo completo entre concentradores, permitindo ainda conexões diretas entre os pontos de demanda, isto é, sem passar por nenhum concentrador.	
H	Envolve problemas de problemas de localização de concentradores com atribuição múltipla permitindo conexões diretas entre os pontos de demanda e a escolha das conexões entre concentradores.	

Essa classificação é muito utilizada na literatura para distinguir uma classe de problemas da outra. No presente trabalho, apenas os problemas de localização de concentradores com atribuição múltipla são considerados, protocolo E, ou *Conc. /D/MA/ $\sum_{fluxo} \sum_{conc.}$*  e *Conc. /D/MA, CONG/ $\sum_{fluxo} \sum_{conc.}$* , sendo *conc.* a abreviação de concentrador.

Na próxima seção 2.3, as principais formulações matemáticas para os problemas do protocolo E são apresentadas.

## 2.3 Formulações Matemáticas

Todos as formulações para o problema de localização de concentradores com atribuição múltipla são baseadas nas seguintes definições: sejam  $N$  o conjunto de pontos de demanda e  $K$  o conjunto de pontos candidatos a se instalar um concentrador, tal que  $K \subseteq N$ . Para qualquer par de pontos  $i$  e  $j$  ( $i, j \in N \mid i \neq j$ ), têm-se  $w_{ij} > 0$ , a demanda do ponto  $i$  para o ponto  $j$  que deve ser roteada através de um ou dois concentradores instalados. Normalmente, têm-se  $w_{ij} \neq w_{ji}$ . Sejam também  $a_k$ , o custo de instalação de um concentrador no ponto  $k \in K$ , e  $c_{ijkm}$ , o custo unitário de transporte do ponto  $i$  até  $j$  roteado via os concentradores instalados em  $k$  e  $m$  ( $i, j \in N$  e  $k, m \in K$ ), nesta ordem. Se apenas um concentrador for usado, então têm-se  $k = m$ .

O custo unitário de transporte da rota ligando os pontos  $i$  e  $j$ , através dos concentradores em  $k$  e  $m$ , nesta ordem, é a composição dos custos dos segmentos ou  $c_{ijkm} = c_{ik} + \alpha c_{km} + c_{mj}$ . As parcelas  $c_{ik}$  e  $c_{mj}$  são, respectivamente, os custos unitário de transporte do ponto  $i$  até o concentrador em  $k$  e do concentrador em  $m$  até o ponto  $j$ . Têm-se ainda que  $c_{kk} = 0$ , quando  $m = k$ . O fator de desconto  $\alpha$  ( $0 \leq \alpha \leq 1$ ) representa a economia de escala nas conexões entre concentradores, sendo, portanto,  $\alpha c_{km}$  o custo unitário de transporte com desconto entre os concentradores em  $k$  e  $m$ . A figura 2.4 ilustra os segmentos de custos e as quatro rotas possíveis no transporte entre os pontos  $i$  e  $j$ :  $i - k - k - j$  ( $c_{ijkk} = c_{ik} + c_{kj}$ ),  $i - m - m - j$  ( $c_{ijmm} = c_{im} + c_{mj}$ ),  $i - k - m - j$  ( $c_{ijkm} = c_{ik} + \alpha c_{km} + c_{mj}$ ) e  $i - m - k - j$  ( $c_{ijmk} = c_{im} + \alpha c_{mk} + c_{kj}$ ).

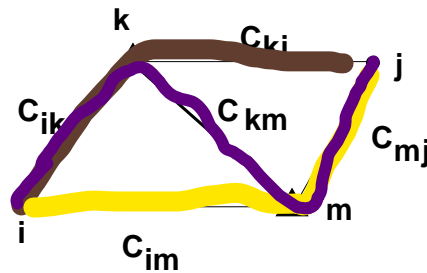


Figura 2.4: Representação dos segmentos de custo

Nas próximas seções, algumas formulações matemáticas para o problema de localização de concentradores com atribuição múltipla são apresentadas. Durante toda a tese, os índices  $i$  e  $j$  ( $i, j \in N$ ) indicam os pontos de origem e de destino de demanda, respectivamente, sendo  $i \neq j$ , enquanto os índices  $k$  e  $m$  ( $k, m \in K$ ) representam os pontos candidatos a



se instalar os concentradores. Portanto, os somatórios mostrados nas formulações são feitos em relação aos índices e a pertinência dos mesmos aos respectivos conjuntos. Além disso, normalmente,  $K \subset N$ , entretanto vai-se considerar que todos os pontos de demanda são candidatos potenciais a se instalar um concentrador, ou seja,  $K \equiv N$ .

### 2.3.1 Formulação de Campbell

Campbell (1994a) propõe uma variável contínua para cada rota possível entre cada par origem-destino originando:

$x_{ijkm} \geq 0$  : é a fração de fluxo transportada entre os pontos  $i$  e  $j$ ,  
através dos concentradores em  $k$  e  $m$ , nesta ordem.

Campbell define ainda as variáveis:

$$y_k = \begin{cases} 1, & \text{se um concentrador é instalado no ponto } k; \\ 0, & \text{caso contrário.} \end{cases}$$

A formulação então é dada por:

$$\min \sum_k a_k y_k + \sum_i \sum_j \sum_k \sum_m w_{ij} c_{ijkm} x_{ijkm} \quad (2.1)$$

sujeito a:

$$x_{ijkm} \leq y_k \quad \forall i, j \in N, k, m \in K \quad (2.2)$$

$$x_{ijkm} \leq y_m \quad \forall i, j \in N, k, m \in K \quad (2.3)$$

$$\sum_k \sum_m x_{ijkm} = 1 \quad \forall i, j \in N \quad (2.4)$$

$$x_{ijkm} \geq 0 \quad \forall i, j \in N, k, m \in K \quad (2.5)$$

$$y_k \in \{0, 1\} \quad \forall k \in K \quad (2.6)$$

A função objetivo (2.1) minimiza os custos de instalação de concentradores e de transporte. As restrições (2.2)-(2.3) garantem que os fluxos só são roteados através de concentradores instalados. As restrições (2.4) asseguram que a demanda ( $w_{ij}$ ) de cada par origem-destino é roteada através de um ou no máximo dois concentradores.

Essa formulação apresenta  $n^4 - n^3 + n$  variáveis e  $2(n^4 - n^3) + n^2$  restrições, sendo sua relaxação linear muito fraca. O seu uso em métodos baseados em *branch and bound*, que utilizam as relaxações lineares como limites inferiores, ou em métodos duais é desaconselhada (Klincewicz, 1996). Uma forma de melhorar a relaxação linear desse modelo é apresentado por Skorin-Kapov et al. (1996) (veja seção 2.3.3).

### 2.3.2 Formulação de Klincewicz

Klincewicz (1996), usando a formulação proposta por Campbell (1994a), indexa cada par de origem e de destino,  $i-j$ , por um único índice  $\omega \in \Omega$  e cada conexão  $k-m$  entre concentradores por um único índice  $r \in R$ , onde  $\Omega$  e  $R$  são os conjuntos de pares de origem-destino e de conexões possíveis entre concentradores, respectivamente.

Klincewicz (1996) apresenta a seguinte formulação, onde  $c_{\omega r} = w_{ij} c_{ijkm}$  e  $R_k$  ( $k \in K$ ) é o conjunto de conexões entre concentradores nas quais o concentrador em  $k$  está presente:

$$\min \sum_k a_k y_k + \sum_{\omega} \sum_r c_{\omega r} x_{\omega r} \quad (2.7)$$

sujeito a:

$$\sum_r x_{\omega r} = 1 \quad \forall \omega \in \Omega \quad (2.8)$$

$$x_{\omega r} \leq y_k \quad \forall \omega \in \Omega, r \in R_k, k \in K \quad (2.9)$$

$$x_{\omega r} \geq 0 \quad \forall \omega \in \Omega, r \in R \quad (2.10)$$

$$y_k \in \{0, 1\} \quad \forall k \in K \quad (2.11)$$

As variáveis de decisão  $x_{\omega r}$  representam a parcela de fluxo transportado desde a origem  $i$  até o destino  $j$ , índice  $\omega$ , que é roteada pela conexão  $k-m$ , índice  $r$ . A função objetivo (2.7) minimiza os custos de instalação de concentradores e de transporte. As restrições (2.8) asseguram que cada par de origem e de destino  $\omega$  é atribuído a alguma conexão  $r$ . As restrições (2.9) garantem que só há fluxo em alguma conexão  $r$  se os concentradores dessa conexão estão instalados.

A reindexação das variáveis resulta em um modelo tradicional de localização de facilidades não capacitadas, permitindo assim o uso de métodos de resolução consagrados da literatura. Entretanto o número de variáveis contínuas continua o mesmo. Klincewicz destaca ainda que, uma vez que não existem restrições de capacidade no fluxo que um concentrador pode rotar, sempre existirá uma solução ótima para a formulação. Essa observação é verdadeira para cada uma das formulações aqui apresentadas, desde que exista pelo menos um concentrador instalado e não haja restrições de capacidade de fluxo.

### 2.3.3 Formulação de Skorin-Kapov et al.

Skorin-Kapov et al. (1996) substituem as restrições (2.2)-(2.3) pelas restrições mais justas (2.13)-(2.14):

$$\min \sum_k a_k y_k + \sum_i \sum_j \sum_k \sum_m w_{ij} c_{ijkm} x_{ijkm} \quad (2.12)$$

sujeito a:

$$\sum_k x_{ijkm} \leq y_m \quad \forall i, j \in N, m \in K \quad (2.13)$$

$$\sum_m x_{ijkm} \leq y_k \quad \forall i, j \in N, k \in K \quad (2.14)$$

$$\sum_k \sum_m x_{ijkm} = 1 \quad \forall i, j \in N \quad (2.15)$$

$$x_{ijkm} \geq 0 \quad \forall i, j \in N, k, m \in K \quad (2.16)$$

$$y_k \in \{0, 1\} \quad \forall k \in K \quad (2.17)$$

As novas restrições (2.13)-(2.14) também garantem que o fluxo entre cada par de origem e de destino  $i - j$  seja atendido apenas por concentradores instalados. Como o número de restrições é reduzido de  $2(n^4 - n^3)$  para  $2(n^3 - n^2)$  e como cada solução não negativa que satisfaz o espaço de viabilidade definido por (2.13)-(2.17) também satisfaz o espaço de viabilidade definido por (2.2)-(2.6), mas não vice-versa, a relaxação linear de (2.12)-(2.17) é mais justa do que a formulação de Campbell (1994a), seção 2.3.1. Uma formulação semelhante à proposta por Skorin-Kapov et al. é apresentada por Aykin (1995).

### 2.3.4 Formulação de Aykin

As variáveis  $x_{ijkm}$ , originalmente propostas por Campbell (1994a) (veja seção 2.3.1), supostamente asseguram a convenção, introduzida por O'Kelly (1987), de permitir no máximo dois concentradores, representados pelos índices  $k$  e  $m$ , em uma determinada rota entre os pontos de origem  $i$  e de destino  $j$ . Entretanto, Aykin (1995) é um dos primeiros autores a questionar o que acontece quando concentradores estão instalados nos pontos  $i$  e  $j$  (esse questionamento é abordado ainda mais claramente por Marín et al. (2006), veja seção 2.3.7).

Quando a origem  $i$  e o destino  $j$  possuem concentradores instalados e dependendo da estrutura de custos  $c_{ijkm}$ , rotas passando por três ou, até mesmo, quatro concentradores podem ocorrer. Contornando esses casos, Aykin propõe uma formulação próxima à apresentada por Skorin-Kapov et al. (1996) (veja seção 2.3.3), dada por:

$$\min \sum_k a_k y_k + \sum_i \sum_j \sum_k \sum_m w_{ij} c_{ijkm} x_{ijkm} \quad (2.18)$$

sujeito a:

$$\sum_i \sum_j \sum_m (x_{ijkm} + x_{ijmk}) \leq M y_k \quad \forall k \in K \quad (2.19)$$

$$\sum_j \sum_{k \neq i} \sum_{m \neq i} (x_{ijkm} + x_{jikm}) \leq M (1 - y_i) \quad \forall i \in N \quad (2.20)$$

$$\sum_{\substack{k \neq i \\ k \neq j}} (x_{ijik} + x_{ijkj}) + \sum_k (x_{ijkk}) \leq M(2 - y_i - y_j) \quad \forall i, j \in N \quad (2.21)$$

$$\sum_k \sum_m x_{ijkm} = 1 \quad \forall i, j \in N \quad (2.22)$$

$$x_{ijkm} \geq 0 \quad \forall i, j \in N, k, m \in K \quad (2.23)$$

$$y_k \in \{0, 1\} \quad \forall k \in K \quad (2.24)$$

Onde  $M$  é uma constante muito grande. As restrições (2.19) garantem que só existem rotas compostas por concentradores instalados. As restrições (2.20)-(2.21) forçam as variáveis de fluxo para zero quando um concentrador é instalado na origem e/ou no destino. As restrições (2.19)-(2.21) são restrições agregadas que resultam em uma relaxação linear fraca, como reconhecido pelo próprio Aykin.

### 2.3.5 A Formulação de Ernst e Krishnamoorthy

Ernst e Krishnamoorthy (1998) apresentam uma abordagem interessante e diferente para o problema original. Ernst e Krishnamoorthy dividem o problema em três níveis: (i) um nível para o fluxo que sai da origem da demanda; (ii) um outro para o fluxo que passa através dos concentradores; e, finalmente, (iii) um outro que chega ao destino da demanda.

As variáveis de fluxo  $z_{ik}$ ,  $f_{ikm}$ ,  $x_{imj}$  são então atribuídas, respectivamente, a cada um dos níveis, significando: o fluxo da origem  $i$  até o concentrador em  $k$ ; o fluxo do concentrador em  $k$  para o concentrador em  $m$  que é originado no ponto  $i$ ; e o fluxo do concentrador em  $m$  para o destino  $j$  que é originado no ponto  $i$ . As variáveis de decisão  $y_k$  continuam tendo o mesmo significado.

Ernst e Krishnamoorthy não usam o custo unitário de transporte agregado, como as formulações anteriores. Eles associam os parâmetros  $\mathcal{X}$  para o segmento de rota  $i-k$  e  $\delta$  para o segmento  $m-j$  para representar custos diferenciados de coleta e de distribuição, respectivamente. O custo unitário de transporte então é dado por:

$$\mathcal{X} c_{ik} + \alpha c_{km} + \delta c_{mj}$$

A formulação é dada por:

$$\min \sum_k a_k y_k + \sum_i \left[ \sum_k \mathcal{X} c_{ik} z_{ik} + \sum_k \sum_m \alpha c_{km} f_{ikm} + \sum_m \sum_j \delta c_{mj} x_{imj} \right] \quad (2.25)$$

s. a:

$$\sum_k z_{ik} = \sum_j w_{ij} \quad \forall i \in N \quad (2.26)$$

$$\sum_m x_{imj} = w_{ij} \quad \forall i, j \in N \quad (2.27)$$

$$\sum_m f_{ikm} + \sum_j x_{ikj} - \sum_m f_{imk} - z_{ik} = 0 \quad \forall i \in N, k \in K \quad (2.28)$$

$$z_{ik} \leq \sum_j w_{ij} y_k \quad \forall i \in N, k \in K \quad (2.29)$$

$$x_{imj} \leq w_{ij} y_m \quad \forall i, j \in N, m \in K \quad (2.30)$$

$$z_{ik} \geq 0 \quad \forall i \in N, k \in K \quad (2.31)$$

$$f_{ikm} \geq 0 \quad \forall i \in N, k, m \in K \quad (2.32)$$

$$x_{imj} \geq 0 \quad \forall i, j \in N, m \in K \quad (2.33)$$

$$y_k \in \{0, 1\} \quad \forall k \in K \quad (2.34)$$

A função objetivo (2.25) minimiza o custo de instalação de concentradores e de transporte. As restrições (2.26) garantem que todo fluxo de cada origem deixa a mesma. As restrições (2.27) fazem com que todo fluxo de cada origem chegue ao seu ponto de destino. As restrições (2.28) são de conservação de fluxo nos concentradores, enquanto as restrições (2.29)-(2.30) asseguram que um concentrador só é usado para coleta ou distribuição, respectivamente, se o mesmo estiver instalado.

Apesar de Campbell et al. (2002) se referirem à formulação de Ernst e Krishnamoorthy como a mais eficiente e geral para problemas de localização de concentradores com atribuição múltipla; Hamacher et al. (2000) e Marín et al. (2006) contestam tal afirmativa mostrando como as restrições (2.29) e (2.30) enfraquecem a relaxação linear da mesma.

### 2.3.6 Formulação de Hamacher et al.

Usando a formulação proposta por Skorin-Kapov et al. (1996) (veja seção 2.3.3), Hamacher et al. (2000) estabelecem a dimensão do politopo da mesma substituindo as restrições (2.13)-(2.14) por (2.36), resultando em:

$$\min \sum_k a_k y_k + \sum_i \sum_j \sum_k \sum_m w_{ij} c_{ijkm} x_{ijkm} \quad (2.35)$$

sujeito a:

$$\sum_{m \neq k} x_{ijkm} + \sum_m x_{ijmk} \leq y_k \quad \forall i, j \in N, k \in K \quad (2.36)$$

$$\sum_k \sum_m x_{ijkm} = 1 \quad \forall i, j \in N \quad (2.37)$$

$$x_{ijkm} \geq 0 \quad \forall i, j \in N, k, m \in K \quad (2.38)$$

$$y_k \in \{0, 1\} \quad \forall k \in K \quad (2.39)$$

As restrições (2.36) garantem que somente há fluxo nas variáveis  $x_{ijkm}$ , se o concentrador em  $k$  for instalado; enquanto que as restrições (2.37) asseguram que a demanda do par origem-destino  $i - j$  é atendido por alguma rota.

Essa formulação possui um limite de programação linear muito justo quando comparado com a formulação de Skorin-Kapov et al. (1996) (veja seção 2.3.3), uma vez que Hamacher et al. demonstram que as restrições (2.36) definem facetas do politopo inteiro. Uma formulação semelhante à essa, que também possui uma relaxação justa, é apresentada por Marín et al. (2006) (veja seção 2.3.7).

### 2.3.7 Formulações de Marín et al.

Marín et al. (2006) fazem um estudo aprofundado das formulações de Skorin-Kapov et al. (1996), Ernst e Krishnamoorthy (1998) e Hamacher et al. (2000), comparando-as segundo a relaxação linear.

Marín et al. destacam ainda claramente que as variáveis  $x_{ijkm}$  nem sempre asseguram que no máximo dois concentradores estão presentes em uma dada rota entre a origem  $i$  e o destino  $j$ . Como mencionado na seção 2.3.4, dependendo da estrutura de custos unitários podem-se ter situações nas quais uma rota pode ter três ou, até mesmo, quatro concentradores, como ilustrado pela figura 2.5.

Na figura 2.5, têm-se três concentradores instalados, 1, 2 e 3, e três conexões. As ligações (1,2) e (1,3) têm custo unitário 1, enquanto a conexão (2,3) tem custo unitário 1.000. Lembrando que  $N \equiv K$  e que há demandas entre concentradores, então considerando uma demanda do ponto 2, isto é, do concentrador em 2, até o ponto 3 (concentrador em 3), as rotas  $i - k - m - j$  de menor custo para envio da demanda são: 2-1-1-3 ou 2-2-1-3 ou 2-1-3-3. Implicando que as variáveis  $x_{2311}$ ,  $x_{2321}$  e  $x_{2313}$  podem ser iguais a  $w_{23}$ . Indiferente de qual dessas variáveis é a escolhida, a solução teria três concentradores em uma mesma rota; contrariando portanto as convenções inicialmente propostas por O’Kelly (1987).

Quando a estrutura de custos desrespeita as desigualdades triangulares de distância e quando se computam os custos de congestionamento nos concentradores, essas situações podem ocorrer com maior frequência. Em função disso, Marín et al. propõem alterações na

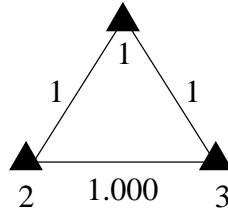


Figura 2.5: Exemplo de uma solução com três concentradores em uma rota.

formulação de Skorin-Kapov et al. (1996), na de Ernst e Krishnamoorthy (1998) e na de Hamacher et al. (2000).

### 2.3.7.1 Reformulação de Skorin-Kapov et al.

Marín et al. adicionam as restrições (2.44) e (2.45) a fim de evitar um número maior do que dois concentradores em uma rota, originando:

$$\min \sum_k a_k y_k + \sum_i \sum_j \sum_k \sum_m w_{ij} c_{ijkm} x_{ijkm} \quad (2.40)$$

sujeito a:

$$\sum_k x_{ijkm} \leq y_m \quad \forall i, j \in N, m \in K \quad (2.41)$$

$$\sum_m x_{ijkm} \leq y_k \quad \forall i, j \in N, k \in K \quad (2.42)$$

$$\sum_k \sum_m x_{ijkm} = 1 \quad \forall i, j \in N \quad (2.43)$$

$$\sum_m \sum_{k \neq i} x_{ijkm} + y_i \leq 1 \quad \forall i, j \in N \quad (2.44)$$

$$\sum_k \sum_{m \neq j} x_{ijkm} + y_j \leq 1 \quad \forall i, j \in N \quad (2.45)$$

$$x_{ijkm} \geq 0 \quad \forall i, j \in N, k, m \in K \quad (2.46)$$

$$y_k \in \{0, 1\} \quad \forall k \in K \quad (2.47)$$

As novas restrições (2.44) garantem que quando há um concentrador instalado na origem  $i$  ( $y_i = 1$ ), então o fluxo entre  $i$  e  $j$  é roteado por apenas um único concentrador adicional ou enviado diretamente ao ponto  $j$ . Da mesma forma, quando há um concentrador instalado no destino  $j$  ( $y_j = 1$ ), as restrições (2.45) asseguram que apenas um único concentrador adicional é usado.

### 2.3.7.2 Reformulação Ernst e Krishnamoorthy

Marín et al. (2006) destacam que a formulação de Ernst e Krishnamoorthy (1998), além de possuir uma relaxação linear fraca, permite a ocorrência de soluções ótimas que possuem

fluxos passando por pontos intermediários que não são concentradores. Através de um exemplo simples, Marín et al. mostram também que é possível ocorrer soluções nas quais o fluxo atravessa até mesmo cinco concentradores. A reformulação sugerida é dada por:

$$\min \sum_k a_k y_k + \sum_i \left[ \sum_k \mathcal{X} c_{ik} z_{ik} + \sum_k \sum_m \alpha c_{km} f_{ikm} + \sum_m \sum_j \delta c_{mj} x_{imj} \right] \quad (2.48)$$

s. a:

$$\sum_m x_{imj} = w_{ij} \quad \forall i, j \in N \quad (2.49)$$

$$\sum_m f_{ikm} + \sum_j x_{ikj} - \sum_m f_{imk} - z_{ik} = 0 \quad \forall i \in N, k \in K \quad (2.50)$$

$$z_{ik} \leq \sum_j w_{ij} y_k \quad \forall i \in N, k \in K \quad (2.51)$$

$$x_{imj} \leq w_{ij} y_m \quad \forall i, j \in N, m \in K \quad (2.52)$$

$$\sum_m f_{ikm} \leq \sum_j w_{ij} y_k \quad \forall i \in N, k \in K \quad (2.53)$$

$$\sum_k f_{ikm} \leq \sum_j w_{ij} y_m \quad \forall i \in N, m \in N \quad (2.54)$$

$$\sum_m f_{ikm} \leq z_{ik} \quad \forall i \in N, k \in N \quad (2.55)$$

$$z_{ii} \geq \sum_j w_{ij} y_i \quad \forall i \in N \quad (2.56)$$

$$x_{ijj} \geq w_{ij} y_j \quad \forall i, j \in N \quad (2.57)$$

$$z_{ik} \geq 0 \quad \forall i \in N, k \in K \quad (2.58)$$

$$f_{ikm} \geq 0 \quad \forall i \in N, k, m \in K \quad (2.59)$$

$$x_{imj} \geq 0 \quad \forall i, j \in N, m \in K \quad (2.60)$$

$$y_k \in \{0, 1\} \quad \forall k \in K \quad (2.61)$$

As restrições (2.26) são redundantes, não sendo portanto incluídas no modelo. As restrições (2.49)-(2.52) possuem o mesmo significado da seção 2.3.5. As restrições (2.53) e (2.54) garantem que só há fluxo nas variáveis  $f_{ikm}$  se há concentradores instalados em  $k$  e em  $m$ , respectivamente. As restrições (2.55) asseguram que o fluxo originado em  $i$  e passando pelos concentradores em  $k$  e em  $m$ , variáveis  $f_{ikm}$ , seja limitado pelo fluxo que sai da mesma origem através do concentrador em  $k$ , variáveis  $z_{ik}$ . As restrições (2.56) implicam que quando o ponto  $i$  possui um concentrador instalado, então existe um fluxo inicial passando por esse concentrador em  $i$ . O mesmo ocorre quando o ponto  $j$  tem um concentrador instalado, nesse caso, o fluxo inicial em  $j$  é assegurado pelas restrições (2.57).



