
