### 2.3.7.3 Reformulação de Hamacher et al.

Interessados em restrições do tipo *set packing*, nas quais a soma de um grande número de variáveis é menor ou igual a 1, Marín et al. obtêm restrições do tipo *clique* a partir da formulação original de Skorin-Kapov et al. (1996), resultando em:

$$\min \sum_k a_k y_k + \sum_i \sum_j \sum_k \sum_m w_{ij} c_{ijkm} x_{ijkm} \quad (2.62)$$

s. a:

$$\sum_{k \neq i} \sum_m x_{ijkm} + y_i \leq 1 \quad \forall i, j \in N \quad (2.63)$$

$$\sum_k \sum_{m \neq j} x_{ijkm} + y_j \leq 1 \quad \forall i, j \in N \quad (2.64)$$

$$x_{ijkk} + \sum_{m \neq k} x_{ijkm} + \sum_{m \neq k} x_{ijmk} \leq y_k \quad \forall i, j \in N, k \in K \quad (2.65)$$

$$\sum_k \sum_m x_{ijkm} = 1 \quad \forall i, j \in N \quad (2.66)$$

$$x_{kimk} = x_{ikkm} = 0 \quad \forall i \in N, k, m \in K \mid k \neq m, i \neq k \quad (2.67)$$

$$x_{ijkm} \geq 0 \quad \forall i, j \in N, k, m \in K \quad (2.68)$$

$$y_k \in \{0, 1\} \quad \forall k \in K \quad (2.69)$$

A formulação (2.62)-(2.69) é semelhante à formulação de Hamacher et al., a menos das restrições (2.63) e (2.64), usadas para limitar em dois o número máximo de concentradores presentes em uma rota, caso haja concentradores instalados nos pontos  $i$  ou  $j$ , respectivamente; e das restrições (2.67) que anulam algumas variáveis da formulação. A relaxação linear dessa formulação é tão boa quanto a de Hamacher et al., porém os tempos de computação são maiores.

Na próxima seção 2.4, um exame dos métodos de resolução do problema de localização de concentradores com atribuição múltipla é apresentado.

## 2.4 Métodos de Resolução

Desde os primeiros trabalhos de O'Kelly (1986, 1987), um número cada vez maior de pesquisadores tem estudado os problemas de sistemas eixo-raio. Alguns exames da literatura (Campbell, 1994a,b; Campbell et al., 2002) têm mostrado a constante evolução das formulações e dos métodos de resolução para os problemas da área.

Uma abordagem comum na resolução do problema é a decomposição do mesmo em dois subproblemas. O primeiro envolve encontrar a melhor configuração de concentradores possível; enquanto o segundo identifica quais as melhores rotas para enviar os fluxos desde as origens até os destinos, através da estrutura de concentradores propostas. Nesse sentido, alguns métodos heurísticos e exatos se destacam na literatura.

Klincewicz (1991) propõe duas heurísticas diferentes e uma meta-heurística, *busca tabu*, para achar boas soluções para o problema em um tempo computacional relativamente curto para a época. Skorin-Kapov e Skorin-Kapov (1994) propõem um algoritmo também baseado em *busca tabu*, porém mais eficiente do que o proposto por Klincewicz. Em seu trabalho, Aykin (1994) desenvolve um método de enumeração para abordar o que viria a ser sua formulação (Aykin, 1995). Campbell (1996), por sua vez, apresenta uma heurística de troca gulosa que obtém bons resultados em face aos tempos computacionais apresentados. Usando *simulated annealing* para obter limites superiores dentro de um algoritmo de *branch-and-bound*, Ernst e Krishnamoorthy (1996) obtêm resultados muito bons a um tempo computacional muito baixo.

Em relação aos métodos exatos, Klincewicz (1996) é um dos primeiros autores a abordar o problema sob uma visão de otimalidade ou  $\varepsilon$ -ótimo; em oposição à grande maioria dos métodos heurísticos presentes na literatura (Campbell et al., 2002). Isso é parcialmente explicado pelo número relativamente grande de variáveis, isto é, variáveis com quatro índices, e da proximidade da área com o problema quadrático de atribuição O’Kelly (1987). Klincewicz, usando a formulação de Campbell (1996), desenvolve uma Relaxação Lagrangeana baseada no método de aproximação de subida dual (*dual ascent approximation*). Pirkul e Schilling (1998) também desenvolvem uma Relaxação Lagrangeana ajustando os multiplicadores via subgradiente, porém o fazem usando a formulação de Skorin-Kapov et al. (1996). Mais recentemente, Mayer e Wagner (2002) empregam o método de Klincewicz (1996) na formulação de Skorin-Kapov et al. (1996) obtendo um desempenho muito superior ao conseguido por Klincewicz.

Conforme mostrado por O’Kelly e Bryan (1998), o uso de um fator constante de desconto (o parâmetro  $\alpha$ ) para representar os efeitos de economia de escala é uma simplificação míope da realidade. Essa simplificação permite soluções ótimas com fluxos muito pequenos e descontos grandes nas conexões entre concentradores; contrariando assim as aplicações reais, nas quais o desconto oferecido depende da quantidade de fluxo sendo transportado entre concentradores.

Com o objetivo de evitar esse tipo de soluções, O’Kelly e Bryan propõem então uma formulação, conhecida como FLOWLOC (veja o capítulo 4), na qual o fator constante de desconto é substituído por uma função côncava linear por partes e dependente da quantidade de fluxo presente nas conexões entre concentradores.

Em razão do baixo desempenho computacional da formulação FLOWLOC proposta por O’Kelly e Bryan (1998), Klincewicz (2002) desenvolve um algoritmo de enumeração implícita e duas heurísticas, uma baseada no procedimento de busca aleatória adaptativa gulosa (GRASP) e outra em uma *busca tabu*. Klincewicz demonstra que, para uma dada configuração de concentradores instalados, o problema pode ser transformado em um problema clássico de localização de facilidades não capacitadas. Resultados desanimadores são apresentados para instâncias com até 25 pontos de demanda e com diferentes parâmetros para a função côncava.

Uma formulação mais elaborada é proposta por Racunica e Wynter (2005). Racunica e Wynter propõem um modelo não-linear com funções côncavas de desconto para todas as conexões da rede e linearizado por um número muito grande de variáveis inteiras, mesmo para instâncias de

pequeno porte. A heurística proposta apresenta baixa eficiência, como mostrado nos poucos resultados computacionais realizados.

Os efeitos de congestionamento são outra consideração importante a ser observada no desenho de sistemas eixo-raio. Grove e O’Kelly (1986) são um dos primeiros a apresentar como os atrasos nos sistemas aéreos do tipo eixo-raio são influenciados pela quantidade de fluxo que passa através de um concentrador (congestionamento). Através da simulação em uma estrutura de concentradores dada, Grove e O’Kelly calculam o atraso médio ocorrido quando o fluxo no sistema é aumentado. Um outro trabalho envolvendo congestionamento é o de Marianov e Serra (2003).

Marianov e Serra modelam as redes eixo-raio aéreas como uma rede de filas M/D/c, propondo restrições de capacidade baseadas na probabilidade de clientes estarem esperando no sistema. Devido à complexidade dessas restrições, os autores propõem uma linearização, resolvendo a formulação resultante por um algoritmo de *busca tabu*. Marianov e Serra propõem também uma formulação para alocação de servidores (pistas de pouso) em cada concentrador instalado, porém não apresentam nem resultados de desempenho computacional, nem métodos de resolução para a mesma. Ainda usando os conceitos de teoria das filas, Kara e Tansel (2003) propõem duas formulações lineares interessantes que minimizam os atrasos excessivos, computando para tal os tempos de tráfego e de transiente durante a carga e descarga nos concentradores.

Muitos pesquisadores abordam ainda os efeitos de congestionamento através de restrições de capacidade de fluxo (Aykin, 1994; Ernst e Krishnamoorthy, 1999; Ebery et al., 2000). Essa simplificação é feita para facilitar o uso de algoritmos do tipo *branch-and-bound* e heurísticas de caminho mínimo. Contudo, as soluções obtidas por essas formulações e por esses métodos nem sempre captam com exatidão a natureza exponencial dos efeitos de congestionamento, isto é, quanto maior o congestionamento, maiores são os custos envolvidos.

Elhedhli e Hu (2005) são um dos primeiros autores a considerarem os efeitos de congestionamento explicitamente na função objetivo do problema de localização de concentradores com atribuição simples. Através de uma função convexa (lei de potência) usada por órgãos gestores de sistemas de transportes (Gillen e Levinson, 1999), que estima os custos de atraso em função do fluxo que passa pelos concentradores instalados, Elhedhli e Hu apresentam uma formulação não-linear, seguido de uma linearização e resolução por Relaxação Lagrangeana. Devido a flexibilidade da abordagem, outras funções convexas de congestionamento podem ser utilizadas, como por exemplo, a função apresentada por Kleinrock (1993), muito empregada em redes de telecomunicações.

Os algoritmos de resolução propostos neste trabalho são todos baseados no método de decomposição de Benders (Benders, 1962), assunto da próxima seção 2.5.

## 2.5 O Método de Decomposição de Benders

Benders (1962) propõe um método de particionamento para resolução de problemas de programação inteira mista linear e não linear. Na abordagem, Benders define um algoritmo

de relaxação que divide o problema original em dois mais simples: o problema mestre e o subproblema.

O problema mestre é uma versão relaxada do problema original com o conjunto de variáveis inteiras e suas respectivas restrições; enquanto que o subproblema é o problema original com os valores das variáveis inteiras temporariamente fixadas pelo problema mestre.

O algoritmo resolve cada um dos dois problemas mais simples de forma iterativa. A cada ciclo, uma nova restrição, conhecida como corte de Benders, é adicionada ao problema mestre. Essa nova restrição, originada a partir da resolução do dual do subproblema, permite estimar limites inferiores para o problema original. O algoritmo cicla até as funções objetivo do problema mestre e do subproblema serem iguais. Ao final do processo, tem-se a solução ótima do problema inteiro misto original.

A eficiência computacional da decomposição de Benders na resolução de alguns problemas de larga escala pode ser verificada no trabalho pioneiro de Geoffrion e Graves (1974), no problema de desenho de redes não capacitadas de Magnanti et al. (1986), no problema de atribuição de vagões e locomotivas de Cordeau et al. (2000, 2001), no problema de gerenciamento de recursos hídricos não convexos de Cai et al. (2001), no desenho de sistemas de manufatura celular de Heragu e Chen (1998), na resolução de problemas estocásticos de transporte e localização de França e Luna (1982) e no problema de desenho de redes *multicasting* de Miranda (2004) e Randazzo e Luna (2001).

De forma genérica, dada a formulação abaixo:

$$\begin{aligned} \min \quad & c^\top x + \bar{c}^\top y \\ \text{sujeito a:} \quad & A x + B y \geq b \\ & D y \geq \bar{b} \\ & x \geq 0 \\ & y \geq 0 \quad \text{e inteiro} \end{aligned}$$

onde  $A$  é uma matriz  $m \times n$ ,  $x$  e  $c$  são vetores de tamanho  $n$ , representando as variáveis contínuas e os custos associados a estas. As matrizes  $B$  e  $D$  têm tamanho igual a  $m \times p$  e a  $q \times p$ , respectivamente, enquanto  $y$  e  $\bar{c}$  são vetores de tamanho  $p$ , representando as variáveis inteiras e os custos associados a estas. Finalmente, os vetores  $b$  e  $\bar{b}$  possuem dimensão  $m$  e  $q$ .

Este problema pode ser expresso como:

$$\min_{y \in Y} \left\{ \bar{c}^\top y + \min_{x \geq 0} \left\{ c^\top x : A x \geq b - B y \right\} \right\} \quad (2.70)$$

onde  $Y = \{y \mid D y \geq \bar{b}, y \geq 0 \text{ e inteiro}\}$ .

Observando a minimização interna, tem-se um problema de programação linear que pode ser dualizado através da associação das variáveis duais  $u$  às restrições  $A x \geq b - B y$ , obtendo-se:

$$\max_{u \geq 0} \left\{ (b - B y)^\top u : A^\top u \leq c \right\}$$

Este problema é conhecido como o subproblema da decomposição de Benders. Através da teoria da dualidade, as formulações duais e primais podem ser usadas de forma intercalada, logo pode-se escrever (2.70) como:

$$\min_{y \in Y} \left\{ \bar{c}^\top y + \max_{u \geq 0} \left\{ (b - B y)^\top u : A^\top u \leq c \right\} \right\} \quad (2.71)$$

O subproblema da decomposição de Benders possui uma característica interessante. Sua região de viabilidade independe da escolha das variáveis  $y$ . Então, definindo-se o conjunto  $\Omega = \{u \mid A^\top u \leq c\}$  como sendo o espaço de viabilidade do subproblema dual e assumindo que  $\Omega$  é não-vazio, caso contrário o problema primal seria inviável ou ilimitado, têm-se que  $\Omega$  é composto de pontos extremos  $u^i$  ( $\forall i = 1, \dots, N$ ), i. e. representando soluções duais limitadas, e raios extremos  $u^r$  ( $\forall r = 1, \dots, Q$ ), i. e. representando soluções duais ilimitadas.

A presença de soluções duais ilimitadas é indesejada uma vez que é o resultado de problemas primais inviáveis. Como  $\Omega$  é um poliedro e possui um número finito de geradores, pode-se então eliminar os valores de  $y$  que geram tais soluções considerando explicitamente as seguintes restrições obtidas a partir do lema de Farkas:

$$(b - B y)^\top u^r \leq 0 \quad r = 1, \dots, Q$$

O problema (2.71) pode ser agora reformulado obtendo-se:

$$\begin{aligned} \min \quad & \bar{c}^\top y + \max \left\{ (b - B y)^\top u^i : i = 1, \dots, N \right\} \\ \text{sujeito a:} \quad & (b - B y)^\top u^r \leq 0 \quad r = 1, \dots, Q \\ & y \in Y \end{aligned}$$

A solução do problema interno de maximização corresponde a um dos pontos extremos de  $\Omega$ . Este problema de maximização pode ser substituído na formulação acima com o auxílio de uma variável contínua  $\eta$ . O seguinte o problema mestre da decomposição de Benders é obtido.

$$\begin{aligned}
& \min \quad \bar{c}^\top y + \eta \\
& \text{sujeito a:} \\
& \quad \eta \geq (b - B y)^\top u^i \quad \forall i = 1, \dots, N \quad (2.72) \\
& \quad (b - B y)^\top u^r \leq 0 \quad \forall r = 1, \dots, Q \quad (2.73) \\
& \quad y \in Y
\end{aligned}$$

Infelizmente, o problema mestre é de difícil resolução uma vez que possui uma restrição para cada ponto e raio extremo do poliedro  $\Omega$ . Essa característica pode resultar em um número muito elevado de restrições dependendo da dimensão do problema. Entretanto, apenas uma pequena parcela dessas restrições estão ativas na solução ótima do problema original. Dessa forma, o problema mestre pode ser resolvido de forma relaxada, gerando-se as restrições (2.72) e (2.73) iterativamente.

A princípio, o problema mestre é resolvido apenas com o conjunto  $Y$  de restrições. A cada ciclo, resolve-se o problema mestre obtendo-se configurações temporárias das variáveis  $y$ . A cada configuração, um subproblema é resolvido, podendo-se obter ou um raio extremo, neste caso uma restrição (2.73) é adicionada ao problema mestre, ou um ponto extremo, neste caso uma restrição (2.72) é inserida no problema mestre. Completa-se assim um ciclo ou uma iteração do método de decomposição de Benders.

A solução ótima do problema mestre fornece um limite inferior, enquanto a conjunção de soluções do problema mestre e do primal do subproblema provê um limite superior do problema original. O problema mestre e o subproblema são resolvidos de forma cíclica (iterativamente) até os limites inferior e superior convergirem, obtendo-se então a solução ótima do problema original.

As restrições (2.72) e (2.73) são conhecidas, respectivamente, como os cortes tipo I e II de Benders. Devido a geração iterativa das restrições a decomposição de Benders é considerada como sendo a versão dual do método de geração de colunas Dantzig-Wolfe.

No presente trabalho, a localização de concentradores com atribuição múltipla e com o fator constante de desconto é abordado no capítulo 3. Neste, métodos exatos baseados na decomposição de Benders (Benders, 1962) são apresentados na resolução da formulação de Hamacher et al. (2000). A escolha desse modelo se deve à qualidade de seu limite de programação linear e ao número reduzido de restrições. Além disso, a presença de apenas um ou dois concentradores em uma rota Um modelo mais interessante para abordar função côncava linear por partes da economia de escala é apresentado no capítulo 4. Um algoritmo, também baseado na decomposição de Benders, é usado na resolução de instâncias consideradas “*fora do alcance*” de métodos exatos. Finalmente, os efeitos de congestionamento são considerados explicitamente na função objetivo, como a abordagem feita por Elhedhli e Hu (2005), no capítulo 5.

## Capítulo 3

# Localização de Concentradores com Atribuição Múltipla

*“Continuous effort - not strength or intelligence - is the key to unlocking our potential.”*

Winston Churchill

No presente capítulo, três algoritmos eficientes, baseado no método de decomposição de Benders (Benders, 1962), são apresentados para resolver o problema de localização de concentradores com atribuição múltipla. Apesar da existência de formulações que corrijam a presença de mais de dois concentradores em uma dada rota para um par de origem-destino  $i - j$ , veja seção 2.3.7, utiliza-se aqui a formulação proposta por Hamacher et al. (2000). Essa escolha se deve à qualidade do limite de programação linear e dos testes computacionais terem sido realizados usando-se instâncias padrões da literatura que respeitam as desigualdades triangulares.

Os três algoritmos desenvolvidos são baseados em diferentes versões do método de decomposição de Benders. O primeiro é uma especialização do método clássico proposto por Benders (1962); enquanto que o segundo é uma implementação da versão multicortes apresentada por Birge e Louveaux (1988); sendo, finalmente, o terceiro uma variante semelhante à desenvolvida por Geoffrion e Graves (1974) que obtém soluções  $\varepsilon$ -ótimas.

Os algoritmos propostos resolvem o problema original de forma iterativa: em um nível superior, conhecido como problema mestre (PM), as decisões de localização de concentradores são realizadas, isto é, os “eixos” são instalados; enquanto, que em um nível inferior, chamado de subproblema (SP), a determinação dos “raios” é feita. O primeiro é um problema inteiro misto de programação matemática; enquanto que o segundo é um problema de transporte com variáveis contínuas, resolvido facilmente por inspeção.

Além disso, a grande maioria dos trabalhos presentes na literatura limita o número de pontos candidatos a se instalar um concentrador. Em outras palavras, fazem com que  $|K| \ll |N|$ . Neste capítulo, essa limitação é desconsiderada, ou seja, considera-se  $K \subseteq N$ .

Esse capítulo está organizado da seguinte forma. Na seção 3.1, os três algoritmos são desenvolvidos, sendo mostrado também como a solução ótima do subproblema pode ser facilmente obtida por inspeção. Em seguida, na seção 3.2, os testes computacionais realizados e a

discussão dos resultados são apresentados. Os comentários finais encerram o capítulo, seção 3.3.

### 3.1 Algoritmos de Decomposição de Benders

A chave para uma boa implementação do método de decomposição de Benders está em ter uma formulação matemática com um bom limite de programação linear, em ajustar o corte de Benders e resolver o subproblema de forma eficiente e rápida, assuntos da próxima seção.

#### 3.1.1 Problema Mestre e Subproblema

A formulação (2.35)-(2.39), proposta por Hamacher et al. (2000), possui um limite de programação linear muito bom, sendo o *gap* de integralidade igual a zero para muitas instâncias de pequeno porte. Além disso, ao se fixar o vetor de variáveis inteiras  $y$ , de forma que uma solução viável seja obtida para a formulação original, pode-se decompor o problema resultante em problemas de caminho mínimo, um para cada par de origem-destino  $i - j$ . Essas duas características contribuem para o bom desempenho do método de decomposição de Benders para resolver o problema de localização de concentradores com atribuição múltipla.

Em outras palavras, fixado um  $y = y^h$  viável, em uma dada iteração  $h$  do algoritmo, o seguinte problema linear primal é obtido:

$$\min \sum_i \sum_j \sum_k \sum_m w_{ij} c_{ijkm} x_{ijkm} + s^h \quad (3.1)$$

sujeito a:

$$\sum_{m \mid m \neq k} x_{ijkm} + \sum_m x_{ijmk} \leq y_k^h \quad \forall i, j \in N, k \in K \quad (3.2)$$

$$\sum_k \sum_m x_{ijkm} = 1 \quad \forall i, j \in N \quad (3.3)$$

$$x_{ijkm} \geq 0 \quad \forall i, j \in N, k, m \in K \quad (3.4)$$

onde  $s^h$  é o custo de instalação dos concentradores associado ao vetor  $y$ .

Através da função objetivo dual da formulação (3.1)-(3.4), pode-se construir uma restrição conhecida como corte de Benders que é adicionada ao problema mestre a cada iteração. Associando então as variáveis duais  $u_{ijk}$  e  $\rho_{ij}$  às restrições (3.2) e (3.3), respectivamente, tem-se o seguinte problema linear dual para iteração  $h$ :



$$\max \sum_i \sum_j \rho_{ij} - \sum_i \sum_j \sum_k y_k^h u_{ijk} \quad (3.5)$$

sujeito a:

$$\rho_{ij} - u_{ijk} - u_{ijm} \leq w_{ij} c_{ijkm} \quad \forall i, j \in N, k \neq m, k, m \in K \quad (3.6)$$

$$\rho_{ij} - u_{ijk} \leq w_{ij} c_{ijkk} \quad \forall i, j \in N, k \in K \quad (3.7)$$

$$u_{ijk} \geq 0 \quad \forall i, j \in N, k \in K \quad (3.8)$$

$$\rho_{ij} \in \mathbb{R} \quad \forall i, j \in N \quad (3.9)$$

O problema dual (3.5)-(3.9) é um problema linear que pode ser facilmente resolvido por inspeção, assunto da seção 3.1.5. A partir da função objetivo dual (3.5), dado que os valores das variáveis duais de uma dada iteração  $h$  tenham sido calculados, pode-se construir a seguinte restrição (3.10), conhecida como corte de Benders, a ser adicionada ao problema mestre.

$$\eta + \sum_i \sum_j \sum_k u_{ijk}^h y_k \geq \sum_i \sum_j \rho_{ij}^h \quad (3.10)$$

onde  $u_{ijk}^h$  e  $\rho_{ij}^h$  são os valores ótimos duais obtidos após a resolução do subproblema em uma dada iteração  $h$ ; e  $\eta \geq 0$  é um subestimador do custo de transporte. Portanto, o problema mestre é dado por:

$$\min \eta + \sum_k a_k y_k \quad (3.11)$$

sujeito a:

$$\eta + \sum_i \sum_j \sum_k u_{ijk}^h y_k \geq \sum_i \sum_j \rho_{ij}^h \quad h = 1, \dots, H \quad (3.12)$$

$$\sum_k y_k \geq 1 \quad (3.13)$$

$$\eta \geq 0 \quad (3.14)$$

$$y_k \in \{0, 1\} \quad \forall k \in K \quad (3.15)$$

onde  $H$  é o número máximo de iterações do algoritmo. A restrição (3.13) é adicionada ao problema mestre para evitar a geração de soluções inviáveis. Essa restrição garante que ao menos um concentrador será instalado. Na próxima seção 3.1.2, a formalização do algoritmo de decomposição de Benders clássico é apresentada.

### 3.1.2 Algoritmo de Decomposição de Benders Clássico

O algoritmo baseado no método de decomposição clássica de Benders é formalmente organizado abaixo, onde  $LS$  é o limite superior,  $LI$  é o limite inferior,  $z_{PM}^*$  e  $z_{SP}^*$  são as soluções

ótimas obtidas resolvendo-se o problema mestre e o subproblema correntes, respectivamente:

---

<b>Algoritmo BD1 (Benders Clássico)</b>	
<b>PASSO 1</b>	Faça $LS = +\infty$ , $LI = -\infty$ .
<b>PASSO 2</b>	Se $LS = LI$ , <b>PARE</b> . Fim da execução, a solução ótima do problema original (2.35)-(2.39) foi obtida.
<b>PASSO 3</b>	Resolva o problema mestre (3.11)-(3.15), obtendo $z_{PM}^*$ e os valores ótimos das variáveis inteiras $y$ .
<b>PASSO 4</b>	Faça $LI = z_{PM}^*$ e atualize os valores de $y$ no subproblema dual (3.5)-(3.9).
<b>PASSO 5</b>	Resolva o subproblema dual (3.5)-(3.9), obtendo $z_{SP}^*$ e os valores ótimos das variáveis duais.
<b>PASSO 6</b>	Adicione o novo corte de Benders ao problema mestre (3.11)-(3.15) usando (3.10)
<b>PASSO 7</b>	Se $z_{SP}^* + \sum_k a_k y_k < LS$ , faça $LS = z_{SP}^* + \sum_k a_k y_k$ . Vá para o <b>PASSO 2</b>

---

Alguns comentários são necessários sobre o algoritmo acima. Como a restrição (3.13) foi adicionada ao problema mestre, toda solução gerada pelo problema mestre é viável para o problema original. Então, as soluções obtidas pelo problema mestre são limites inferiores do valor da função objetivo do problema original, formulação (2.35)-(2.39). Este limite inferior é melhorado a cada nova adição do corte de Benders (3.10). Ao convergir para o limite superior, tem-se a solução ótima do problema original.

A eficiência computacional do algoritmo acima depende principalmente de três questões: (i) o número de iterações necessárias para convergência global; (ii) o tempo gasto na resolução do subproblema em cada iteração; (iii) o tempo e o esforço computacional demandados para resolução do problema mestre.

Uma maneira de abordar a questão (i) é adicionar mais de um corte de Benders por iteração. Isso pode ser feito quando:

- (a) o problema dual (3.5)-(3.9) possui múltiplas soluções ótimas e consegue-se adicionar cortes que não são dominados por nenhuma outra restrição presente no problema mestre.
- (b) quando se consegue reformular o corte de Benders, decompondo-o. As restrições (3.10) podem ser decompostas para cada par  $i - j$ , originando novos cortes por iteração.

Enquanto (a) é uma tarefa difícil de ser realizada, envolvendo a solução de um programa linear, veja Magnanti et al. (1986); (b) é apresentado na seção 3.1.3.

A questão (ii) é abordada na seção 3.1.5, na qual se mostra como resolver o subproblema de maneira eficiente. A solução do subproblema deve ser feita com cuidado especial, uma vez que o método de decomposição de Benders é muito sensível às escolhas dos valores das variáveis duais. Caso contrário, um desempenho muito ruim será observado.

Finalmente, o esforço computacional e o tempo demandado para resolução do problema mestre, questão (iii), é um fator crítico do desempenho geral do algoritmo. Apesar de o problema mestre ser de programação inteira mista, pode-se reformulá-lo, transformando-o em um problema inteiro puro, assunto da seção 3.1.4. Entretanto, mesmo assim, um tempo considerável é gasto na obtenção de soluções a cada iteração.

### 3.1.3 Algoritmo de Decomposição de Benders Multicorte

O subproblema (3.5)-(3.9) pode ser dividido em problemas menores, um para cada par  $i - j$ . Explorando esse fato, o corte de Benders (3.10) pode ser reformulado, como proposto por Birge e Louveaux (1988), originando o seguinte conjunto de cortes de Benders (3.16) por iteração.

$$\eta_{ij} + \sum_k u_{ijk}^h y_k \geq \rho_{ij}^h \quad \forall i, j \in N, h = 1, \dots, H \quad (3.16)$$

O problema mestre agora é dado por:

$$\min \sum_i \sum_j \eta_{ij} + \sum_k a_k y_k \quad (3.17)$$

sujeito a:

$$\eta_{ij} + \sum_k u_{ijk}^h y_k \geq \rho_{ij}^h \quad \forall i, j \in N, h = 1, \dots, H \quad (3.18)$$

$$\sum_k y_k \geq 1 \quad (3.19)$$

$$\eta_{ij} \geq 0 \quad \forall i, j \in N \quad (3.20)$$

$$y_k \in \{0, 1\} \quad \forall k \in K \quad (3.21)$$

onde as variáveis  $u_{ijk}^h$  e  $\rho_{ij}^h$  continuam sendo os valores ótimos das variáveis duais da iteração  $h$  e  $\eta_{ij}$  são os subestimadores do custo de transporte para o par  $i - j$ .

O algoritmo de decomposição de Benders multicorte é organizado como:

---

**Algoritmo BD 2 (Benders Multicorte)**

---

- |              |          |  |
|--------------|----------|--|
| <b>PASSO</b> | <b>1</b> | Faça $LS = +\infty$ , $LI = -\infty$ .   |
| <b>PASSO</b> | <b>2</b> | Se $LS = LI$ , <b>PARE</b> . Fim de execução, a solução ótima do problema original (2.35)-(2.39) foi obtida. |
| <b>PASSO</b> | <b>3</b> | Resolva o problema mestre (3.17)-(3.21), obtendo $z_{PM}^*$ e os valores ótimos das variáveis inteiras $y$ . |
| <b>PASSO</b> | <b>4</b> | Faça $LI = z_{PM}^*$ e atualize os valores de $y$ no subproblema dual (3.5)-(3.9).                           |
| <b>PASSO</b> | <b>5</b> | Resolva o subproblema dual (3.5)-(3.9), obtendo $z_{SP}^*$ e os valores ótimos das variáveis duais.          |
| <b>PASSO</b> | <b>6</b> | Adicione os novos cortes de Benders ao problema mestre (3.17)-(3.21) usando o conjunto de restrições (3.16)  |
| <b>PASSO</b> | <b>7</b> | Se $z_{SP}^* + \sum_k a_k y_k < LS$ , faça $LS = z_{SP}^* + \sum_k a_k y_k$ . Vá para o <b>PASSO 2</b>       |
- 

O problema mestre (3.17)-(3.21) têm um número muito grande de restrições quando comparado ao método de decomposição de Benders clássico. Entretanto, espera-se que o número de iterações seja menor (Birge e Louveaux, 1988; Nemhauser e Wolsey, 1988).

### 3.1.4 Algoritmo de Decomposição de Benders $\varepsilon$ -Ótimo

Geoffrion e Graves (1974) apresentam uma variante do método de decomposição de Benders aplicado ao problema de desenho de sistemas de distribuição multi-produto. Como a geração de vários cortes de Benders é necessária antes de se encontrar uma solução ótima do problema original, Geoffrion e Graves mostram que é desnecessário resolver o problema mestre no ótimo a cada iteração, parando-o toda vez que uma solução viável, melhor do que a solução incumbente, é obtida. O problema mestre é então transformado em um problema de busca de viabilidade.

Infelizmente, essa estratégia não fornece mais um limite inferior para o valor da solução ótima do problema original (2.35)-(2.39), requerendo uma abordagem alternativa para o critério de parada do algoritmo. Esse novo critério de parada envolve a reformulação do corte de Benders (3.10) através da introdução um erro marginal predeterminado e aceitável  $\varepsilon > 0$ , e da remoção da variável contínua  $\eta$  do problema mestre.

Observando a função objetivo (3.11) do problema mestre, tem-se que:

$$\begin{aligned} \eta + \sum_k a_k y_k &\leq (LS - \varepsilon) \\ \eta &\leq (LS - \varepsilon) - \sum_k a_k y_k \end{aligned} \quad (3.22)$$

Substituindo  $\eta$  de (3.22) no corte de Benders (3.10), obtém-se:

$$\begin{aligned} (LS - \varepsilon) - \sum_k a_k y_k + \sum_i \sum_j \sum_k u_{ijk}^h y_k &\geq \sum_i \sum_j \rho_{ij}^h \\ \sum_k (\sum_i \sum_j u_{ijk}^h - a_k) y_k &\geq \sum_i \sum_j \rho_{ij}^h - (LS - \varepsilon) \end{aligned} \quad (3.23)$$

A restrição (3.23) é o novo corte de Benders. Este permite o algoritmo terminar toda vez que o problema mestre não conseguir encontrar uma solução viável melhor do que o valor  $(LS - \varepsilon)$ . Então, o algoritmo pára, sendo a solução incumbente uma solução  $\varepsilon$ -ótima do problema original.

A introdução de uma função objetivo mais interessante, como o lado direito da restrição (3.23), pode ser feita, resultando no problema mestre (3.24)-(3.27):

$$\min \sum_k \left( \sum_i \sum_j u_{ijk}^h - a_k \right) y_k \quad (3.24)$$

sujeito a:

$$\sum_k \left( \sum_i \sum_j u_{ijk}^h - a_k \right) y_k \geq \sum_i \sum_j \rho_{ij}^h - (LS - \varepsilon) \quad \forall h = 1, \dots, H \quad (3.25)$$

$$\sum_k y_k \geq 1 \quad (3.26)$$

$$y_k \in \{0, 1\} \quad \forall k \in K \quad (3.27)$$

Geoffrion e Graves (1974) demonstram que, quando  $\varepsilon > 0$ , o corte de Benders (3.25) não gera uma solução inteira já visitada. O algoritmo de decomposição de Benders  $\varepsilon$ -ótimo é assim organizado:

---

**Algoritmo BD3 (Benders  $\varepsilon$ -ótimo)**

---

**PASSO 1** Faça  $LS = +\infty$ .

**PASSO 2** Resolva o problema mestre (3.24)-(3.27). Se a solução é inviável, **PARE!** A solução incumbente é uma solução  $\varepsilon$ -ótima. Senão obtenha os valores ótimos das variáveis inteiras  $y$ .

**PASSO 3** Atualize  $y$  no novo problema dual (3.5)-(3.9).

**PASSO 4** Resolva o subproblema dual (3.5)-(3.9), obtendo  $z_{SP}^*$  e os valores ótimos das variáveis duais.

**PASSO 5** Adicione um novo corte de Benders ao problema mestre (3.24)-(3.27) usando (3.23)

**PASSO 6** Se  $z_{SP}^* + \sum_k a_k y_k < LS$ , faça  $LS = z_{SP}^* + \sum_k a_k y_k$  e salve a solução incumbente. Vá para o **PASSO 2**.

---

Em todos os algoritmos de decomposição de Benders, a resolução rápida e eficiente do problema dual (3.5)-(3.9) é crítica, sendo o assunto da próxima seção 3.1.5.

### 3.1.5 Subproblema

Ao invés de resolver o subproblema (3.5)-(3.9) através dos métodos simplex ou de pontos interiores, um algoritmo especializado pode ser desenvolvido baseando-se nas condições de complementaridade de folga e na seguinte proposição:

**Proposição 3.1** *O problema primal (3.1)-(3.4) é sempre viável e limitado para qualquer  $y_k$  viável fixado que satisfaça as restrições (2.35)-(2.38).*

**Prova.** O problema original (2.35)-(2.39) é modelado para achar o menor custo de instalação dos concentradores e de transporte entre os pares  $i - j$ . Desde que exista pelo menos um concentrador instalado, pode-se facilmente visualizar que a formulação (2.35)-(2.39) é sempre viável para qualquer configuração  $y$  que satisfaça as restrições (2.36)-(2.39).

Através de um algoritmo de caminho mínimo, pode-se determinar qual a melhor rota a ser usada através dos concentradores instalados por cada par  $i - j$ . Os pontos de demanda são portanto implicitamente atribuídos aos concentradores instalados.

Uma vez que os custos de transporte são finitos e não negativos, qualquer solução viável para o problema primal deve ser limitado. Portanto o problema primal é viável e limitado para qualquer  $y$  dado que satisfaça o conjunto de restrições (2.36)-(2.39).  $\square$

**Corolário.** O problema dual (3.5)-(3.9) é viável e limitado, possuindo, pelo menos, uma solução ótima.

**Proposição 3.2** *A solução ótima do problema dual (3.5)-(3.9) pode ser encontrada por inspeção.*

**Prova.** Em transporte de carga, suponha que um transportador cobre pelo frete de carga  $w_{ij}$  e pague pelo uso dos nós de transbordo e pelo custeio do transporte entre os nós de origem e de destino  $i$  e  $j$ . Em transporte aéreo de passageiros, por exemplo,  $\rho_{ij}/w_{ij}$  pode ser interpretado pelo preço da passagem aérea entre  $i$  e  $j$ , enquanto  $u_{ijk}/w_{ij}$  corresponde à taxa de embarque pelo uso da conexão em  $k$ .

Nesses casos, as variáveis duais  $\rho_{ij}$  e  $u_{ijk}$  podem ser interpretadas como o valor total do frete e o valor de transbordo no nó  $k$  de toda a carga  $w_{ij}$  a ser transportada entre os nós  $i$  e  $j$ , respectivamente.

Nessa interpretação, o conjunto de soluções do problema dual (3.5)-(3.9) limita os preços e os custos de forma a não haver lucro para nenhum transportador entre quaisquer pares origem-destino  $i - j$ . Pela complementaridade de folga, os valores de frete e de taxas de transbordo são encontrados em uma solução ótima do problema dual tal que haja um equilíbrio financeiro para todos os transportadores.

Então, da teoria de dualidade e da propriedade de folgas complementares, para qualquer  $y_k^h$  viável dado, em uma iteração  $h$ , tem-se:

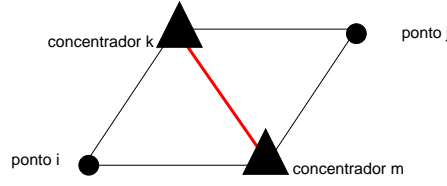
$$\begin{aligned} & \min \left\{ \left( \sum_i \sum_j \sum_k \sum_m w_{ij} c_{ijkm} x_{ijkm}^h \right) \mid (3.2)-(3.4) \text{ é válido} \right\} \\ & = \max \left\{ \sum_i \sum_j \rho_{ij}^h - \sum_i \sum_j \sum_k u_{ijk}^h y_k \mid (3.6)-(3.9) \text{ é válido} \right\} \end{aligned}$$

Para um dado par  $i - j$ , considerando os concentradores instalados, tem-se um problema simples de roteamento a ser resolvido que pode ser visto na figura 3.1:

A rota ótima é dada por:

$$\min_{k, m \in K} \left\{ w_{ij} c_{ijkm} \mid y_k^h = y_m^h = 1 \right\}$$

Dessa forma, as variáveis duais  $\rho_{ij}^h$  podem ser calculadas por:

Figura 3.1: Rotas possíveis para o par  $i - j$ .

$$\rho_{ij}^h = \min_{k,m \in K} \left\{ w_{ij} c_{ijkm} \mid y_k^h = y_m^h = 1 \right\} \quad (3.28)$$

Resultando em um *gap* de dualidade igual a *zero*. Os cálculos dos valores das variáveis duais  $u_{ijk}^h$  são feitos também através da interpretação econômica, sendo consideradas como taxas adicionais que devem ser pagas para se rotear o fluxo do par  $i - j$  através de um concentrador alternativo ainda não instalado em  $k$ . Para se computar os valores mais justos para essas variáveis duais, lembrando que os valores de  $\rho_{ij}^h$  já foram determinados usando (3.28), tem-se o problema remanescente para cada par  $i - j$ :

$$\min \sum_k y_k^h u_{ijk}^h \quad (3.29)$$

sujeito a:

$$u_{ijk}^h + u_{ijm}^h \geq \rho_{ij}^h - w_{ij} c_{ijkm} \quad \forall k \neq m, k, m \in K \quad (3.30)$$

$$u_{ijk}^h \geq \rho_{ij}^h - w_{ij} c_{ijkk} \quad \forall k \in K \quad (3.31)$$

$$u_{ijk}^h \geq 0 \quad \forall k \in K \quad (3.32)$$

De acordo com Magnanti e Wong (1981), a solução do programa (3.29)-(3.32) produz cortes de Benders mais justos. Uma maneira de resolver esse problema eficientemente é observar as restrições (3.31) e (3.32). Elas permitem o cálculo de valores iniciais para  $u_{ijk}^h$ , ou:

$$u_{ijk}^h = \max\{0, (\rho_{ij}^h - w_{ij} c_{ijkk})\} \quad \forall y_k^h = 0, k \in K$$

Respeitando a complementaridade de folga, têm-se ainda  $u_{ijk}^h = 0$  para todo  $y_k^h = 1$ .

As possíveis combinações de rotas usando dois concentradores distintos devem ser agora consideradas para se calcular os valores ótimos das variáveis  $u_{ijk}^h$ . Isto implica em respeitar as restrições (3.30) e em inspecionar, para cada par  $i - j$ , caminhos usando um, caso (i), ou dois, caso (ii), concentradores alternativos ainda não instalados. Então, usando os valores iniciais de  $u_{ijk}^h$ , têm-se para o caso (i):

$$u_{ijk}^h = \max \left\{ u_{ijk}^h, \max_{\substack{m \in K \\ m \neq k, y_m^h = 1}} \{ (\rho_{ij}^h - w_{ij} c_{ijkm}), (\rho_{ij}^h - w_{ij} c_{ijmk}) \} \right\} \quad \forall y_k^h = 0, k \in K$$

Com os valores calculados para o caso (i), analisa-se o caso (ii):

$$u_{ijk}^h = \max \left\{ u_{ijk}^h, \max_{\substack{m \in K \\ m \neq k, y_m^h = 0}} \{ (\rho_{ij}^h - w_{ij} c_{ijkm} - u_{ijm}^h), (\rho_{ij}^h - w_{ij} c_{ijmk} - u_{ijm}^h) \} \right\} \quad \forall y_k^h = 0, k \in K$$

Dessa forma, os valores ótimos das variáveis duais são calculados.  $\square$

O algoritmo de resolução do subproblema é apresentado abaixo:

---

**Algoritmo do Subproblema**

---

**Para cada par  $i - j$**

**PASSO 1** Calcule  $\rho_{ij}^h = \min_{k, m \in K} \{ w_{ij} c_{ijkm} \mid y_k^h = y_m^h = 1 \}$ .

**PASSO 2** Calcule  $u_{ijk}^h = \max\{0, (\rho_{ij}^h - w_{ij} c_{ijkk})\}$ ,  $\forall y_k^h = 0, k \in K$ .

Faça  $u_{ijk}^h = 0$ ,  $\forall y_k^h = 1, k \in K$ .

**PASSO 3** Calcule

$$u_{ijk}^h = \max \left\{ u_{ijk}^h, \max_{\substack{m \in K \\ m \neq k, y_m^h = 1}} \{ (\rho_{ij}^h - w_{ij} c_{ijkm}), (\rho_{ij}^h - w_{ij} c_{ijmk}) \} \right\}, \forall y_k^h = 0, k \in K.$$

**PASSO 4** Calcule

$$u_{ijk}^h = \max \left\{ u_{ijk}^h, \max_{\substack{m \in K \\ m \neq k, y_m^h = 0}} \{ (\rho_{ij}^h - w_{ij} c_{ijkm} - u_{ijm}^h), (\rho_{ij}^h - w_{ij} c_{ijmk} - u_{ijm}^h) \} \right\}, \forall y_k^h = 0, k \in K.$$

**Fim Para**

---

O esforço computacional de resolução do subproblema, em cada iteração  $h$ , é considerável. A complexidade do algoritmo proposto é igual a  $o(n^4)$ . Apesar da teoria indicar que os programas inteiros, como o problema mestre, são mais difíceis; o tempo computacional de resolução do subproblema, verificado nos testes computacionais, é maior do que o tempo gasto na solução do problema mestre.

## 3.2 Resultados Computacionais

Os testes computacionais foram realizados usando dois conjuntos padrões de instâncias da literatura: (a) CAB do Conselho de Aviação Civil dos Estados Unidos da América (O’Kelly, 1987) e (b) AP do serviço Postal Australiano (Ernst e Krishnamoorthy, 1996, 1999). As instâncias CAB incluem testes com 10, 15, 20 e 25 pontos de demanda, enquanto que testes de 10 até 200 pontos de demanda podem ser gerados a partir das instâncias AP.

Muitos estudos da literatura (O’Kelly e Bryan, 1998; Elhedhli e Hu, 2005; Ebery et al., 2000; Boland et al., 2004) usam essas instâncias para medir o desempenho dos métodos de



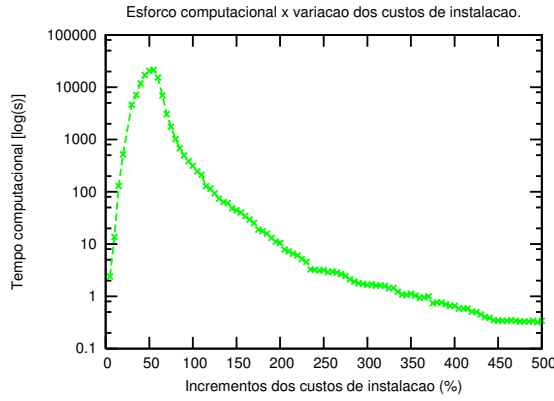


Figura 3.2: Esforço computacional versus variação dos custos de instalação.

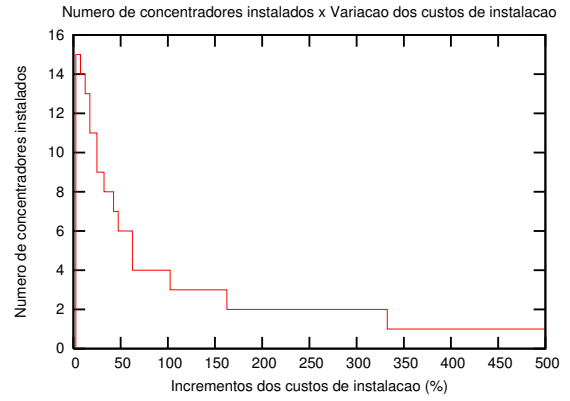


Figura 3.3: Número de concentradores instalados versus variação dos custos de instalação.

resolução. Entretanto, enquanto o conjunto CAB não possui os custos de instalação dos concentradores; o conjunto AP possui apenas os cinquenta primeiros custos de instalação.

Em função dessa limitação, custos fixos foram gerados para todas as instâncias através de uma distribuição normal com média igual a  $f_o$  e coeficiente de variação igual a 40% desse valor. Isso foi feito para representar a variação dos custos de instalação em problemas reais.

O valor  $f_o$  foi calculado como sugerido por Ebery et al. (2000) e representa a diferença normalizada entre um cenário que possui apenas um concentrador imaginário localizado no centro de massa das demandas e outro tendo todos os pontos de demanda como concentradores.

$$f_o = \left( \frac{\sum_i \sum_j (c_{i\ell} + c_{\ell j}) w_{ij} - \sum_i \sum_j \alpha c_{ij} w_{ij}}{p} \right)$$

onde  $\ell$  é o local mais próximo do centro de massa das demandas e  $p$  é um valor inteiro que pode ser usado para estimar o número de concentradores instalados na solução ótima.

Ainda como sugerido por Ebery et al. (2000), todos os maiores custos de instalação gerados foram atribuídos aos pontos de maior demanda. Em geral, essa atribuição dificulta a seleção dos pontos de demanda nos quais se instalará um concentrador. Além disso, como as matrizes de demanda do conjunto de dados AP são fracionárias, elas foram normalizadas para se manter a coerência com os dados do conjunto CAB.

Os 96 testes gerados foram nomeados como CAB $n.\alpha$  e AP $n.\alpha$ , indicando a origem dos dados usados, o número de pontos de demanda,  $n$ , e o desconto usado  $\alpha$ . O símbolo  $\alpha$  pode assumir valores iguais a 2, 4, 6 e 8 significando descontos de 20%, 40%, 60% e 80%, respectivamente.

Os testes computacionais foram realizados em uma *Sun Blade 100* com um processador *Ultra-SPARC* de 500 MHz e 1 *Gbyte* de memória RAM, executando o sistema operacional Solaris 5.8. Todos os algoritmos foram codificados em C++, sendo o CPLEX 7.0, usando a configuração padrão de instalação, utilizado para resolver os problemas mestres. Um limite

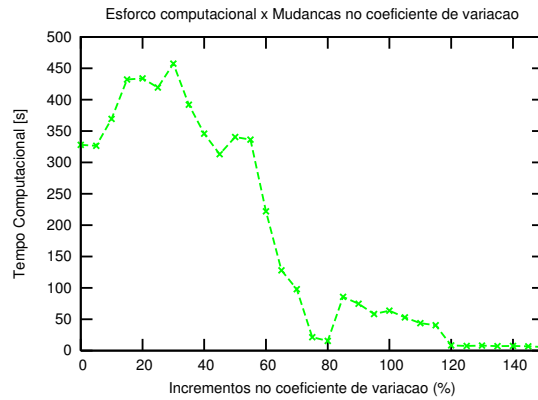


Figura 3.4: Esforço computacional versus mudança no coeficiente de variação dos custos fixos.

de 10 horas de execução de máquina foi estabelecido como um critério adicional de parada. O parâmetro  $\varepsilon$  do algoritmo baseado na variante de Geoffrion e Graves foi configurado em 1%.

Os testes realizados apontam a influência dos custos de instalação no esforço computacional do CPLEX 7.0 ao resolver as instâncias de forma monolítica usando a formulação (2.35)-(2.39). A figura 3.2 destaca este esforço ao resolver a instância CAB15.2 variando-se os custos de instalação desde 5% até 500% dos valores originais, em incrementos de 5%. A figura 3.3 apresenta o número de concentradores instalados em cada incremento.

Um aumento de apenas 20% nos custos de instalação, ou seja, uma variação de 50% para 60% dos valores de referência, representa uma mudança de configuração de 6 para 4 concentradores instalados e uma diferença no esforço computacional de uma ordem de grandeza. Em outras palavras, ao se lidar com aplicações reais de larga escala, deve-se conceber uma boa métrica para calcular os custos de instalação, caso esses não estejam disponíveis; ao contrário de prescrevê-los *ad hoc*. Eles têm um impacto muito grande na resolução do problema.

A figura 3.4 apresenta como o esforço computacional também é afetado quando o coeficiente de variação é aumentado desde 0% até 150%. Novamente, a instância CAB15.2 foi usada tendo o valor  $f_o$  como média. O coeficiente de variação representa como os custos de instalação se diferenciam uns dos outros. Pode-se destacar, observando a figura, que o esforço computacional diminui ao se aumentar o coeficiente de variação.

Se os custos de instalação se diferem pouco, o número de locais mais prováveis para se instalar um concentrador aumenta muito, tornando a instância difícil de ser resolvida e vice-versa. A heterogeneidade dos custos de instalação pode induzir a uma possível hierarquização da escolha dos locais como a idealizada por Mateus e Thizy (1999) para o problema de localização em redes.

As tabelas 3.1-3.5 mostram os resultados computacionais dos testes realizados. As instâncias CAB aparentam ser mais difíceis de resolver do que as baseadas no conjunto AP. Isso é verdadeiro tanto para a abordagem monolítica, quanto para os algoritmos baseados na decomposição de Benders.

O CPLEX 7.0 se apresenta como uma ferramenta bastante eficiente para a resolução de

Tabela 3.1: Esforço computacional para o conjunto CAB.

Instância	CPLEX 7.0		Benders Clássico (BD1)		Benders Multicortes (BD2)		$\varepsilon$ -Benders (BD3)	
	Solução	Tempo	#	Tempo	Nr.	Tempo	Nr.	Tempo
	Ótima	CPU [s]	Iter.	CPU [s]	Iter.	CPU [s]	Iter.	CPU [s]
CAB10.2	685709	0.58	43	2.05	20	22.07	88	4.70
CAB10.4	669854	0.56	58	13.53	18	39.07	124	7.45
CAB10.6	597781	0.44	69	15.70	14	13.59	159	12.97
CAB10.8	458985	0.35	62	30.81	9	9.54	75	3.52
CAB15.2	2.71877e+06	5.68	47	46.21	56	5248.60	332	14.86
CAB15.4	2.64529e+06	17.90	51	55.00	54	1341.90	289	25.66
CAB15.6	2.35277e+06	11.20	50	64.95	38	376.75	339	45.86
CAB15.8	1.90899e+06	10.07	62	73.92	19	183.75	234	40.89
CAB20.2	5.56714e+06	17.40	100	75.98	28	3058.00	160	51.83
CAB20.4	5.61875e+06	22.72	130	131.97	34	2155.40	242	22.54
CAB20.6	5.2409e+06	20.82	117	78.93	32	2381.30	343	33.39
CAB20.8	4.56789e+06	25.88	190	89.10	29	2125.80	370	36.50
CAB25.2	8.57026e+06	52.22	170	89.84	25	2429.50	214	27.17
CAB25.4	8.95641e+06	65.74	196	129.20	22	2132.50	191	72.58
CAB25.6	8.5161e+06	93.95	125	188.40	19	3900.00	205	74.50
CAB25.8	7.21453e+06	39.41	152	288.40	20	3501.10	462	88.60

instâncias menores do que 40 pontos de demanda. Entretanto, para instâncias maiores, as limitações de memória de computador podem ser um fator preponderante.

Tabela 3.2: Esforço computacional para o conjunto AP.

Instância	CPLEX 7.0		Benders Clássico (BD1)		Benders Multicortes (BD2)		$\varepsilon$ -Benders (BD3)	
	Solução	Tempo	Nr.	Tempo	Nr.	Tempo	Nr.	Tempo
	Ótima	CPU [s]	Iter.	CPU [s]	Iter.	CPU [s]	Iter.	CPU [s]
AP10.2	68015	1.13	13	0.91	6	2.22	69	3.19
AP10.4	63805	0.87	8	0.29	5	0.78	21	0.71
AP10.6	59596	0.90	7	0.20	5	0.73	8	0.22
AP10.8	55386	0.86	12	0.49	7	1.37	9	0.19
AP20.2	127333	25.69	10	1.10	6	18.57	82	10.79
AP20.4	118151	10.65	9	0.94	6	9.51	86	10.49
AP20.6	108310	9.30	11	1.13	7	17.76	60	6.41
AP20.8	96800	7.59	12	1.25	7	16.42	9	0.75
AP30.2	185713	88.68	12	4.45	9	289.63	54	20.96
AP30.4	171317	85.42	13	4.77	7	130.49	63	24.12
AP30.6	156351	85.54	17	6.27	9	219.60	80	29.52
AP30.8	130898	68.81	10	3.47	6	52.06	37	12.20

Os algoritmos baseados no método de decomposição de Benders resolvem instâncias com até 200 pontos de demanda respeitando o limite de 10 horas proposto. É importante mencionar que é a primeira vez na literatura que se resolve instâncias desses tamanhos no ótimo.

A adequação do método de decomposição de Benders para essa classe de problemas não é uma surpresa, em razão da proximidade dos mesmos com os problemas clássicos de localização de facilidades (Klincewicz, 1996; Hamacher et al., 2000) e o histórico do método Geoffrion e Graves (1974); Magnanti e Wong (1981); Magnanti et al. (1986). Por outro lado, isso não aparenta ser um senso comum na literatura de sistemas eixo-raio, uma vez que são poucos os trabalhos usando essa estratégia de resolução (Racunica e Wynter, 2005).

As figuras 3.5-3.8 mostram uma comparação do desempenho dos três algoritmos na resolução das instâncias do conjunto AP. Comparando as três variantes implementadas, pode-se constatar que os tempos computacionais da versão multicortes (BD2) são maiores do que as outras duas.

Tabela 3.3: Resultandos computacionais para as instâncias do conjunto AP.

Instância	Solução Ótima	Benders Clássico (BD1)		Benders Multicortes (BD2)		$\epsilon$ -Benders (BD3)	
		Nr.	Tempo	Nr.	Tempo	Nr.	Tempo
		Iter.	CPU [s]	Iter.	CPU [s]	Iter.	CPU [s]
AP40.2	307925	18	20.40	7	981.29	69	81.81
AP40.4	287261	10	10.39	6	493.11	50	53.70
AP40.6	266597	13	13.87	7	416.55	88	94.47
AP40.8	239986	15	16.05	12	9.70	12	10.29
AP50.2	368629	16	45.12	16	30588.90	525	1974.38
AP50.4	350496	12	28.95	7	3682.55	68	173.09
AP50.6	325278	11	26.33	8	2478.18	61	148.05
AP50.8	291650	11	26.15	9	3179.42	49	107.59
AP60.2	386250	15	78.07	10	17002.80	18	83.40
AP60.4	353215	13	66.63	8	5897.72	12	52.68
AP60.6	312972	15	77.33	9	3928.76	72	361.34
AP60.8	244630	11	55.45	6	806.55	91	428.98
AP70.2	430065	21	208.24	10	32333.60	14	143.50
AP70.4	381045	16	156.27	7	9130.93	11	107.17
AP70.6	321994	18	176.72	8	4341.53	102	1099.60
AP70.8	236992	14	135.57	7	4665.56	124	1259.21
AP80.2	482852	23	393.21	8	35770.80	13	233.64
AP80.4	428380	19	321.68	9	34492.90	12	211.32
AP80.6	355785	15	249.90	8	12309.40	95	1829.29
AP80.8	261915	14	231.81	6	2228.47	148	2669.24
AP90.2	515012	19	518.06	8	44338.80	27	723.29
AP90.4	454901	14	374.41	8	35785.80	24	633.07
AP90.6	379827	14	376.04	7	10404.80	14	345.54
AP90.8	276640	11	288.37	6	7537.26	152	3786.78
AP100.2	565395	17	707.32	6	33079.40	35	1454.67

Apesar de apresentar um número inferior de iterações para se obter a solução ótima, a adição de  $(n^2 - n)$  cortes por iteração sobrecarrega o problema mestre do algoritmo BD2, tornando os tempos de computação proibitivos. A figura 3.11 ilustra a proporção de tempo gasto pelo problema mestre e pelo subproblema do algoritmo BD2, versão multicortes, na resolução das instâncias AP20.4, AP30.4, AP40.4, AP50.4 e AP60.4. Ao se comparar essas

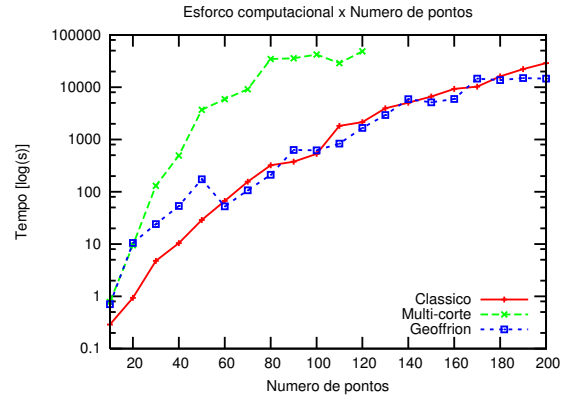
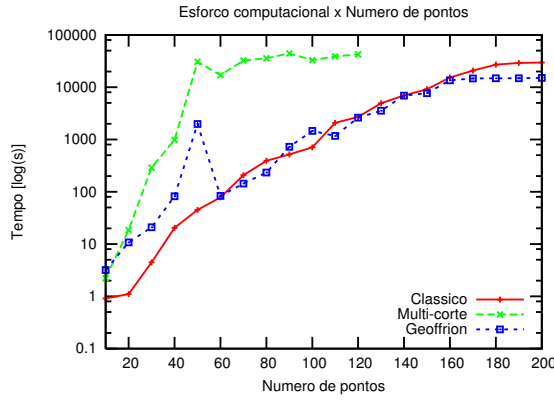


Figura 3.5: Esforço computacional ( $\alpha = 20\%$ ). Figura 3.6: Esforço computacional ( $\alpha = 40\%$ ).

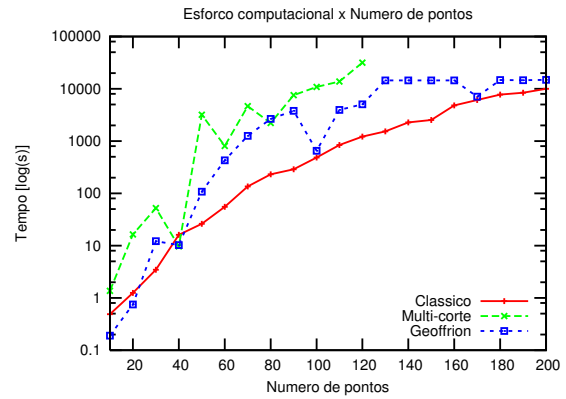
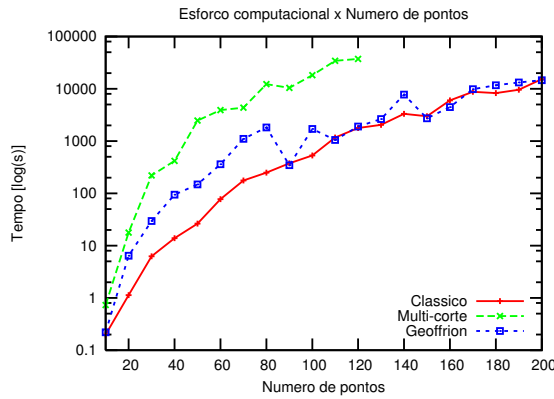


Figura 3.7: Esforço computacional ( $\alpha = 60\%$ ). Figura 3.8: Esforço computacional ( $\alpha = 80\%$ ).

proporções com as apresentadas na figura 3.9 pelo algoritmo BD1, versão clássica, constata-se que os  $(n^2 - n)$  cortes fortalecem o limite inferior (veja figura 3.10), porém a um custo de computação muito grande. Uma boa estratégia, a ser investigada futuramente, pode ser a combinação dos algoritmos BD1 e BD2.

Os cortes de Benders (3.10) são fortes o suficiente para tornar o uso da variante BD3 (Geoffrion e Graves, 1974) desinteressante. A convergência do limite inferior apresentada pelo algoritmo BD1 (veja figura 3.10), o fato do BD3 gastar algumas iterações sem a melhora do limite superior e em razão do problema mestre gastar um tempo menor do que o tempo de resolução do subproblema (veja figura 3.9) justificam a resolução do problema mestre no ótimo.

Em muito dos testes realizados, a versão (BD1) possui tempos de resolução uma ordem de magnitude menor do que as outras duas variantes. Entretanto, o algoritmo BD1 apresenta o efeito de “achatamento” (*flattening*) para *gaps* de otimalidade menores do que 5%, enquanto que a versão BD2 se mostra menos suscetível a esse efeito. Mais uma vez, uma combinação das versões BD1 e BD2 pode ser muito promissora.

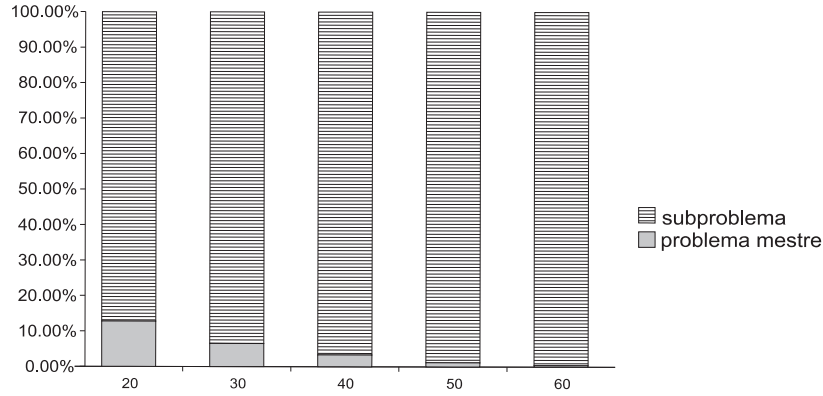


Figura 3.9: Proporção do tempo gasto pelo algoritmo BD1 executando o problema mestre e o subproblema para as instâncias AP20.4, AP30.4, AP40.4, AP50.4 e AP60.4

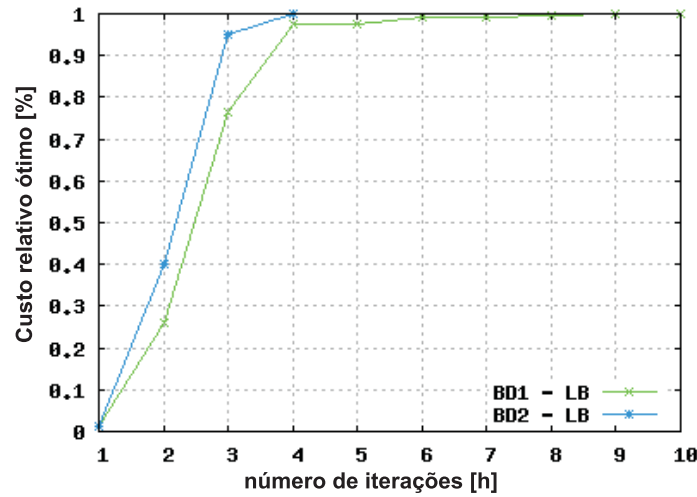


Figura 3.10: Taxa de convergência do limite inferior dos algoritmos BD1 e BD2 para a instância AP30.6

### 3.3 Comentários finais

Nesse capítulo, o papel dos custos de instalação pôde ser observado na resolução do problema de localização de concentradores com atribuição múltipla. Os valores desses custos não devem ser definidos *ad hoc*, mas determinados o mais próximo da realidade possível.

Três variantes do método de decomposição de Benders foram apresentadas. A versão clássica BD1 se mostrou como a mais interessante enquanto ferramenta de engenharia, uma vez que é capaz de resolver instâncias de tamanho 200 em menos de dez horas de computação. Instâncias desse tamanho eram consideradas “fora de alcance” dos métodos exatos. Do ponto de vista do projetista, esperar dez horas para se ter uma solução ótima não é muito, em face da decisão não ser uma tarefa rotineira e envolver grandes quantidades de recursos.

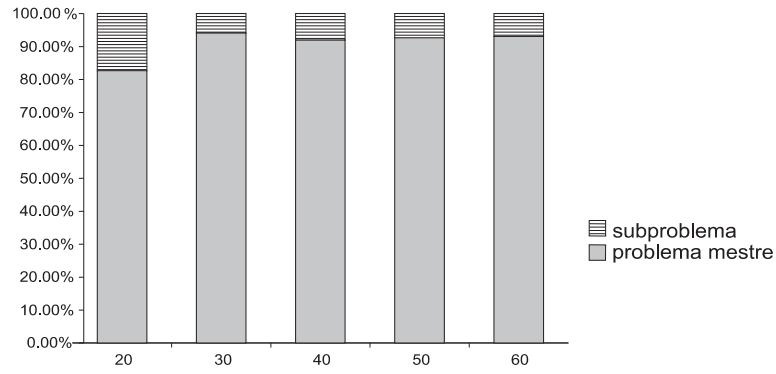


Figura 3.11: Proporção do tempo gasto pelo algoritmo BD2 executando o problema mestre e o subproblema para as instâncias AP20.4, AP30.4, AP40.4, AP50.4 e AP60.4

Tabela 3.4: Resultandos computacionais para as instâncias do conjunto AP.

Instância	Solução Ótima	Benders Clássico (BD1)		Benders Multicortes (BD2)		$\epsilon$ -Benders (BD3)	
		Nr. Iter.	Tempo CPU [s]	Nr. Iter.	Tempo CPU [s]	Nr. Iter.	Tempo CPU [s]
AP100.4	499926	13	530.15	7	42084.50	16	618.65
AP100.6	417504	13	530.41	6	18162.70	41	1706.33
AP100.8	300453	12	485.93	6	10895.30	17	652.48
AP110.2	597530	33	2073.29	6	38783.10	20	1166.61
AP110.4	519661	29	1817.38	6	28978.00	15	836.88
AP110.6	418531	19	1167.57	8	34327.20	20	1047.26
AP110.8	279354	14	843.05	7	13789.80	66	3956.18
AP120.2	643378	30	2706.98	6	42477.20	30	2607.51
AP120.4	559106	24	2149.31	6	48691.10	20	1671.06
AP120.6	447050	20	1776.16	7	37309.50	23	1901.45
AP120.8	296716	14	1218.94	9	31748.20	62	5049.88
AP130.2	721862	40	4948.79	*	*	30	3562.44
AP130.4	622797	32	3953.02	*	*	25	2925.62
AP130.6	478471	17	2049.22	*	*	23	2643.72
AP130.8	318359	13	1535.92	*	*	130	14469.50
AP140.2	772636	41	6973.30	*	*	42	6928.06
AP140.4	665403	30	5069.62	*	*	36	5920.21
AP140.6	518544	20	3331.82	*	*	47	7763.31
AP140.8	345468	14	2281.46	*	*	98	14537.50
AP150.2	823854	41	9132.24	*	*	36	7659.90
AP150.4	709951	30	6652.07	*	*	25	5146.31
AP150.6	529332	14	2992.48	*	*	14	2731.39
AP150.8	345224	12	2534.82	*	*	76	14436.60
AP160.2	815066	52	15286.30	*	*	48	13529.70
AP160.4	697254	32	9320.59	*	*	22	5971.96

(\*) Tempo limite de 10 horas excedido.

A combinação das versões BD1 e BD2 aparenta ser uma oportunidade promissora. Contudo, outras variantes devem ser também exploradas, como o método de decomposição de

Tabela 3.5: Resultandos computacionais para as instâncias do conjunto AP.

Instância	Solução Ótima	Benders Clássico (BD1)		Benders Multicortes (BD2)		$\varepsilon$ -Benders (BD3)	
		Nr.	Tempo	Nr.	Tempo	Nr.	Tempo
		Iter.	CPU [s]	Iter.	CPU [s]	Iter.	CPU [s]
AP160.6	531393	21	6029.17	*	*	17	4473.44
AP160.8	338266	17	4823.31	*	*	60	14466.10
AP170.2	869440	56	20833.80	*	*	42	14697.50
AP170.4	742202	28	10270.00	*	*	41	14616.00
AP170.6	586161	24	8756.42	*	*	28	9845.21
AP170.8	369088	17	6093.01	*	*	22	7052.51
AP180.2	927744	57	27150.00	*	*	33	14875.10
AP180.4	796991	34	16051.60	*	*	30	13629.10
AP180.6	596958	18	8271.67	*	*	26	11724.10
AP180.8	378423	17	7778.87	*	*	33	14727.00
AP190.2	991540	50	29173.10	*	*	27	14881.60
AP190.4	857546	38	22148.60	*	*	27	14982.10
AP190.6	626671	17	9565.40	*	*	24	13249.90
AP190.8	395724	15	8386.79	*	*	27	14602.00
AP200.2	1.08422e+06	40	29524.90	*	*	22	15028.90
AP200.4	925895	39	28967.60	*	*	21	14585.50
AP200.6	685370	21	15298.50	*	*	21	14563.70
AP200.8	425069	14	9970.83	*	*	22	14857.40

(\*) Tempo limite de 10 horas excedido.

Benders Regularizado (Ruszczynski, 1986).

O próximo passo é propor modelos mais realísticos. No capítulo 4, os efeitos de economia de escala dependentes do fluxo são abordados como uma função côncava linear por partes. No capítulo 5, os custos de congestionamento, representados por uma lei de potência em função do fluxo que passa através dos concentradores, são incorporados explicitamente na função objetivo.



## Capítulo 4

# Localização de Concentradores com Economia de Escala Dependente do Fluxo

*“Just remember, once you’re over the hill you begin to pick up speed.”*

Arthur Shopenhauer

O uso de um fator  $\alpha$  constante para representar a economia de escala é uma simplificação da realidade que pode permitir soluções ótimas com fluxos pequenos premiados por descontos elevados. Em muitas aplicações reais, o desconto dado é em função do tráfego que passa pela conexão entre concentradores.

O’Kelly e Bryan (1998) são um dos primeiros a verificar essa anomalia e a propor uma formulação conhecida como FLOWLOC, na qual o fator constante de desconto é substituído por uma função côncava linear por partes. Essa função permite uma melhor representação dos efeitos de economia de escala, uma vez que os mesmos passam a ser então dependentes do fluxo na conexão entre concentradores. A adequação dessa proposta é demonstrada através da resolução de uma instância com 20 pontos de demanda usando um pacote comercial de otimização de programação inteira mista .

Usando a formulação FLOWLOC e preocupado com as mais de 20 horas de computação gastas por O’Kelly e Bryan (1998), Klinecicz (2002) propõe três estratégias de solução para abordar instâncias de até 25 pontos de demanda: duas heurísticas, uma baseada em *busca tabu* e outra na busca aleatória adaptativa gulosa (GRASP), e uma enumeração implícita especializada. Klinecicz mostra que, ao se fixar a configuração de concentradores no modelo FLOWLOC, a formulação de custo côncava resultante pode ser reduzida a um problema clássico de localização de facilidades e podendo, portanto, ser resolvida pelos métodos disponíveis da área.

Neste capítulo, uma formulação mais justa, usando as idéias iniciais de O’Kelly e Bryan (1998) e o modelo proposto por Hamacher et al. (2000), é apresentada para abordar os efeitos de economia de escala em função do tráfego nas conexões entre concentradores. Devido à

qualidade do limite de programação linear, pacotes comerciais de otimização podem agora ser utilizados para resolver instâncias de tamanho médio em um tempo aceitável. Além disso, mais uma vez, o método de decomposição de Benders (Benders, 1962) é empregado, de forma competitiva e eficiente, na resolução de problemas testes com até 50 pontos de demanda.

Neste capítulo, a seção 4.1 apresenta as definições usadas, o modelo FLOWLOC de O’Kelly e Bryan (1998) e a nova formulação. O algoritmo de decomposição de Benders é formalizado na seção 4.2, sendo os experimentos computacionais realizados e os comentários finais mostrados nas seções 4.3 e 4.4, respectivamente.

## 4.1 Definições e Formulações

Os custos unitários de transporte  $c_{ik}$  e  $c_{mj}$  têm o mesmo significado da seção 2.3, porém os custos  $c_{km}$  são agora representados por uma função côncava linear por partes dependente do fluxo passando pela conexão  $k - m$  ( $k \neq m, k, m \in K$ ) (O’Kelly e Bryan, 1998).

Como observado por Klineciewicz (2002), a função côncava de custo pode ser interpretada como o “envelope inferior” de um conjunto  $L$  de funções lineares, onde cada elemento  $l \in L$  é uma das “partes”. Então, na formulação FLOWLOC de O’Kelly e Bryan (1998), os custos das conexões entre concentradores são dados por funções lineares que minimizam:

$$\varphi_{km}(x) = \sum_{l \in L} (b_{km}^l c_{km} + \alpha_{km}^l c_{km} \sum_{i \in N} \sum_{j \in N} w_{ij} x_{ijkm}) \quad \forall k \neq m, k, m \in K$$

onde  $b_{km}^l c_{km}$  e  $\alpha_{km}^l c_{km}$  são o intercepto e a inclinação da conexão  $k - m$  e segmento linear  $l$ , respectivamente. A figura 4.1 ilustra um exemplo de  $\varphi_{km}$  com três partes lineares.

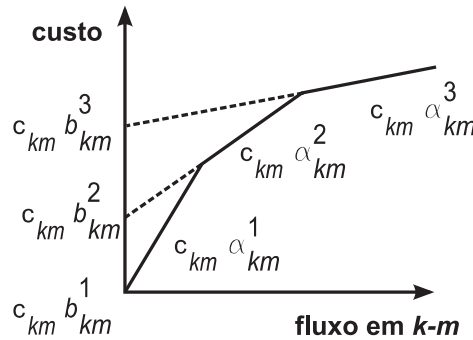


Figura 4.1: Função côncava linear por partes da conexão  $k - m$ . Extraído de Klineciewicz (2002)

O’Kelly e Bryan (1998) definem ainda as variáveis  $z_{km}^l$  que assumem o valor igual a 1, se os parâmetros de custo  $\alpha_{km}^l$  e  $b_{km}^l$  são usados pelo fluxo total da conexão  $k - m$  e segmento  $l$ , representado por  $f_{km}^l \geq 0$ , e 0 caso contrário. No restante do texto, para simplificar a apresentação, considera-se  $l \in L$  nos somatórios das formulações abaixo. Portanto, O’Kelly e Bryan formulam o FLOWLOC como:

$$\begin{aligned} \min \sum_k a_k y_k + \sum_i \sum_j \sum_k \sum_m w_{ij} (c_{ik} + c_{mj}) x_{ijkm} \\ + \sum_l \sum_k \sum_{m \neq k} c_{km} (\alpha_{km}^l f_{km}^l + b_{km}^l z_{km}^l) \end{aligned} \quad (4.1)$$

s. t.:

$$f_{km}^l \leq z_{km}^l \sum_i \sum_j w_{ij} \quad \forall l \in L, k \neq m, k, m \in K \quad (4.2)$$

$$\sum_l f_{km}^l = \sum_i \sum_j w_{ij} x_{ijkm} \quad \forall k \neq m, k, m \in K \quad (4.3)$$

$$x_{kmkm} \geq y_k + y_m - 1 \quad \forall k \neq m, k, m \in K \quad (4.4)$$

$$\sum_l z_{km}^l = x_{kmkm} \quad \forall k \neq m, k, m \in K \quad (4.5)$$

$$\sum_k \sum_m x_{ijkm} = 1 \quad \forall i, j \in N \quad (4.6)$$

$$\sum_m x_{ijkm} \leq y_k \quad \forall i, j \in N, k \in K \quad (4.7)$$

$$\sum_k x_{ijkm} \leq y_m \quad \forall i, j \in N, m \in K \quad (4.8)$$

$$x_{ijkm} \geq 0 \quad \forall i, j \in N, k, m \in K \quad (4.9)$$

$$f_{km}^l \geq 0 \quad \forall l \in L, k \neq m, k, m \in K \quad (4.10)$$

$$z_{km}^l \in \{0, 1\} \quad \forall l \in L, k \neq m, k, m \in K \quad (4.11)$$

$$y_k \in \{0, 1\} \quad \forall k \in K \quad (4.12)$$

A função objetivo (4.1) minimiza o custo total de instalação dos concentradores e de transporte. As restrições (4.2) associam o intercepto  $b_{km}^l$  com a respectiva inclinação  $\alpha_{km}^l$  da conexão  $k - m$  e segmento  $l$ , enquanto que as restrições (4.3) determinam a quantidade de fluxo total passando pela mesma.

As restrições (4.4) e (4.5) forçam toda a demanda  $w_{km}$  a usar a conexão  $k - m$  ao se ter concentradores instalados nos pontos de demanda  $i$  e  $j$ , i. e. quando  $i = k$  e  $j = m$ , e o uso de apenas um dos segmentos da função côncava linear por partes, respectivamente. As restrições (4.6) asseguram que a demanda  $w_{ij}$  seja roteada através de um ou no máximo dois concentradores, enquanto que as restrições (4.7) e (4.8) permitem apenas o roteamento através de concentradores instalados.

Como relatado por Klinecicz (2002), a formulação FLOWLOC (4.1)-(4.12) apresenta *gaps* de integralidade muito grandes, em média 28%. Ao se agregar o fluxo de uma conexão  $k - m$  em uma única variável  $f_{km}^l$ , restrições (4.3), a decomposição natural do problema por origem-destino  $i - j$  é destruída. Como consequência, restrições do tipo “*M grande*”, como (4.2) que acoplam fracamente os conjuntos de variáveis  $f_{km}^l$  e  $z_{km}^l$ , se fazem necessárias, resultando, portanto, em uma relaxação linear de baixa qualidade.

Um *gap* de integralidade mais justo pode ser obtido ao se modelar o problema como uma

formulação de desconto multinível, mantendo-se, assim, a decomposição natural por origem-destino  $i - j$ . Nesse caso, dois novos conjuntos de variáveis,  $\ddot{x}_{ijkm}^l \geq 0$  e  $\bar{x}_{ijk} \geq 0$ , são necessários, representando, respectivamente, as quantidades de tráfego transportadas entre os pontos de demanda  $i$  e  $j$ , roteadas via os concentradores instalados em  $k$  e em  $m$  ( $k \neq m$ ), nesta ordem, utilizando o segmento linear  $l$ ; e via um único concentrador operando em  $k$ . A figura 4.2 ilustra as variáveis  $\ddot{x}_{ijkm}^l$  com três segmentos lineares.

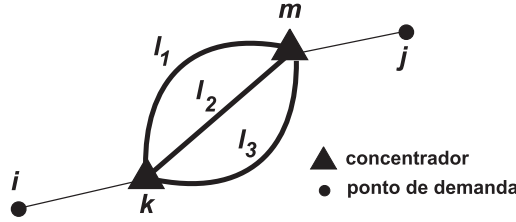


Figura 4.2: Representação das variáveis  $\ddot{x}_{ijkm}^l$ .

Como não há restrições de capacidade nas conexões e nos concentradores instalados,  $\ddot{x}_{ijkm}^l = 1$  ( $\bar{x}_{ijk} = 1$ ) se a rota  $i - k - l - m - j$  ( $i - k - j$ ) é usada para transportar o fluxo  $w_{ij}$ , e  $\ddot{x}_{ijkm}^l = 0$  ( $\bar{x}_{ijk} = 0$ ) caso contrário. O fluxo total em uma conexão  $k - m$  é agora determinado por  $\sum_i \sum_j \sum_l w_{ij} \ddot{x}_{ijkm}^l$  e o custo dado pelo segmento  $l$  que minimiza  $(b_{km}^l + \alpha_{km}^l \sum_i \sum_j w_{ij} \ddot{x}_{ijkm}^l) c_{km}$ . Tem-se então a nova formulação:

$$\begin{aligned} \min \quad & \sum_k a_k y_k + \sum_l \sum_k \sum_{m \neq k} c_{km} b_{km}^l z_{km}^l + \sum_i \sum_j \sum_l \sum_k \sum_{m \neq k} w_{ij} c_{ijkm}^l \ddot{x}_{ijkm}^l \\ & + \sum_i \sum_j \sum_k w_{ij} (c_{ik} + c_{kj}) \bar{x}_{ijk} \end{aligned} \quad (4.13)$$

s. a:

$$\sum_l \sum_{m \neq k} \ddot{x}_{ijkm}^l + \sum_l \sum_{m \neq k} \ddot{x}_{ijmk}^l + \bar{x}_{ijk} \leq y_k \quad \forall i, j \in N, k \in K \quad (4.14)$$

$$\sum_l \sum_k \sum_{m \neq k} \ddot{x}_{ijkm}^l + \sum_k \bar{x}_{ijk} = 1 \quad \forall i, j \in N \quad (4.15)$$

$$\sum_l z_{km}^l \leq y_k \quad \forall k \neq m, k, m \in K \quad (4.16)$$

$$\sum_l z_{km}^l \leq y_m \quad \forall k \neq m, k, m \in K \quad (4.17)$$

$$\ddot{x}_{ijkm}^l \leq z_{km}^l \quad \forall i, j \in N, l \in L, k \neq m, k, m \in K \quad (4.18)$$

$$\ddot{x}_{ijkm}^l \geq z_{km}^l \quad \forall l \in L, k \neq m, k, m \in K \quad (4.19)$$

$$\sum_l z_{km}^l \geq y_k + y_m - 1 \quad \forall k, m \in K \quad (4.20)$$

$$\ddot{x}_{ijkm}^l \geq 0 \quad \forall i, j \in N, l \in L, k \neq m, k, m \in N \quad (4.21)$$

$$\bar{x}_{ijk} \geq 0 \quad \forall i, j \in N, k \in N \quad (4.22)$$

$$z_{km}^l \in \{0, 1\} \quad \forall l \in L, k \neq m, k, m \in K \quad (4.23)$$

$$y_k \in \{0, 1\} \quad \forall k \in K \quad (4.24)$$

onde  $c_{ijkm}^l = (c_{ik} + \alpha_{km}^l c_{km} + c_{mj})$ .

A função objetivo (4.13) minimiza o custo total de instalação de concentradores e de transporte. As restrições (4.14) garantem que o tráfego é roteado apenas por concentradores instalados. As restrições (4.15) asseguram que a demanda  $w_{ij}$  é roteada através de um ou no máximo dois concentradores. As restrições (4.16) e (4.17) permitem a ativação dos parâmetros  $b_{km}^l$  e  $\alpha_{km}^l$  se, e somente se, há concentradores instalados em  $k$  e em  $m$ . O tráfego só usa a rota  $i - k - l - m - j$  se os parâmetros corretos da função côncava linear por partes são utilizados, restrições (4.18). Se a origem e o destino possuem concentradores instalados, i. e.  $i = k$  e  $j = m$ , então as restrições (4.19) forçam a demanda  $w_{km}$  a utilizar a conexão  $k - m$ , sendo equivalentes às restrições (4.4) e (4.5) da formulação FLOWLOC.

**Proposição 4.1** *A formulação (4.13)-(4.24) possui uma relaxação linear mais justa do que a formulação FLOWLOC (4.1)-(4.12).*

Dado uma solução ótima  $\{x^*, f^*, y^*, z^*\}$  da programação linear da formulação FLOWLOC, pode-se observar que as restrições (4.2) permitem fluxos muito grandes em  $f_{km}^{l*}$ , mesmo para pequenos valores fracionários de  $z_{km}^{l*}$ .

Como as restrições (4.18) evitam  $\ddot{x}_{ijkm}^l$  maiores do que  $z_{km}^{l*}$ ,  $\forall l \in L, k \neq m, k, m \in K$ , o valor equivalente a  $f_{km}^{l*}$ , na formulação (4.13)-(4.24), seria então limitado a  $\sum_i \sum_j w_{ij} \ddot{x}_{ijkm}^l$ .

Excluindo-se os casos nos quais apenas um componente de  $z^*$  é estritamente positivo e apenas um concentrador é instalado,  $\sum_i \sum_j w_{ij} \ddot{x}_{ijkm}^l$  é estritamente menor do que  $z_{km}^{l*} \sum_i \sum_j w_{ij}$ , uma vez que apenas os pares  $i - j$  que realmente usam a conexão  $k - m$  e o segmento linear  $l$  são avaliados.

Em outras palavras, os valores  $f_{km}^{l*}$ , computados pela formulação FLOWLOC (4.1)-(4.12), nem sempre são viáveis na formulação (4.13)-(4.24). Entretanto, fazendo-se  $f_{km}^l = \sum_i \sum_j w_{ij} \ddot{x}_{ijkm}^l$ , pode-se sempre encontrar uma solução viável na formulação FLOWLOC. Portanto, a formulação (4.13)-(4.24) possui uma relaxação linear mais justa do que a formulação FLOWLOC (4.1)-(4.12).  $\square$

A formulação (4.13)-(4.24) possui duas propriedades interessantes: a diferença de uma ordem de grandeza entre o número de variáveis contínuas e inteiras, e a sua decomposição natural por par origem-destino  $i - j$ .

Observando a formulação (4.13)-(4.24), têm-se  $q n^4 + (1 - 2q) n^3 + (q - 1) n^2$  variáveis contínuas e  $q n^2 + (1 - q) n$  inteiras, onde  $q = |L|$  e  $n = |N|$ . Mesmo para pequenos valores de

$n$ , tem-se um número considerável de variáveis de fluxo. Em função disso, pacotes comerciais de otimização de programação matemática não conseguem resolver instâncias de grande porte em computadores pessoais.

Além disso, devido a sua decomposição natural por origem-destino  $i - j$ , ao se fixar os valores das variáveis inteiras  $y_k$  e  $z_{km}^l$ , a formulação resultante pode ser reduzida a um problema de transporte e resolvido por meio de algoritmos de caminho mínimo.

Em razão dessas características, abordagens baseadas em decomposição, explorando a estrutura da nova formulação, possuem uma forte motivação. Como o método de decomposição de Benders (Benders, 1962) já foi empregado com sucesso no capítulo 3, é aqui utilizado na resolução do problema de localização de concentradores com a economia de escala dependente do fluxo, assunto da seção 4.2.

Antes de se prosseguir, uma ressalva é feita. Uma das grandes motivações do desenvolvimento de sistemas do tipo eixo-raio é a idéia de rateio (compartilhamento) dos custos de transporte. O uso de políticas que estabelecem conexões antieconômicas para se garantir a interconexão entre todos os concentradores é, em algum sentido, artificial. Em outras palavras, a consideração de se ter todas as conexões entre concentradores ativas, com pelo menos a demanda  $w_{km}$ , quando  $i = k$  e  $j = m$  (restrições (4.4), (4.5) e (4.19)), se opõe ao conceito de se buscar soluções que melhor exploram as economias de escala da rede. Conexões entre concentradores que são muito utilizadas e operam, portanto, com custos unitários marginais  $\alpha_{km}^l c_{km}$  pequenos, tendem a atrair cada vez mais fluxo. Neste sentido, as conexões entre concentradores seriam função natural da estrutura de custos. Portanto, no presente trabalho, as restrições (4.19) são desconsideradas da formulação (4.13)-(4.24) e ao se desenvolver o algoritmo baseado no método de decomposição de Benders (Benders, 1962), seção 4.2.

## 4.2 Algoritmo de Decomposição de Benders

O problema mestre é uma versão relaxada do problema original com o conjunto de variáveis inteiras e suas respectivas restrições a ser formalizado na seção 4.2.2; enquanto que o subproblema é o problema original com os valores das variáveis inteiras temporariamente fixadas, assunto da próxima seção 4.2.1.

### 4.2.1 Subproblema

Desconsiderando as restrições (4.19) da formulação (4.13)-(4.24), o seguinte problema linear primal para cada par origem-destino  $i - j$  ( $i, j \in N$ ) é obtido ao se fixar o vetor de variáveis inteiras,  $y = y^h$  e  $z = z^h$ , em uma dada iteração  $h$ .

$$\min \sum_l \sum_k \sum_{m \neq k} w_{ij} c_{ijk}^l \ddot{x}_{ijk}^l + \sum_k w_{ij} (c_{ik} + c_{kj}) \bar{x}_{ijk} \quad (4.25)$$

s. a:

$$\sum_l \sum_{m \neq k} \ddot{x}_{ijk}^l + \sum_l \sum_{m \neq k} \ddot{x}_{ijm}^l + \bar{x}_{ijk} \leq y_k^h \quad \forall k \in K \quad (4.26)$$

$$\sum_l \sum_k \sum_{m \neq k} \ddot{x}_{ijk}^l + \sum_k \bar{x}_{ijk} = 1 \quad (4.27)$$

$$\ddot{x}_{ijk}^l \leq (z_{km}^l)^h \quad \forall l \in L, k \neq m, k, m \in K \quad (4.28)$$

$$\ddot{x}_{ijk}^l \geq 0 \quad \forall l \in L, k \neq m, k, m \in K \quad (4.29)$$

$$\bar{x}_{ijk} \geq 0 \quad \forall k \in K \quad (4.30)$$

Um limite superior do problema original pode ser obtido somando-se os valores ótimos de (4.25), para todos os pares  $i - j$ , às quantidades  $\sum_k a_k y_k^h$  e  $\sum_l \sum_k \sum_{m \neq k} c_{km} b_{km}^l (z_{km}^l)^h$ . Associando então as variáveis duais  $u_{ijk}$ ,  $\rho_{ij}$  e  $s_{ijk}^l$  às restrições (4.26), (4.27) e (4.28), respectivamente, obtêm-se o seguinte subproblema dual da iteração  $h$  para o par  $i - j$ :

$$\max \rho_{ij} - \sum_k y_k^h u_{ijk} - \sum_l \sum_k \sum_{m \neq k} (z_{km}^l)^h s_{ijk}^l \quad (4.31)$$

sujeito a:

$$\rho_{ij} - u_{ijk} - u_{ijm} - s_{ijk}^l \leq w_{ij} c_{ijk}^l \quad \forall l \in L, k \neq m, k, m \in K \quad (4.32)$$

$$\rho_{ij} - u_{ijk} \leq w_{ij} (c_{ik} + c_{kj}) \quad \forall k \in K \quad (4.33)$$

$$\rho_{ij} \in \mathbb{R} \quad (4.34)$$

$$u_{ijk} \geq 0 \quad \forall k \in K \quad (4.35)$$

$$s_{ijk}^l \geq 0 \quad \forall l \in L, k \neq m, k, m \in K \quad (4.36)$$

Ao invés de se usar os métodos simplex ou de pontos interiores na resolução do subproblema (4.31)-(4.36), pode-se resolvê-lo por inspeção utilizando a teoria da dualidade, a propriedade das folgas complementares e as duas proposições abaixo.

**Proposição 4.2** *O problema primal (4.25)-(4.30) é sempre viável e limitado para quaisquer  $y_k$  e  $z_{km}^l$  viáveis que satisfaçam as restrições (4.14)-(4.24).*

O problema (4.25)-(4.30) é modelado para minimizar os custos de transporte dos pares  $i - j$ ,  $i, j \in N$ . Para qualquer configuração de concentradores instalados, na qual há pelo menos um concentrador operando, e quaisquer parâmetros de custos das conexões  $k - m$ , que satisfaçam às restrições (4.14)-(4.24), pode-se facilmente observar que o modelo (4.25)-(4.30) é sempre viável.

Um algoritmo de caminho mínimo pode ser usado para calcular as melhores rotas para cada par  $i - j$ , fazendo-se a atribuição dos pontos de demanda aos concentradores de forma implícita.

Além disso, qualquer solução do problema primal é limitada, uma vez que os custos de transporte são finitos e não-negativos. Portanto, o problema primal é viável e limitado para quaisquer vetores  $y$  e  $z$  que satisfaçam às restrições (4.14)-(4.24).  $\square$

**Corolário 4.1** *O problema dual (4.31)-(4.36) é viável e limitado, possuindo pelo menos uma solução ótima.*

**Proposição 4.3** *A solução ótima do problema dual (4.31)-(4.36) pode ser encontrada por inspeção.*

Devido à teoria de dualidade e à propriedade das folgas complementares, para quaisquer  $y_k^h$  e  $(z_{km}^l)^h$  viáveis da iteração  $h$  tem-se:

$$\begin{aligned} & \min \left\{ \sum_k w_{ij} \left( (c_{ik} + c_{kj}) \bar{x}_{ijk} + \sum_l \sum_{m \neq k} c_{ijkml}^l \bar{x}_{ijkml}^l \right) \mid (4.26)-(4.30) \text{ é válido} \right\} \\ & = \max \left\{ \rho_{ij} - \sum_k y_k^h u_{ijk} - \sum_l \sum_k \sum_{m \neq k} (z_{km}^l)^h s_{ijkml}^l \mid (4.32)-(4.36) \text{ é válido} \right\} \end{aligned}$$

Para qualquer par  $i - j$ ,  $i, j \in N$ , um problema simples de roteamento deve ser resolvido considerando que os concentradores estão instalados e que os parâmetros de custos estão propriamente estabelecidos. Então, o custo da rota ótima é dado por:

$$w_{ij} \min \left\{ \min_{\substack{l \in L \\ k \neq m, k, m \in K}} \{c_{ijkml}^l \mid y_k^h = y_m^h = (z_{km}^l)^h = 1\}; \min_{k \in K} \{(c_{ik} + c_{kj}) \mid y_k^h = 1\} \right\}$$

Para se zerar o *gap* de dualidade, é possível escrever os valores ótimos de  $\rho_{ij}^h$  da iteração  $h$  como:

$$\rho_{ij}^h = w_{ij} \min \left\{ \min_{\substack{l \in L \\ k \neq m, k, m \in K}} \{c_{ijkml}^l \mid y_k^h = y_m^h = (z_{km}^l)^h = 1\}; \min_{k \in K} \{(c_{ik} + c_{kj}) \mid y_k^h = 1\} \right\}$$

Uma vez calculados os valores ótimos de  $\rho_{ij}^h$ , os valores ótimos de  $u_{ijk}^h$  da iteração  $h$  podem ser obtidos através das restrições (4.33):

$$u_{ijk}^h = \max \left\{ 0, \rho_{ij}^h - w_{ij} (c_{ik} + c_{kj}) \mid y_k^h = 0 \right\}$$



Devido à complementaridade de folga, toda vez que  $y_k^h = 1$ , então tem-se  $u_{ikj}^h = 0$ .

Após os cálculos de  $\rho_{ij}^h$  e  $u_{ijk}^h$ , os valores ótimos de  $(s_{ijk}^l)^h$  da iteração  $h$  podem ser determinados usando-se as restrições (4.32):

$$(s_{ijk}^l)^h = \max \left\{ 0, \rho_{ij}^h - u_{ijk}^h - u_{ijm}^h - w_{ij} c_{ijk}^l \mid (z_{km}^l)^h = 0 \right\}$$

Mais uma vez, devido à complementaridade de folga, toda vez que  $(z_{km}^l)^h = 1$ , então tem-se  $(s_{ijk}^l)^h = 0$ . Deve-se lembrar que para se ter  $(z_{km}^l)^h = 1$  é necessário, mas não suficiente, se ter  $y_k^h = y_m^h = 1$ .

Portanto, a solução ótima do problema dual (4.31)-(4.36) é encontrada por inspeção.  $\square$

O algoritmo de resolução do subproblema é apresentado abaixo:

<b>Algoritmo do Subproblema</b>	
	<b>Para cada par <math>i - j</math></b>
<b>PASSO 1</b>	Calcule $\rho_{ij}^h = w_{ij} \min \left\{ \min_{l \in L, k \neq m, k, m \in K} \{ c_{ijk}^l \mid y_k^h = y_m^h = (z_{km}^l)^h = 1 \}; \min_{k \in K} \{ (c_{ik} + c_{kj}) \mid y_k^h = 1 \} \right\}$ .
<b>PASSO 2</b>	Calcule $u_{ijk}^h = \max \left\{ 0, \rho_{ij}^h - w_{ij} (c_{ik} + c_{kj}) \mid y_k^h = 0 \right\} \quad \forall k \in K$ . Faça $u_{ijk}^h = 0, \quad \forall y_k^h = 1, k \in K$ .
<b>PASSO 3</b>	Calcule
	$(s_{ijk}^l)^h = \max \left\{ 0, \rho_{ij}^h - u_{ijk}^h - u_{ijm}^h - w_{ij} c_{ijk}^l \mid (z_{km}^l)^h = 0 \right\} \quad \forall l \in L, k \neq m, k, m \in K$ .
	Faça $(s_{ijk}^l)^h = 0, \quad \forall (z_{km}^l)^h = 1, k \neq m, k, m \in K$ .
	<b>Fim Para</b>

A partir da função objetivo (4.31) dos subproblemas duais determina-se o corte de Benders:

$$\eta + \sum_i \sum_j \sum_k u_{ijk}^h y_k + \sum_i \sum_j \sum_l \sum_k \sum_{m \neq k} (s_{ijk}^l)^h z_{km}^l \geq \sum_i \sum_j \rho_{ij}^h \quad (4.37)$$

onde  $u_{ijk}^h$ ,  $\rho_{ij}^h$  e  $(s_{ijk}^l)^h$  são os valores ótimos das variáveis duais da iteração  $h$  e  $\eta$  é uma variável contínua representando um subestimador dos custos de transporte. A cada iteração  $h$ , o corte de Benders (4.37) é adicionado ao problema mestre, assunto da próxima seção 4.2.2.

#### 4.2.2 Problema Mestre

O problema mestre é constituído da variável  $\eta$  e dos conjuntos de variáveis inteiras e suas respectivas restrições. O problema mestre é uma relaxação do problema original (4.13)-(4.24) fornecendo portanto um limite inferior. Este limite é melhorado pela adição do corte de Benders (4.37) a cada iteração. Quando o limite inferior converge para o limite superior, a solução ótima é obtida. O problema mestre pode ser assim formulado.

$$\min \eta + \sum_k a_k y_k + \sum_l \sum_k \sum_{m \neq k} c_{km} b_{km}^l z_{km}^l \quad (4.38)$$

s. a:

$$\eta + \sum_i \sum_j \sum_k u_{ijk}^h y_k + \sum_i \sum_j \sum_l \sum_k \sum_{m \neq k} (s_{ikmj}^l)^h z_{km}^l \geq \sum_i \sum_j \rho_{ij}^h \quad h=1, \dots, H \quad (4.39)$$

$$\sum_l z_{km}^l \leq y_k \quad \forall k \neq m, k, m \in K \quad (4.40)$$

$$\sum_l z_{km}^l \leq y_m \quad \forall k \neq m, k, m \in K \quad (4.41)$$

$$\sum_k y_k \geq 1 \quad (4.42)$$

$$\eta \geq 0 \quad (4.43)$$

$$y_k \in \{0, 1\} \quad \forall k \in K \quad (4.44)$$

$$z_{km}^l \in \{0, 1\} \quad \forall l \in L, k \neq m, k, m \in K \quad (4.45)$$

onde  $H$  é o número máximo de iterações de Benders. A restrição (4.42) é adicionada para se evitar soluções inviáveis. Essa restrição assegura a instalação de pelo menos um concentrador, evitando-se assim a geração de raios extremos.

O algoritmo baseado na decomposição Benders é formalmente organizado abaixo, onde  $LS$  é o limite superior,  $LI$  é o limite inferior,  $z_{PM}^*$  e  $z_{SP}^*$  são as soluções ótimas obtidas resolvendo-se o problema mestre e o subproblema correntes, respectivamente:

---

<b>Algoritmo Decomposição de Benders</b>	
<b>PASSO 1</b>	Faça $LS = +\infty$ , $LI = -\infty$ .
<b>PASSO 2</b>	Se $LS = LI$ , <b>PARE</b> . Fim da execução, a solução ótima do problema original (4.13)-(4.24) foi obtida.
<b>PASSO 3</b>	Resolva o problema mestre(4.38)-(4.45), obtendo $z_{PM}^*$ e os valores ótimos das variáveis inteiras $y$ e $z$ .
<b>PASSO 4</b>	Faça $LI = z_{PM}^*$ e atualize os valores de $y$ e $z$ no subproblema dual (4.31)-(4.36).
<b>PASSO 5</b>	Resolva o subproblema dual (4.31)-(4.36), obtendo $z_{SP}^*$ e os valores ótimos das variáveis duais.
<b>PASSO 6</b>	Adicione o novo corte de Benders ao problema mestre (4.38)-(4.45) usando (4.37)
<b>PASSO 7</b>	Se $z_{SP}^* + \sum_k a_k y_k + \sum_l \sum_k \sum_{m \neq k} c_{km} b_{km}^l z_{km}^l < LS$ , faça $LS = z_{SP}^* + \sum_k a_k y_k + \sum_l \sum_k \sum_{m \neq k} c_{km} b_{km}^l z_{km}^l$ . Vá para o <b>PASSO 2</b>

---

Na próxima seção 4.3, os resultados computacionais são apresentados.

### 4.3 Resultados Computacionais

Os resultados computacionais realizados usaram os conjuntos de instâncias padrão CAB e AP, sendo os custos fixos gerados como na seção (3.2). As três funções côncavas lineares por parte utilizadas, propostas por Klineciewicz (2002), possuem quatro inclinações diferentes (veja tabela 4.1), propondo diferentes estratégias de economia de escala: modesta, média e agressiva.

Os 27 testes gerados foram nomeados como CAB $n$ -fq e AP $n$ -fq, indicando a origem dos dados usados, o número de pontos de demanda,  $n$ , e o tipo de função côncava linear por partes utilizada,  $q$  (veja tabela 4.1).

Tabela 4.1: Parâmetros das funções côncavas lineares por partes. (Extraído de Klineciewicz (2002))

Fluxo em $k - m$ (x1,000)	Função		
	1	2	3
	$\alpha_{km}^l$		
$0 \leq f_{km}^1 < 50$	1.0	1.0	0.8
$50 \leq f_{km}^2 < 100$	0.9	0.8	0.6
$100 \leq f_{km}^3 < 200$	0.8	0.6	0.4
$200 \leq f_{km}^4$	0.7	0.4	0.2

Apesar de ter o mesmo número de variáveis inteiras, veja tabela 4.2, a nova formulação (4.13)-(4.24) possui quase quatro vezes o número de variáveis contínuas e, em média, 300 vezes o número de restrições da formulação FLOWLOC de O’Kelly e Bryan (1998). Entretanto a qualidade da relaxação linear é muito superior, como mostrado pela proposição 4.1 e demonstrado pelos testes realizados nas instâncias CAB, tabela 4.3, e ilustrado pela figura 4.3

A figura 4.3 destaca também um comportamento muito interessante. Ao se usar uma estratégia mais agressiva de economia de escala ou ao se aumentar o tamanho das instâncias, a qualidade da relaxação linear da formulação FLOWLOC (O’Kelly e Bryan, 1998) se deteriora, enquanto que a nova formulação se mostra quase indiferente. Isso é uma consequência direta da proposição 4.1.

Todos os testes computacionais foram feitos em um *Pentium* 4 com 2.4 GHz de processador e 1 *Gbyte* de memória principal, executando o sistema operacional Linux, versão Kurumin 5.0. O algoritmo de decomposição de Benders foi implementado em C++ usando a interface do pacote comercial de otimização XPRESS-MP para resolver o problema mestre. Mais uma vez, o limite de dez horas de máquina foi imposto.

A tabela 4.4 mostra a comparação do algoritmo de decomposição de Benders e da formulação (4.13)-(4.24) sendo resolvida de forma monolítica pelo XPRESS. As entradas marcadas com (\*) indicam que as instâncias não foram resolvidas devido a limitação de memória principal do computador.

O impacto da decomposição de Benders para essa classe de problemas é considerável, como visto no capítulo 3. O algoritmo proposto foi capaz de resolver no ótimo instâncias de até 24 milhões de variáveis contínuas e 10 mil variáveis inteiras em um tempo aceitável. Entretanto, o

Tabela 4.2: Tamanho das instâncias

Nome da Instância	FLOWLOC			(4.13)-(4.24)		
	Variáveis		Número de Restrições	Variáveis		Número de Restrições
	Contínuas	Inteiras		Contínuas	Inteiras	
CAB10- <i>fp</i>	9.360	370	930	33.300	370	32.770
CAB15- <i>fp</i>	48.090	855	2.145	179.550	855	177.255
CAB20- <i>fp</i>	153.520	1.540	3.860	585.200	1.540	579.140
CAB25- <i>fp</i>	377.400	2.425	6.075	1.455.000	2.425	1.442.425
AP10- <i>fp</i>	9.360	370	930	33.300	370	32.770
AP20- <i>fp</i>	153.520	1.540	3.860	585.200	1.540	579.140
AP30- <i>fp</i>	786.480	3.510	8.790	3.053.700	3.510	3.031.110
AP40- <i>fp</i>	2.502.240	6.280	15.720	9.796.800	6.280	9.740.680
AP50- <i>fp</i>	6.134.800	9.850	24.650	24.132.500	9.850	24.019.850

método apresentou uma tendência peculiar. Ao se usar estratégias mais agressivas de economia de escala, o desempenho se deteriora bem, porém o número de iterações não se altera muito.

Tabela 4.3: Comparação das relaxações lineares

Instância	Solução Ótima	FLOWLOC			(4.13)-(4.24)		
		Relaxação	Qualidade	Tempo	Relaxação	Qualidade	Tempo
		linear	do Limite	[s]	linear	do Limite	[s]
CAB10-f1	74.110,44	68.605,82	0,926	0,69	73.570,31	0,993	1,48
CAB10-f2	74.110,44	54.300,67	0,733	0,45	73.570,31	0,993	1,66
CAB10-f3	70.937,13	42.019,66	0,592	0,34	70.937,13	1,000	1,97
CAB15-f1	255.161,09	226.180,02	0,886	11,04	255.161,09	1,000	10,05
CAB15-f2	255.161,09	176.049,96	0,690	4,91	255.161,09	1,000	11,39
CAB15-f3	239.482,17	133.692,02	0,558	2,85	239.482,17	1,000	20,16
CAB20-f1	577.791,01	499.067,21	0,864	148,55	576.051,86	0,997	50,24
CAB20-f2	577.791,01	379.892,56	0,657	88,39	576.051,86	0,997	58,69
CAB20-f3	531.125,26	287.741,52	0,542	31,95	531.125,26	1,000	137,85

## 4.4 Comentários Finais

Neste capítulo, uma formulação mais justa foi proposta para o problema de localização de concentradores com a economia de escala dependente do fluxo. A nova formulação claramente apresenta um desempenho muito melhor do que o modelo FLOWLOC de O’Kelly e Bryan (1998).

O uso de pacotes comerciais de otimização são capazes agora de resolver instâncias de médio porte, em um tempo aceitável, usando computadores pessoais. Além disso, o algoritmo de decomposição de Benders desenvolvido se mostrou competitivo e eficiente para abordar o problema, solucionando instâncias com até 50 pontos de demanda. Além disso, como a nova formulação é decomponível por par  $i - j$ , pode-se investigar ainda a utilização de uma versão multicortes (Birge e Louveaux, 1988) como visto na seção 3.1.3.

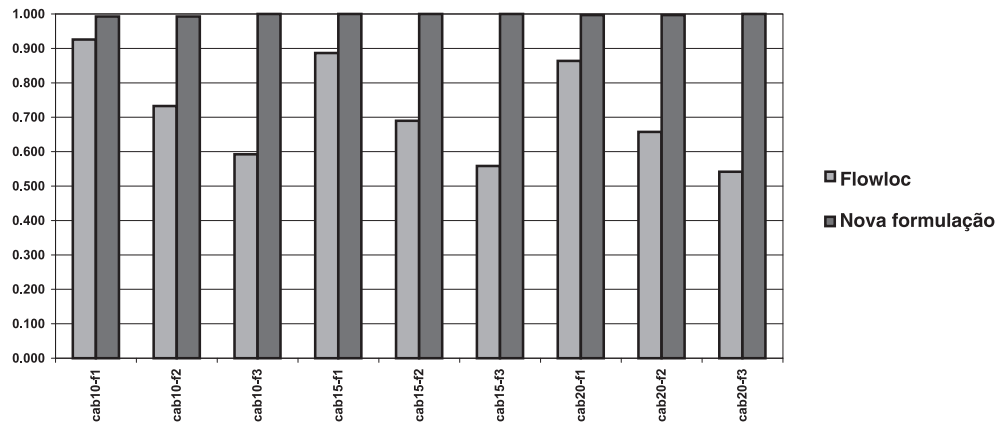


Figura 4.3: Comparação da qualidade das relaxações lineares.

Finalmente, as formulações FLOWLOC e (4.13)-(4.24) são mais próximas da realidade dos projetistas de sistemas eixo-raio, entretanto os efeitos de congestionamento da infraestrutura do sistema não podem ser negligenciados, uma vez que são responsáveis pela degradação da própria economia de escala, sendo, portanto, assunto do próximo capítulo 5.

Tabela 4.4: Resultados computacionais

Instância	Formulação (4.13)-(4.24)	Algoritmo de	
	retirando-se (4.19) e (4.20)	decomposição de Benders	
	XPRESS-MP	Número de	Tempo
	Tempo [s]	Iterações	[s]
CAB10-f1	1,82	17	1,62
CAB10-f2	2,54	17	1,62
CAB10-f3	1,06	40	42,88
CAB15-f1	19,84	24	5,30
CAB15-f2	19,71	24	5,24
CAB15-f3	18,65	9	2,72
CAB20-f1	9.692,31	68	60,14
CAB20-f2	4.788,94	68	58,34
CAB20-f3	10.309,68	18	95,15
CAB25-f1	*	80	257,20
CAB25-f2	*	79	228,97
CAB25-f3	*	21	362,04
AP10-f1	1,68	7	0,13
AP10-f2	1,62	8	0,18
AP10-f3	1,68	8	0,24
AP20-f1	42,79	9	1,59
AP20-f2	44,10	14	2,84
AP20-f3	46,50	13	7,37
AP30-f1	*	60	200,79
AP30-f2	*	83	393,09
AP30-f3	*	64	649,61
AP40-f1	*	102	3.557,46
AP40-f2	*	119	7.257,97
AP40-f3	*	115	17.564,62
AP50-f1	*	57	593,84
AP50-f2	*	113	3.241,73
AP50-f3	*	107	21.668,69

(\*) Memória principal do computador excedida.

## Capítulo 5

# Localização de Concentradores com Congestionamento

*“The pessimist sees difficulty in every opportunity. The optimist sees the opportunity in every difficulty.”*

Winston Churchill

Em sistemas eixo-raio, o congestionamento é comumente modelado de forma implícita, isto é, através de restrições de capacidade de fluxo (Aykin, 1994; Ernst e Krishnamoorthy, 1999; Ebery et al., 2000; Campbell et al., 2002; Marianov e Serra, 2003). Entretanto, apesar de facilitar o desenvolvimento de algoritmos de resolução, essa estratégia de modelagem não capta a natureza explosiva dos efeitos de congestionamento, isto é, quanto maior o volume de tráfego passando por um concentrador, maiores são os custos de operação e/ou as conseqüências, e. g. a degradação do nível de serviço (Hubbard, 1978; Kara e Tansel, 2003; Elhedhli e Hu, 2005).

Uma maneira mais interessante, porém mais complexa de se resolver, é modelar os custos de congestionamento explicitamente na função objetivo, como realizado por Elhedhli e Hu (2005). Usando uma função de uso relativo do concentrador, empregada pelos órgãos gestores do transporte aéreo norte-americano para estimar os custos de atrasos (Gillen e Levinson, 1999), que aumenta os custos segundo uma *lei de potência*, conforme o fluxo passando pelo concentrador é aumentado, Elhedhli e Hu propõem um modelo não-linear para o problema de localização de concentradores com atribuição simples. A formulação é então linearizada e instâncias de até 25 pontos de demanda são resolvidas através de Relaxação Lagrangeana.

Usando a abordagem de Elhedhli e Hu (2005), nesse capítulo, propõe-se uma formulação não-linear para o problema de localização de concentradores com atribuição múltipla na qual os custos de congestionamento estão explicitamente representados na função objetivo. Entretanto, ao invés de se usar Relaxação Lagrangeana, um algoritmo baseado na decomposição de Benders generalizado (Geoffrion, 1972) é desenvolvido para resolver o problema não-linear.

A seção 5.1 apresenta as definições usadas e a formulação proposta. O algoritmo de decomposição de Benders generalizado é formalizado na seção 5.2, sendo os experimentos computacionais realizados e os comentários finais mostrados nas seções 5.3 e 5.4, respectivamente.

## 5.1 Definições e Formulação

O custo de congestionamento em cada concentrador pode ser representado por uma função  $\tau_k(g_k)$  contínua, crescente em  $[0, +\infty)$ , e convexa, onde  $g_k \geq 0$  é o fluxo total passando pelo concentrador instalado em  $k \in K$ .

Essa função de custo representa a natureza explosiva dos efeitos de congestionamento, sendo responsável pela distribuição do fluxo total ao longo da rede de concentradores instalados e, conseqüentemente, pela redução do congestionamento total do sistema, expresso em termos monetários.

A função usada por Elhedhli e Hu (2005), empregada pelos órgãos gestores do transporte aéreo norte-americano para estimar os custos de atrasos (Gillen e Levinson, 1999), é uma lei de potência dada por  $\tau_k = d_k g_k^{b_k}$ , onde  $d_k$  e  $b_k > 1$  são escalares positivos e estão relacionados às capacidades de cada concentrador. A figura 5.1 ilustra essa função para diferentes valores de  $b$ .

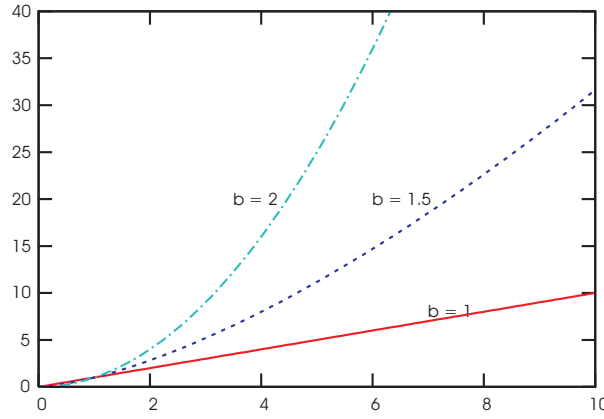


Figura 5.1: Função de congestionamento para diferentes valores de  $b_k$  e  $d_k = 1$

Então, estendendo a formulação de Hamacher et al. (2000), veja seção 2.3.6, e acrescentando  $\tau_k$  à função objetivo, obtém-se o seguinte modelo:

$$\min \sum_k a_k y_k + \sum_k \tau_k(g_k) + \sum_i \sum_j \sum_k \sum_m c_{ijkm} x_{ijkm} \quad (5.1)$$

s. a:

$$\sum_m x_{ijkm} + \sum_{m \neq k} x_{ijmk} \leq w_{ij} y_k \quad \forall i, j \in N, k \in K \quad (5.2)$$

$$\sum_i \sum_j \sum_m x_{ijkm} + \sum_i \sum_j \sum_{m \neq k} x_{ijmk} = g_k \quad \forall k \in K \quad (5.3)$$

$$\sum_k \sum_m x_{ijkm} = w_{ij} \quad \forall i, j \in N \quad (5.4)$$



$$x_{ijkm} \geq 0 \quad \forall i, j \in N, k, m \in K \quad (5.5)$$

$$g_k \geq 0 \quad \forall k \in K \quad (5.6)$$

$$y_k \in \{0, 1\} \quad \forall k \in K \quad (5.7)$$

Neste capítulo, as variáveis  $x_{ijkm}$  não são mais definidas como a fração de fluxo; e sim como a quantidade de fluxo transportada entre os pontos  $i$  e  $j$ , através dos concentradores em  $k$  e  $m$ , nesta ordem. Essa redefinição é feita para facilitar resolução do problema não-linear de atribuição de fluxo gerado ao se fixar as variáveis  $y_k$ , veja seção 5.2.2.1.

A função objetivo (5.1) minimiza os custos de instalação, de congestionamento e de transporte. As restrições (5.2) garantem que as demandas são transportadas através de concentradores instalados. As restrições (5.3) determinam a quantidade de fluxo que passa pelo concentrador em  $k$ . As restrições (5.4) asseguram que toda demanda é roteada através de um ou dois concentradores instalados. A função objetivo (5.1) permite ainda que a carga de fluxo seja balanceada através de toda a infraestrutura de concentradores instalados.

Como a capacidade de cada concentrador é implicitamente considerada na função objetivo, através do parâmetro  $b_k$ , considera-se que a formulação (5.1)-(5.7) sempre possui uma solução viável, desde que exista pelo menos um concentrador instalado. Além disso, para qualquer estrutura de concentradores instalados, isto é, para um vetor  $y$  viável não-nulo, a formulação transforma-se em um problema de transporte não-linear, podendo ser resolvido por métodos de busca de direções viáveis, tais como o método de Frank e Wolfe (1956), o algoritmo de decomposição proximal de Mahey et al. (1998) ou o algoritmo de atribuição de Bar-Gera (2002).

Tais características e as experiências obtidas nos capítulos 3 e 4 motivaram o emprego do método de decomposição de Benders generalizado (Benders, 1962; Geoffrion, 1972) para abordar o problema.

## 5.2 Método de Decomposição de Benders Generalizado

Geoffrion (1972) estende o método de decomposição de Benders (Benders, 1962) para lidar com problemas inteiro mistos não-lineares, formulando o método de decomposição de Benders generalizado.

O método de decomposição de Benders generalizado tem sido usado em diferentes problemas tais como no gerenciamento de recursos hídricos (Cai et al., 2001), no projeto de redes baseadas em modelos não-lineares (Hoang, 2005) e na atribuição de capacidades e fluxo em redes de telecomunicações (Mahey et al., 2001).

Ao ser empregado no problema de localização de concentradores com atribuição múltipla e considerando os custos de congestionamento, o método de decomposição de Benders generalizado divide o problema original em dois outros menores, problema mestre e subproblema, que são resolvidos iterativamente, um após o outro.

Em termos gerais, o problema mestre escolhe uma estrutura de concentradores a ser instalada através das variáveis  $y$ . O subproblema, dado que a estrutura de concentradores tenha

sido fixada pelo problema mestre, resulta em um problema de transporte não-linear. Sua solução dual é então usada na geração de um corte de Benders, que é adicionado ao problema mestre, auxiliando portanto na escolha de uma nova estrutura de concentradores na próxima iteração do método. Enquanto a solução do problema mestre fornece um limite inferior, a solução do subproblema fornece um limite superior. O método converge para a solução ótima, quando o limite inferior se iguala ao limite superior.

Na seções 5.2.1 e 5.2.2, o problema mestre e o subproblema são apresentados, respectivamente, sendo o algoritmo de decomposição de Benders generalizado formalizado na seção 5.2.3.

### 5.2.1 Problema Mestre

A projeção do problema (5.1)-(5.7) no espaço das variáveis estruturais  $y$  resulta no seguinte problema a ser resolvido:

$$\min_{y \in Y} \sum_k a_k y_k + v(y) \quad (5.8)$$

onde  $Y = \{y \mid \text{para } y \text{ fixo, há fluxos viáveis satisfazendo (5.2) - (5.6)}\}$  e  $v(y)$  é dado pela resolução do subproblema abaixo:

$$v(y) = \min_{(x,g) \in G} \sum_k \tau_k(g_k) + \sum_i \sum_j \sum_k \sum_m c_{ijkm} x_{ijkm} \quad (5.9)$$

sujeito a (5.3) para  $y$  fixo e onde  $G = \{(x, g) \mid x \geq 0 \text{ e } g \geq 0 \text{ satisfazem (5.3)-(5.6)}\}$ .

Para qualquer configuração de concentradores, na qual há pelo menos um concentrador instalado, isto é, há pelo menos um  $y_k = 1$ ,  $\forall k \in K$ , há um caminho viável conectando o ponto de origem de demanda  $i$  ao destino  $j$ ; não havendo, portanto, a necessidade de nenhuma restrição adicional de viabilidade para o domínio do problema projetado 5.8.

Além disso, uma vez que o subproblema 5.9 possui uma função objetivo convexa diferenciável e restrições lineares, as condições de Karush-Kuhn-Tucker são necessárias e suficientes para otimalidade, sendo 5.9, conseqüentemente, apto a técnicas de dualização (Geoffrion, 1972; Lasdon, 1970).

Após a associação do vetor de variáveis duais  $u \geq 0$  às restrições de acoplamento (5.2) e como não há *gap* de dualidade, o valor ótimo do subproblema 5.9, para qualquer  $y \in Y$ , pode ser dado por:

$$v(y) = \max_{u \geq 0} \left\{ \min_{(x,g) \in G} \sum_k \tau_k(g_k) + \sum_i \sum_j \sum_k \sum_m c_{ijkm} x_{ijkm} \right. \\ \left. + \sum_i \sum_j \sum_k u_{ijk} \left( \sum_{\substack{m \\ m \neq k}} x_{ijmk} + \sum_m x_{ijkm} - w_{ij} y_k \right) \right\}$$

ou

$$v(y) = \max_{u \geq 0} \left\{ - \sum_i \sum_j \sum_k u_{ijk} w_{ij} y_k + \min_{(x,g) \in G} \left\{ \sum_k \tau_k(g_k) \right. \right. \\ \left. \left. + \sum_i \sum_j \sum_k \left[ (c_{ijkk} + u_{ijk}) x_{ijkk} + \sum_{m \neq k} (c_{ijkm} + u_{ijk} + u_{ijm}) x_{ijkm} \right] \right\} \right\} \quad (5.10)$$

Como o supremo é o menor dos limites superiores, a formulação (5.1)-(5.7) pode ser representada pelo seguinte problema mestre com o auxílio da variável  $\eta \geq 0$ :

$$\min_{y \in Y} \sum_k a_k y_k + \eta \quad (5.11)$$

s. a:

$$\eta \geq - \sum_i \sum_j \sum_k u_{ijk} w_{ij} y_k + \min_{(x,g) \in G} \left\{ \sum_k \tau_k(g_k) + \sum_i \sum_j \sum_k (c_{ijkk} + u_{ijk}) x_{ijkk} \right. \\ \left. + \sum_i \sum_j \sum_k \sum_{m \neq k} (c_{ijkm} + u_{ijk} + u_{ijm}) x_{ijkm} \right\} \quad \forall u \geq 0 \quad (5.12)$$

$$\eta \geq 0 \quad (5.13)$$

O método de decomposição de Benders generalizado resolve o problema (5.11)-(5.13) através de uma estratégia de relaxação. As restrições (5.12) são ignoradas inicialmente, sendo adicionadas iterativamente, uma por vez, quando necessárias, ao problema mestre. As restrições adicionadas são chamadas de cortes de Benders e são formadas a partir da resolução do subproblema (5.9).

Em uma dada iteração  $h$ , o valor ótimo  $v(y^h)$  pode ser obtido resolvendo-se (5.10), para uma dada configuração viável de concentradores instalados  $y^h$ , e recuperando-se o valor ótimo do vetor de variáveis duais  $u^h$ . Então o valor ótimo de  $v(y^h)$  é dado por:

$$\begin{aligned}
v(y^h) = & - \sum_i \sum_j \sum_k u_{ijk}^h w_{ij} y_k^h + \min_{(x,g) \in G} \left\{ \sum_k \tau_k(g_k) \right. \\
& \left. + \sum_i \sum_j \sum_k \left[ (c_{ijkk} + u_{ijk}^h) x_{ijkk} + \sum_{m \neq k} (c_{ijkm} + u_{ijk}^h + u_{ijm}^h) x_{ijkm} \right] \right\} \quad (5.14)
\end{aligned}$$

isolando o mínimo de (5.14), tem-se:

$$\begin{aligned}
v(y^h) + \sum_i \sum_j \sum_k u_{ijk}^h w_{ij} y_k^h = & \min_{(x,g) \in G} \left\{ \sum_k \tau_k(g_k) + \sum_i \sum_j \sum_k (c_{ijkk} + u_{ijk}^h) x_{ijkk} \right. \\
& \left. + \sum_i \sum_j \sum_k \sum_{m \neq k} (c_{ijkm} + u_{ijk}^h + u_{ijm}^h) x_{ijkm} \right\} \quad (5.15)
\end{aligned}$$

Substituindo o mínimo de (5.15) em (5.12), tem-se a seguinte restrição associada a  $u^h$ :

$$\eta \geq - \sum_i \sum_j \sum_k u_{ijk}^h w_{ij} y_k + v(y^h) + \sum_i \sum_j \sum_k u_{ijk}^h w_{ij} y_k^h$$

Reorganizando, obtém-se o seguinte corte de Benders baseado em  $y^h$  e  $u^h$ :

$$\eta + \sum_i \sum_j \sum_k u_{ijk}^h w_{ij} (y_k - y_k^h) \geq v(y^h) \quad (5.16)$$

O problema mestre é agora substituído pelo seguinte problema mestre relaxado:

$$\min \sum_k a_k y_k + \eta \quad (5.17)$$

s. a:

$$\eta + \sum_i \sum_j \sum_k u_{ijk}^h w_{ij} (y_k - y_k^h) \geq v(y^h) \quad \forall h = 1, \dots, H \quad (5.18)$$

$$\sum_k y_k \geq 1 \quad (5.19)$$

$$\eta \geq 0 \quad (5.20)$$

$$y_k \in \{0, 1\} \quad \forall k \in K \quad (5.21)$$

onde  $H$  é o número máximo de iterações de Benders para se obter a solução ótima e a restrição (5.19) força a instalação de pelo menos um concentrador, evitando assim a geração de soluções inviáveis para o problema original.

A cada iteração  $h$ , um novo corte de Benders (5.16) pode ser gerado dado que os valores ótimos de  $v(y^h)$  e do vetor de variáveis duais  $u^h$  tenham sido recuperados a partir da resolução do subproblema (5.9), assunto da seção (5.2.2).

### 5.2.2 Subproblema

Para uma estrutura de concentradores  $y^h$ , fixada pelo problema mestre relaxado em uma iteração  $h$ , o subproblema (5.9) a ser resolvido é:

$$\min \sum_k \tau_k(g_k) + \sum_i \sum_j \sum_k \sum_m c_{ijkm} x_{ijkm} \quad (5.22)$$

s. t.:

$$\sum_m x_{ijkm} + \sum_{m \neq k} x_{ijmk} \leq w_{ij} y_k^h \quad \forall i, j \in N, k \in K \quad (5.23)$$

$$\sum_i \sum_j \sum_m x_{ijkm} + \sum_i \sum_j \sum_{m \neq k} x_{ijmk} = g_k \quad \forall k \in K \quad (5.24)$$

$$\sum_k \sum_m x_{ijkm} = w_{ij} \quad \forall i, j \in N \quad (5.25)$$

$$x_{ijkm} \geq 0 \quad \forall i, j \in N, k, m \in K \quad (5.26)$$

$$g_k \geq 0 \quad \forall k \in K \quad (5.27)$$

Observando (5.22)-(5.27), têm-se apenas dois tipos de vetores de variáveis,  $x$  e  $g$ , e, como a função  $\tau_k(g_k)$  é crescente e convexa, as restrições de acoplamento (5.24) podem ser dualizadas através dos multiplicadores Lagrangeanos  $\beta \geq 0$ . Portanto, pode-se dividir o problema em dois menores: um problema linear de transporte  $d_L(\beta)$ , possuindo apenas as variáveis  $x_{ijkm}$ ; e um problema não-linear de atribuição de fluxo,  $d_{NL}(\beta)$ , tendo somente as variáveis  $g_k$ .

$$d_L(\beta) = \min \sum_i \sum_j \sum_k \sum_{m \neq k} (c_{ijkm} + \beta_k + \beta_m) x_{ijkm} + \sum_i \sum_j \sum_k (c_{ijkk} + \beta_k) x_{ijkk}$$

sujeito às restrições (5.23), (5.25) e (5.26).

$$d_{NL}(\beta) = \min_{g \geq 0} \sum_k (\tau_k(g_k) - \beta_k g_k)$$

Com os valores correspondentes das variáveis duais  $\beta \geq 0$ , o valor ótimo  $v(y^h)$  pode ser calculado como:

$$v(y^h) = \max_{\beta \geq 0} \{d_L(\beta) + d_{NL}(\beta)\}$$

Lembrando que  $d_{NL}(\beta)$  possui uma função objetivo diferenciável e restrições lineares, as condições de Karush-Kuhn-Tucker são necessárias e suficientes para otimalidade. Então, para

todo  $k \in K$  e uma configuração de concentradores  $y^h$  fixadas, a solução ótima  $\beta_k^h$  é aquela que minimiza a parcela correspondente de  $d_{NL}(\beta)$  da função objetivo, ou, para iteração  $h$ :

$$\beta_k^h = \tau'_k(g_k^h) \quad \forall k \in K \quad (5.28)$$

Os valores ótimos de  $g_k^h$  podem ser obtidos através de um algoritmo baseado no método de Frank e Wolfe (1956), a ser descrito na seção 5.2.2.1

### 5.2.2.1 Problema Não-linear de Atribuição de Fluxo

O método proposto por Frank e Wolfe (1956) permite calcular os valores ótimos de  $g_k$ , porém a rede de concentradores fixados pelo problema mestre relaxado deve ser antes reformulada.

Em uma dada iteração  $h$ , a rede composta pelos concentradores instalados e os pontos de demanda pode ser modelada através de um grafo não-orientado  $\bar{G} = (N, E_1^h \cup E_2^h)$  onde  $N$  é conjunto de pontos de demanda, conforme descrito na seção 2.3,  $E_1^h$  é o conjunto de arestas conectando os pontos de demanda aos pontos onde há concentradores instalados na iteração  $h$  e  $E_2^h$  é o conjunto das conexões entre concentradores instalados. Seja ainda também  $K_1^h \subseteq K$  o conjunto de pontos onde há concentradores instalados na iteração  $h$ . A figura 5.2 (a) ilustra um exemplo de um grafo  $\bar{G}$ , no qual as arestas em negrito representam as conexões entre concentradores instalados.

O grafo  $\bar{G}$  pode ser transformado em um grafo direcionado  $D$  dividindo-se cada ponto  $k \in K_1^h$ , onde há um concentrador instalado, em dois pontos  $k'$  e  $k''$  e conectando-os por um arco  $(k', k'')$ . Esses novos arcos criados compõem o conjunto  $A_K^h$ . Além disso, cada aresta  $(i, k) \in E_1^h$  é convertida nos arcos  $(i, k')$  e  $(k'', i)$ , formando o conjunto de arcos  $A_N^h$ , e cada aresta  $(k, m) \in E_2^h$  é substituída pelos arcos  $(k'', m')$  e  $(m'', k')$  constituindo o conjunto  $A_D^h$  de arcos. A figura 5.2 (b) mostra o grafo transformado.

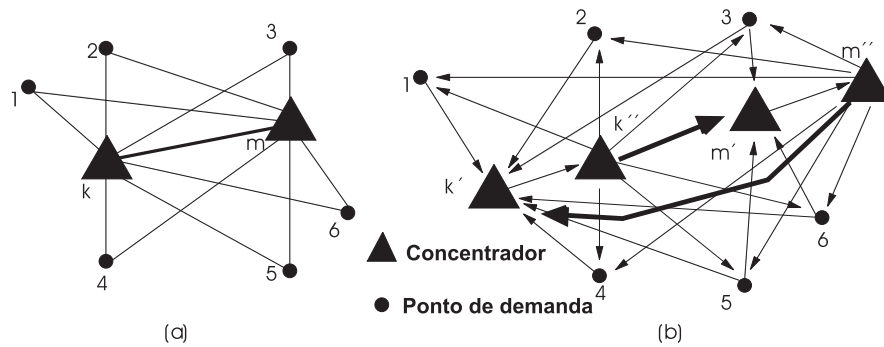


Figura 5.2: Modificação da representação da rede.

Definem-se ainda as variáveis  $f_{uv}^i \geq 0$ , representando a quantidade de fluxo passando pelo arco  $(u, v) \in A$ , onde  $A = A_D^h \cup A_N^h \cup A_K^h$ , na iteração  $i$  do método de Frank e Wolfe. Em termos gerais, após a transformação do grafo, o método consiste em: dado um fluxo viável

$f_{uv}^i$ ,  $(u, v) \in A$ , calcula-se o comprimento  $\ell_{uv}^i$  de cada arco  $(u, v) \in A$  como sendo a derivada parcial  $\ell_{uv}^i = \frac{\partial \varsigma_{uv}(f_{uv}^i)}{\partial f_{uv}^i}$ , onde  $\varsigma_{uv}(f_{uv}^i)$  é a função de custo de transporte dada por:

$$\varsigma_{uv}(f_{uv}^i) = \begin{cases} c_{uv} f_{uv}^i, & \forall (u, v) \in A_N^h \\ \alpha c_{uv} f_{uv}^i, & \forall (u, v) \in A_D^h \\ \tau_{u,v}(f_{uv}^i), & \forall (u, v) \in A_K^h \end{cases}$$

e  $c_{uv}$  é o custo unitário de transporte entre os pontos  $u$  e  $v$ .

Então, para cada par origem-destino  $i - j$ ,  $i, j \in N$ , um problema de caminho mínimo é resolvido usando o grafo  $D$  e os comprimentos dos arcos  $\ell_{uv}^i$ ,  $(u, v) \in A$ , obtendo-se o vetor de fluxo  $\tilde{f}^i$ .

Uma busca de interpolação  $\min_{\lambda \in [0,1]} f^i + \lambda(\tilde{f}^i - f^i)$  é então realizada. Como esse problema possui uma única variável, o valor ótimo  $\lambda^{*i}$  pode ser obtido através de uma busca unidimensional, por exemplo o método da seção áurea (Bazaraa et al., 1992).

Devido à convexidade do conjunto de viabilidade, uma nova solução viável  $f^{i+1}$  pode ser calculada usando  $f^{i+1} = f^i + \lambda^{*i}(\tilde{f}^i - f^i)$ . O procedimento é repetido até que algum critério de parada seja satisfeito, como por exemplo,  $\frac{\|f^{i+1} - f^i\|}{\|f^i\|} < \varepsilon$ , onde  $\varepsilon$  é um valor pequeno, como por exemplo 0,01.

Os valores de  $g^h$  podem ser agora recuperados, uma vez que valores ótimos ou asseguradamente próximos do ótimo  $f^*$  estão disponíveis:

$$g_k^h = \{f_{uv}^* \mid u = k' \text{ e } v = k''\} \quad \forall k \in K_1^h$$

Uma vez calculados, os valores de  $\beta^h$  podem ser então determinados através de (5.28) e utilizados no subproblema  $d_L(\beta)$ , assunto da seção 5.2.2.2.

### 5.2.2.2 Problema Linear de Transporte

Após os cálculos de  $g^h$  e, conseqüentemente, de  $\beta^h$ , pode-se agora resolver o subproblema de transporte  $d_L(\beta)$  e determinar portanto os valores ótimos de  $u^h$  e de  $v(y^h)$ .

Ao invés resolver  $d_L(\beta)$  por meio dos métodos simplex ou de pontos interiores, uma algoritmo especializado, semelhante ao da seção 3.1.5, pode ser desenvolvido baseando-se nas condições de complementaridade de folga e na seguinte proposição:

**Proposição 5.1** *O problema primal  $d_L(\beta)$  é sempre viável e limitado para qualquer  $y_k$  viável fixado que satisfaça as restrições (5.23), (5.25) e (5.26).*

**Prova.** O problema é modelado para achar o menor custo de transporte entre os pares  $i - j$ . Desde que exista pelo menos um concentrador instalado, pode-se facilmente visualizar que  $d_L(\beta)$  é sempre viável para qualquer configuração  $y$  que satisfaça as restrições (5.23), (5.25) e (5.26).

Através de um algoritmo de caminho mínimo, pode-se determinar qual a melhor rota a ser usada através dos concentradores instalados por cada par  $i - j$ . Os pontos de demanda são portanto implicitamente atribuídos aos concentradores instalados.

Uma vez que os custos de transporte são finitos e não negativos, qualquer solução viável para o problema primal deve ser limitado. Portanto o problema primal é viável e limitado para qualquer  $y$  dado que satisfaça as restrições (5.23), (5.25) e (5.26).  $\square$

**Corolário.** O problema dual de  $d_L(\beta)$  é viável e limitado, possuindo, pelo menos, uma solução ótima.

Associando as variáveis duais  $u_{ijk}$  e  $\rho_{ij}$  às restrições (5.23) e (5.25), respectivamente, obtém-se o seguinte problema dual de  $d_L(\beta)$  da iteração  $h$ .

$$\max \sum_i \sum_j w_{ij} \rho_{ij} - \sum_i \sum_j \sum_k w_{ij} y_k^h u_{ijk} \quad (5.29)$$

sujeito a:

$$\rho_{ij} - u_{ijk} - u_{ijm} \leq c_{ijkm} + \beta_k + \beta_m \quad \forall i, j \in N, k \neq m, k, m \in K \quad (5.30)$$

$$\rho_{ij} - u_{ijk} \leq c_{ijkk} + \beta_k \quad \forall i, j \in N, k \in K \quad (5.31)$$

$$u_{ijk} \geq 0 \quad \forall i, j \in N, k \in K \quad (5.32)$$

$$\rho_{ij} \in \mathbb{R} \quad \forall i, j \in N \quad (5.33)$$

**Proposição 5.2** *A solução ótima do problema dual (5.29)-(5.33) pode ser encontrada por inspeção.*

**Prova.** Da teoria de dualidade e da propriedade de folgas complementares, para qualquer  $y_k^h$  viável dado, em uma iteração  $h$ , tem-se:

$$\begin{aligned} & \min \left\{ \sum_i \sum_j \sum_k \sum_{m \neq k} (c_{ijkm} + \beta_k + \beta_m) x_{ijkm}^h \right. \\ & \quad \left. + \sum_i \sum_j \sum_k (c_{ijkk} + \beta_k) x_{ijkk}^h \mid (5.23), (5.25) \text{ e } (5.26) \right\} \\ & = \max \left\{ \sum_i \sum_j w_{ij} \rho_{ij}^h - \sum_i \sum_j \sum_k w_{ij} u_{ijk}^h y_k^h \mid (5.30)-(5.33) \text{ é válido} \right\} \end{aligned}$$

A rota ótima do par  $i - j$  é dada por:

$$\min \left\{ \min_{\substack{k \neq m \\ k, m \in K}} \{c_{ijkm} + \beta_k + \beta_m\}; \min_{k \in K} \{c_{ijkk} + \beta_k\} \mid y_k^h = y_m^h = 1 \right\}$$

Dessa forma, é possível interpretar as variáveis duais  $\rho_{ij}^h$  como a maior diferença de preço entre os pontos  $i$  e  $j$ , considerando os concentradores instalados. Então os valores ótimos



podem ser calculados por:

$$\rho_{ij}^h = \min \left\{ \min_{\substack{k \neq m \\ k, m \in K}} \{c_{ijkm} + \beta_k + \beta_m\}; \min_{k \in K} \{c_{ijkk} + \beta_k\} \mid y_k^h = y_m^h = 1 \right\} \quad (5.34)$$

Essa interpretação resulta em um *gap* de dualidade igual a *zero*. Os cálculos dos valores das variáveis duais  $u_{ijk}^h$  são feitos também através da interpretação econômica, sendo consideradas como taxas adicionais que devem ser pagas para se rotear o fluxo do par  $i - j$  através de um concentrador alternativo ainda não instalado em  $k$ . Para se computar os valores mais justos para essas variáveis duais, lembrando que os valores de  $\rho_{ij}^h$  já foram determinados usando (5.34), tem-se o problema remanescente para cada par  $i - j$ :

$$\min \sum_k w_{ij} y_k^h u_{ijk}^h \quad (5.35)$$

sujeito a:

$$u_{ijk}^h + u_{ijm}^h \geq \rho_{ij}^h - (c_{ijkm} + \beta_k + \beta_m) \quad \forall k \neq m, k, m \in K \quad (5.36)$$

$$u_{ijk}^h \geq \rho_{ij}^h - (c_{ijkk} + \beta_k) \quad \forall k \in K \quad (5.37)$$

$$u_{ijk}^h \geq 0 \quad \forall k \in K \quad (5.38)$$

Uma maneira de resolver esse problema eficientemente é observar as restrições (5.37) e (5.38). Elas permitem o cálculo de valores iniciais para  $u_{ijk}^h$ , ou:

$$u_{ijk}^h = \max\{0, (\rho_{ij}^h - c_{ijkk})\} \quad \forall y_k^h = 0, k \in K$$

Respeitando a complementaridade de folga, têm-se ainda  $u_{ijk}^h = 0$  para todo  $y_k^h = 1$ .

As possíveis combinações de rotas usando dois concentradores distintos devem ser agora consideradas para se calcular os valores ótimos das variáveis  $u_{ijk}$ . Como na seção 3.1.5, isto implica em respeitar as restrições (5.36) e em inspecionar, para cada par  $i - j$ , caminhos usando um, caso (i), ou dois, caso (ii), concentradores alternativos ainda não instalados. Então, usando os valores iniciais de  $u_{ijk}^h$ , têm-se para o caso (i):

$$u_{ijk}^h = \max \left\{ u_{ijk}^h, \max_{\substack{m \in K \\ m \neq k, y_m^h = 1}} \{ (\rho_{ij}^h - c_{ijkm} - \beta_m), (\rho_{ij}^h - c_{ijmk} - \beta_m) \} \right\} \quad \forall y_k^h = 0, k \in K$$

Com os valores calculados para o caso (i), analisa-se o caso (ii):

$$u_{ijk}^h = \max \left\{ u_{ijk}^h, \max_{\substack{m \in K \\ m \neq k, y_m^h = 0}} \{ (\rho_{ij}^h - c_{ijkm} - u_{ijm}^h), (\rho_{ij}^h - c_{ijmk} - u_{ijm}^h) \} \right\} \quad \forall y_k^h = 0, k \in K$$

Dessa forma, os valores ótimos das variáveis duais são calculados.  $\square$

O algoritmo de resolução do subproblema  $d_L(\beta)$  é apresentado abaixo:

---

<b>Algoritmo do Subproblema</b>	
<b>Para cada par <math>i - j</math></b>	
<b>PASSO 1</b>	Calcule $\rho_{ij}^h = \min \left\{ \min_{\substack{k \neq m \\ k, m \in K}} \{c_{ijkm} + \beta_k + \beta_m\}; \min_{k \in K} \{c_{ijkk} + \beta_k\} \mid y_k^h = y_m^h = 1 \right\}$ .
<b>PASSO 2</b>	Calcule $u_{ijk}^h = \max\{0, (\rho_{ij}^h - c_{ijkk})\}$ , $\forall y_k^h = 0, k \in K$ . Faça $u_{ijk}^h = 0$ , $\forall y_k^h = 1, k \in K$ .
<b>PASSO 3</b>	Calcule
	$u_{ijk}^h = \max \left\{ u_{ijk}^h, \max_{\substack{m \in K \\ m \neq k, y_m^h = 1}} \{(\rho_{ij}^h - c_{ijkm} - \beta_m), (\rho_{ij}^h - c_{ijmk} - \beta_m)\} \right\}, \forall y_k^h = 0, k \in K.$
<b>PASSO 4</b>	Calcule
	$u_{ijk}^h = \max \left\{ u_{ijk}^h, \max_{\substack{m \in K \\ m \neq k, y_m^h = 0}} \{(\rho_{ij}^h - c_{ijkm} - u_{ijm}^h), (\rho_{ij}^h - c_{ijmk} - u_{ijm}^h)\} \right\}, \forall y_k^h = 0, k \in K.$
<b>Fim Para</b>	

---

Após os cálculos de  $d_{NL}(\beta)$ ,  $d_L(\beta)$  e das variáveis duais  $u_{ijk}$  pode-se determinar  $v(y^h)$  da iteração  $h$  e, conseqüentemente, construir o corte de Benders (5.16) a ser adicionado ao problema mestre relaxado. O algoritmo de decomposição de Benders generalizado é então formalizado na seção 5.2.3.

### 5.2.3 Algoritmo de Decomposição de Benders Generalizado

O algoritmo de decomposição de Benders generalizado é organizado a seguir, onde  $LS$  é o limite superior,  $LI$  é limite inferior,  $\varepsilon$  é um valor previamente estipulado, por exemplo 0,01,  $z_{RMP}^*$  e  $z_v^*$  são os valores ótimos obtidos resolvendo-se o problema mestre relaxado e o subproblema  $v(y^h)$  correntes, respectivamente:

**Algoritmo de Decomposição de Benders Generalizado**

- 
- PASSO 1** Faça  $LS = +\infty$ ,  $LI = -\infty$ .
- PASSO 2** Se  $\frac{(LS-LI)}{LS} < \varepsilon$ , **PARE**. Fim da execução, uma solução próxima da ótima foi obtida.
- PASSO 3** Resolva o problema mestre relaxado (5.17)-(5.21), obtendo  $z_{RMP}^*$  e os valores ótimos das variáveis inteiras  $y$ .
- PASSO 4** Faça  $LI = z_{RMP}^*$  e atualize os valores de  $y$  nos subproblemas  $d_L(\beta)$  e  $d_{NL}(\beta)$ .
- PASSO 5** Use o algoritmo baseado no método de Frank e Wolfe (1956), descrito na seção 5.2.2.1, e determine os valores ótimos de  $g^h$ .
- PASSO 6** Calcule os valores de  $\beta^h$  usando a equação (5.28) e os valores de  $g^h$ .
- PASSO 7** Resolva o subproblema  $d_L(\beta)$  como descrito na seção 5.2.2.2, obtendo os valores ótimos das variáveis  $u^h$ .
- PASSO 8** Calcule o valor ótimo  $v(y^h)$  e faça:  

$$z_v^* = v(y^h) + \sum_k a_k y_k^h.$$
- PASSO 9** Adicione o novo corte de Benders ao problema mestre relaxado (5.17)-(5.21) usando (5.16).
- PASSO 10** Se  $z_v^* < LS$ , faça  $LS = z_v^*$ . Vá para o **PASSO 2**
- 

Na próxima seção 5.3, os resultados computacionais são apresentados.

### 5.3 Resultados Computacionais

Os resultados computacionais realizados usaram os conjuntos de instâncias padrão CAB, sendo os custos fixos gerados como na seção (3.2). Os 12 testes gerados foram nomeados como CAB $n.\alpha$ , indicando a origem dos dados usados, o número de pontos de demanda,  $n$ , e o desconto usado entre concentradores,  $\alpha$ . O símbolo  $\alpha$  pode assumir valores iguais a 2, 4, 6 e 8 significando descontos de 20%, 40%, 60% e 80%, respectivamente.

Todos os testes computacionais foram feitos em um *Pentium* 4 com 2.4 GHz de processador e 1 *Gbyte* de memória principal, executando o sistema operacional Windows XP SP2. O algoritmo de decomposição de Benders generalizado foi implementado em C++ usando a interface do pacote comercial de otimização CPLEX 9.1 para resolver o problema mestre. Dessa vez, o limite de dez horas de máquina não foi imposto.

Os parâmetros da função  $\tau_k = d_k g_k^{b_k}$  foram estabelecidos em  $d_k = 1$  e  $b_k = \{1, 1; 1, 5; 1, 7; 2, 0\}$  para se medir diferentes níveis de balanceamento de fluxo pela rede de concentradores. O valor de  $\varepsilon$  foi preestabelecido em 0,01.

Os resultados computacionais são resumidos na tabela 5.1. Para cada valor de  $b_k$ , são mostradas três colunas com o número de iterações, o tempo de execução e o *gap* final obtido pelo algoritmo de decomposição de Benders generalizado.

Quanto mais não-linear é a função  $\tau_k$ , isto é, quanto maior for os valores de  $b$ , pior é o comportamento do algoritmo. O mesmo comportamento é observado ao se aumentar o desconto  $\alpha$ . Para as instâncias CAB20. $\alpha$  e  $b = 2, 0$ , o método apresenta um tempo de computação superior a 15 horas. Isso é parcialmente explicado pelo efeito de *zig-zag* do método de Frank e Wolfe (1956) ao se aproximar do ótimo, tornando a resolução dos subproblemas

Tabela 5.1: Resultados computacionais

Instância	$b_k = 1.1$			$b_k = 1.5$			$b_k = 1.7$			$b_k = 2$		
	Nr.	Tempo	<i>gap</i>	Nr.	Tempo	<i>gap</i>	Nr.	Tempo	<i>gap</i>	Nr.	Tempo	<i>gap</i>
	Iter.	[s]	[%]	Iter.	[s]	[%]	Iter.	[s]	[%]	Iter.	[s]	[%]
CAB10.2	12	0,39	0,23	18	4,72	0,00	30	14,48	0,97	50	28,80	0,74
CAB10.4	12	0,38	0,21	25	7,36	0,01	46	28,33	0,52	82	51,14	0,16
CAB10.6	15	0,55	0,67	27	10,77	1,00	43	23,33	0,89	88	58,89	0,87
CAB10.8	17	2,28	0,59	31	18,55	0,86	42	25,00	0,90	65	41,02	0,95
CAB15.2	14	2,86	0,98	31	29,91	0,75	99	118,24	0,98	321	187,78	0,98
CAB15.4	12	1,13	0,67	46	35,62	0,21	89	132,05	0,96	415	324,75	0,97
CAB15.6	8	0,25	0,58	57	42,47	0,18	91	86,39	0,99	236	299,06	0,75
CAB15.8	36	5,61	0,98	263	324,14	0,99	811	3.434,17	1,00	1.891	6.608,73	0,97
CAB20.2	24	3,47	0,94	162	185,03	0,99	844	4.652,81	0,98	2.589	61.701,34	0,97
CAB20.4	20	5,16	0,92	103	135,56	0,89	782	2.564,48	0,97	2.532	58.450,14	0,98
CAB20.6	16	2,41	0,95	86	101,65	0,92	513	867,59	0,99	2.458	54.560,08	0,88
CAB20.8	7	0,38	0,90	51	20,86	0,83	317	367,61	0,98	2.636	62.710,81	0,92

$d_{NL}(\beta)$  proibitivos. Entretanto, para valores de  $b$  menores do que 2,0, o algoritmo se mostra como uma boa ferramenta de projeto e de balanceamento de redes do tipo eixo-raio.

A tabela 5.2 apresenta os fluxos das conexões entre os três concentradores instalados para os quatro valores de  $b_k$  e sem considerar o congestionamento para a instância CAB10.8. As porcentagens dos fluxos são apresentados nas colunas 2 – 4, sendo a razão do maior pelo menor fluxo mostrado na coluna 5. Essa razão permite comparar a distribuição do fluxo pela rede. Uma taxa elevada indica uma má atribuição de tráfego pelo sistema. Apesar de serem instalados o mesmo número de concentradores, a atribuição dos pontos de demandas aos concentradores e, conseqüentemente, a distribuição do tráfego pela rede são diferentes para cada um dos valores de  $b$ .

Ao se aumentar o valor de  $b$ , a razão max/min diminui, sugerindo que ao se considerar os efeitos de congestionamento, sistemas mais equilibrados e balanceados são esperados. Desconsiderando os efeitos de congestionamento, a formulação de (2.35)-(2.39) explora ao máximo a economia de escala, permitindo conexões entre concentradores com pouco fluxo, apenas 6,73 %, e outras sobrecarregadas com 79% do tráfego, para esse exemplo.

Tabela 5.2: Distribuição de fluxo para uma rede com 10 pontos de demanda,  $d = 1$  e  $b = \{1, 1; 1, 5; 1, 7; 2, 0\}$ 

$b$	Fluxo entre concentradores (%)			razão max/min
sem congestionamento	6,73	14,27	79,00	11,74
1,1	9,80	23,80	66,40	6,78
1,5	21,70	24,50	53,80	2,48
1,7	29,21	31,12	39,67	1,36
2	31,40	31,90	36,70	1,17

A variação da distribuição do fluxo de  $b = 1,7$  para  $b = 2$  é muito pequena, menor do que 3% para o exemplo da tabela (5.2). Essa variação é ligeiramente maior para as demais instâncias, em média 6%, desde que o mesmo número de concentradores sejam instalados. Esse comportamento é interessante uma vez que o algoritmo apresenta um desempenho melhor quando  $b < 2$ .

## 5.4 Comentários Finais

Neste capítulo, o problema de localização de concentradores considerando o congestionamento foi abordado. Uma extensão da formulação (2.35)-(2.39) é feita, considerando os custos de congestionamento explicitamente na função objetivo.

Os custos de congestionamento são modelados por meio de uma função convexa (*lei de potência*) que aumenta exponencialmente quanto maior for a quantidade de fluxo passando pelos concentradores instalados.

O algoritmo de decomposição de Benders generalizado proposto se mostra como uma boa ferramenta de projeto de sistemas de eixo-raio, desde que  $b < 2$ , permitindo o desenho de redes balanceadas. Para valores  $b \geq 2$ , outras abordagens (Mahey et al., 1998; Bar-Gera, 2002) de resolução do problema  $d_{NL}(\beta)$ , mais eficientes do que o método de Frank e Wolfe (1956), devem ser melhor investigadas.

Uma abordagem bastante interessante seria reunir as características do capítulo 4 com os efeitos de congestionamento apresentados neste, propondo uma formulação com a economia de escala e o congestionamento representados por uma função côncava e outra convexa, respectivamente. Outra abordagem também interessante seria permitir a possibilidade de expansão de capacidade dos concentradores, como proposto por Luna e Mahey (2000) e usando em problemas de atribuição de fluxo por Ferreira e Luna (2003). Apesar de serem bem realistas, a resolução dessas abordagens podem ser um verdadeiro desafio.

## Capítulo 6

# Conclusão

*“Every day, you may make progress. Every step may be fruitful. Yet, there will stretch out before you an ever-lengthening, ever-ascending, ever-improving path. You know, you will never get to the end of the journey. But this, so far from discouraging, only adds to the joy and glory of the climb.”*

Winston Churchill

O objetivo central deste trabalho foi estudar o problemas de localização de concentradores com atribuição múltipla e duas de suas variantes. O método de decomposição de Benders foi implementado com sucesso para o problema tradicional, capítulo 3. Os resultados computacionais obtidos demonstram que os algoritmos desenvolvidos são competitivos quando comparados a qualidade da solução obtida e o tamanho dos problemas resolvidos.

Infelizmente, a abordagem clássica do problema permite soluções que gratificam baixos volumes de tráfego entre concentradores com grandes descontos. Uma alternativa para se evitar isso é tratar a economia de escala como uma função côncava dependente do fluxo. Uma formulação tratando a economia de escala como uma função côncava linear por partes foi proposta, capítulo 4. A nova formulação não só apresenta uma relaxação linear superior a existente na literatura, como também mantém a decomposição original do problema por par de origem-destino  $i-j$ . Essa propriedade permite o desenvolvimento de um algoritmo bastante eficiente baseado no método de decomposição de Benders. Instâncias, até então consideradas *fora do alcance* de métodos exatos, com até 50 pontos de demanda são resolvidas de forma competitiva em um tempo aceitável.

Uma formulação ainda mais realista, considerando os custos de congestionamento explicitamente na função objetivo, foi também proposta, capítulo 5. O modelo não-linear não só capta os efeitos de congestionamento como também permite balancear o tráfego pelo sistema. O algoritmo implementado, baseado no método de decomposição de Benders Generalizado, é capaz de resolver instâncias de médio porte para valores de  $b < 2$ .

A principal conclusão do trabalho é que a metodologia é eficiente e eficaz na solução dos problemas de localização de concentradores com atribuição múltipla.

Com relação às perspectivas de estudos futuros relacionados com o presente trabalho, pode-se citar:

- Desenvolver um algoritmo híbrido combinando as versões BD1 e BD2 para o problema de localização de concentradores com atribuição múltipla, capítulo 3.
- Investigar o método de decomposição de Benders regularizado (Ruszczynski, 1986) para o problema de localização de concentradores com atribuição múltipla com a economia de escala dependente do fluxo. Semelhante à versão multicortes de Birge e Louveaux (1988), porém o método de decomposição de Benders regularizado permite a remoção dos cortes de Benders que não estão ativos.
- Implementar métodos mais agressivos para resolver o problema não-linear de atribuição de fluxo  $d_{NL}(\beta)$  do capítulo 5. A investigação de outras funções de congestionamento, tais como a proposta por Kleinrock (1993), também é uma alternativa a ser explorada.
- Estudar a localização de concentradores ao longo de multiperíodos. Mais precisamente, o objetivo é desenvolver, ao longo de um horizonte de planejamento de  $T$  períodos, um plano de localização e relocação de concentradores considerando as previsões de demanda dos diversos pares de origem-destino para cada período. A consideração de multiperíodos ao se localizar concentradores tem sua motivação no projeto de redes de sistemas logísticos do tipo *cross-docking*. Nessas redes, em razão da facilidade de se abrir e fechar os centros de consolidação e da volatilidade das demandas ao longo do tempo, é normal, além da própria localização dos centros, a relocação dos mesmos para locais mais favoráveis e até mesmo o fechamento definitivo ao longo do horizonte de planejamento. Uma forma de abordar o projeto de redes logísticas do tipo *cross-docking* é usar o conceito de localização de concentradores com atribuição múltipla usando uma previsão da evolução das demandas no tempo. Uma abordagem semelhante, porém focando o problema de localização de facilidades, foi feita por Wesolowsky e Truscott (1975).
- Formular uma versão estocástica para problema de localização de concentradores com atribuição múltipla. Ainda não há na literatura versões estocásticas do problema.

# Referências Bibliográficas

- Aykin, T. (1994). Lagrangian relaxation based approaches to capacitated hub-and-spoke network design problem. *European Journal of Operational Research*, 79:501–523.
- Aykin, T. (1995). Network policies for hub-and-spoke systems with applications to the air transportation system. *Transportation Science*, 29:201–221.
- Bar-Gera, H. (2002). Origin-based algorithm for the traffic assignment problem. *Transportation Science*, 36(4):398–417.
- Bazaraa, M. S.; Sherali, H. D. e Shetty, C. M. (1992). *NonLinear Programming*. John Wiley & Sons, New York.
- Benders, J. F. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerisch Mathematik*, 4:238–252.
- Birge, J. R. e Louveaux, F. V. (1988). A multicut algorithm for two-stage stochastic linear programs. *European Journal of Operational Research*, 34:384–392.
- Boland, N.; Krishnamoorthy, M.; Ernst, A. e Ebery, J. (2004). Preprocessing and cutting for multiple allocation hub location problem. *European Journal of Operational Research*, 155:638–653.
- Bryan, D. L. e O’Kelly, M. E. (1999). Hub-and-spoke networks in air transportation: an analytical review. *Journal of Regional Science*, 2(30):275–295.
- Cai, X.; McKinney, D. C. e Jr., L. S. L. D. W. W. (2001). Solving large nonconvex water resources management models using generalized Benders decomposition. *Operations Research*, 49(2):235–245.
- Campbell, J. F. (1994a). Integer programming formulations of discrete hub location problems. *European Journal of Operational Research*, 72:387–405.
- Campbell, J. F. (1994b). A survey of network hub location. *Studies in Locational Analysis*, 6:31–49.
- Campbell, J. F. (1996). Hub location and the  $p$ -hub median problem. *Operations Research*, 44(6):923–935.



- Campbell, J. F.; Ernst, A. T. e Krishnamoorthy, M. (2002). Hub location problems. In Drezner, Z. e Hamacher, H. W., editores, *Facility Location*, pp. 373–407. Springer.
- Cordeau, J. F.; Soumis, F. e Desrosiers, J. (2000). A Benders decomposition approach for the locomotive and car assignment problem. *Transportation Science*, 34:133–149.
- Cordeau, J. F.; Soumis, F. e Desrosiers, J. (2001). Simultaneous assignment of locomotives and cars to passenger trains. *Operations Research*, 49(4):531–548.
- Donaldson, H.; Johnson, E. L.; Ratliff, H. D. e Zhang, M. (1999). Schedule driven cross-docking networks. <http://www.isye.gatech.edu/research/files/misc9904.pdf>.
- Drezner, T. e Drezner, Z. (2001). A note on applying the gravity rule to the airline hub problem. *Journal of Regional Science*, 41:447–458.
- Ebery, J.; Krishnamoorthy, M.; Ernst, A. e Boland, N. (2000). The capacitated multiple allocation hub location problema: Formulations and algorithms. *European Journal of Operational Research*, 120:614–631.
- Elhedhli, S. e Hu, F. X. (2005). Hub-and-spoke network design with congestion. *Computers & Operations Research*. To appear.
- Ernst, A. T. e Krishnamoorthy, M. (1996). Efficient algorithms for the uncapacitated single allocation p-hub median problem. *Location Science*, 4:139–154.
- Ernst, A. T. e Krishnamoorthy, M. (1998). Exact and heuristic algorithms for the uncapacitated multiple allocation p-hub median problem. *European Journal of Operational Research*, 104:100–112.
- Ernst, A. T. e Krishnamoorthy, M. (1999). Solution algorithms for the capacitated single allocation hub location problem. *Annals of Operations Research*, 86:141–159.
- Ferreira, R. P. M. e Luna, H. P. L. (2003). Discrete capacity and flow assignment algorithms with performance guarantee. *Computer Communications*, 26(10):1056–1069.
- França, P. e Luna, H. (1982). Solving stochastic transportation-location problem by generalized benders decomposition. *Transportation Science*, 16(2):113–126.
- Frank, M. e Wolfe, P. (1956). An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3:95–110.
- Gavish, B. (1987a). Optimization models for configuring distributed computer systems. *IEEE Transactions on Computers*, C-36(7):773–793.
- Gavish, B. (1987b). Topological design of computer communication networks - the overall design problem. *European Journal of Operational Research*, 58(2):149–172.

- Geoffrion, A. (1972). Generalized Benders decomposition. *Journal of optimization Theory and Applications*, 10(4):237–260.
- Geoffrion, A. M. e Graves, G. W. (1974). Multicommodity distribution system design by Benders decomposition. *Management Science*, 20:822–844.
- Gillen, D. e Levinson, D. (1999). The full cost of air travel in the california corridor. In *Presented in the 78th Annual Meeting of Transportation Research Board*, Washington, DC.
- Grove, G. P. e O’Kelly, M. E. (1986). Hub networks and simulated schedule delay. *Papers of the Regional Science Association*, 59:103–119.
- Hamacher, H. W.; Labbé, M.; Nickel, S. e Sonneborn, T. (2000). Polyhedral properties of the uncapacitated multiple allocation hub location problem. Technical Report 20, Institut für Techno und Wirtschaftsmathematik (ITWM), <http://www.itwm.fhg.de/zentral/berichte/bericht20.pdf>.
- Hamacher, H. W. e Nickel, S. (1998). Classification of location models. *Location Science*, 6:229–242.
- Heragu, S. S. e Chen, J.-S. (1998). Optimal solution of cellular manufacturing system design: Benders decomposition approach. *European Journal of Operational Research*, 107:175–192.
- Hoang, H. H. (2005). Topological optimization of networks: A nonlinear mixed integer model employing generalized benders decomposition. *IEEE Transactions on Automatic Control*, AC-27:164–169.
- Hubbard, H. B. (1978). Terminal airspace/airport congestion delays. *Interfaces*, 8(2):1–13.
- Jaillet, P.; Song, G. e Yu, G. (1996). Airline network design and hub location problems. *Location Science*, 4:195–211.
- Kara, B. Y. e Tansel, B. C. (2003). The latest arrival hub location problem. *Management Science*, 47:1408–1420.
- Kimms, A. (2006). Economies of scale in hub and spoke network design models: We have it all wrong. In Morlock, M.; Schwindt, C.; Trautmann, N. e Zimmermann, J., editores, *Perspectives on Operations Research*, pp. 293–317. Wiesbaden.
- Kleinrock, L. (1993). On the modeling and analysis of computer networks. *Proceedings of the IEEE*, 81(8):1179–1191.
- Klincewicz, J. G. (1991). Heuristics for the  $p$ -hub location problem. *European Journal of Operational Research*, 53:25–37.
- Klincewicz, J. G. (1996). A dual algorithm for the uncapacitated hub location problem. *Location Science*, 4(3):173–184.

- Klincewicz, J. G. (1998). Hub location in backbone/tributary network design: a review. *Location Science*, 6:307–335.
- Klincewicz, J. G. (2002). Enumeration and search procedures for a hub location problem with economies of scale. *Annals of Operations Research*, 110:107–122.
- Lasdon, L. (1970). *Optimization Theory for Large Systems*. Macmillan, London.
- Luna, H. P. L. e Mahey, P. (2000). Bounds for global optimization of capacity expansion and flow assignment problems. *Operations Research Letters*, 26:211–216.
- Magnanti, T. L.; Mirchandani, P. e Wong, R. T. (1986). Tailoring Benders decomposition for uncapacitated network design. *Mathematical Programming Study*, 26:112–154.
- Magnanti, T. L. e Wong, R. T. (1981). Accelerating benders decomposition: Algorithmic enhancement and model selection criteria. *Operations Research*, 29(3):464–483.
- Mahey, P.; Benchakroun, A. e Boyer, F. (2001). Capacity and flow assignment of data networks by generalized benders decomposition. *Journal of Global Optimization*, 20(2):169–189.
- Mahey, P.; Ouorou, A.; Leblanc, L. e Chifflet, J. (1998). A new proximal decomposition algorithm for routing in telecommunication networks. *Networks*, 31:227–238.
- Marianov, V.; d. Serra e Revelle, C. (1999). Location of hubs in a competitive environment. *European Journal of Operations Research*, 114:363–371.
- Marianov, V. e Serra, D. (2003). Location models for airline hubs behaving as m/d/c queues. *Computer and Operations Research*, 30(7):983–1003.
- Marín, A.; Cánovas, L. e Landete, M. (2006). New formulations for the uncapacitated multiple allocation hub location problem. *European Journal of Operations Research*, 172:274–292.
- Mateus, G. R. e Thizy, J. M. (1999). Exact sequential choice of locations in a network. *Annals of Operations Research*, 86:199–219.
- Mayer, G. e Wagner, B. (2002). Hublocator: an exact solution method for the multiple allocation hub location problem. *Computer & Operations Research*, 29:715–739.
- Miranda, G. J. (2004). *Facility Location and Network Design with Congestion Costs and Interdependency*. PhD thesis, Universidade Federal de Minas Gerais, Departamento de Ciência da Computação.
- Nagy, G. e Salhi, S. (2005). Heuristic algorithms for single and multiple depot vehicle routing problems with pickups and deliveries. *European Journal of Operations Research*, 162:126–145.
- Nemhauser, G. L. e Wolsey, L. A. (1988). *Integer and Combinatorial Optimization*. Wiley, New York.

- Nickel, S.; Schobel, A. e Sonneborn, T. (2000). Hub location problems in urban traffic networks. In Niittymähi e Pursula, editores, *Mathematical Methods and Optimisation in Transportation Systems*, pp. 1–12. Kluwer Academic Publisher.
- O’Kelly, M. E. (1986). The location of interacting hub facilities. *Transportation Science*, 20:92–106.
- O’Kelly, M. E. (1987). A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research*, 32:393–404.
- O’Kelly, M. E. (1998). A geographer’s analysis of hub-and-spoke networks. *Journal of Transport Geography*, 3(6):171–186.
- O’Kelly, M. E. e Bryan, D. L. (1998). Hub location with flow economies of scale. *Transportation Research Part B*, 32(8):605–616.
- O’Kelly, M. E. e Miller, H. J. (1994). The hub network design problem. *Journal of Transport Geography*, 2:31–40.
- Pirkul, H. e Schilling, D. A. (1998). An efficient procedure for designing single allocation hub and spoke systems. *Management Science*, 44(12):235–242.
- Racunica, I. e Wynter, L. (2005). Optimal location of intermodal freight hubs. *Transportation Research B*, 39:453–477.
- Randazzo, C. e Luna, H. (2001). A comparison of optimal methods for local access uncapacitated network design. *The Annals of Operations Research*, 106:263–286.
- Ruszczynski, A. (1986). A regularized decomposition method for minimizing a sum of polyhedral functions. *Mathematical Programming*, 35:309–333.
- Sasaki, M.; Suzuki, A. e Drezner, Z. (1999). On the selection of hub airports for an airline hub-and-spoke. *Computer & Operations Research*, 26:1411–1422.
- Skorin-Kapov, D. (1998). Hub network games. *Networks*, 4(31):293–302.
- Skorin-Kapov, D. e Skorin-Kapov, J. (1994). On tabu search for the location of interacting hub facilities. *European Journal of Operational Research*, 73:501–508.
- Skorin-Kapov, D.; Skorin-Kapov, J. e O’Kelly, M. (1996). Tight linear programming relaxations of uncapacitated  $p$ -hub median problems. *European Journal of Operational Research*, 94:582–593.
- Taylor, G. D.; Harit, S.; English, J. R. e Whicker, G. (1995). Hub and spoke networks in truck-load trucking: Configuration, testing and operational concerns. *Logistics & Transportation Review*, 31(3):209–238.
- Wesolowsky, G. O. e Truscott, W. G. (1975). The multiperiod location-allocation problem with relocation of facilities. *Management Science*, 22:57–65.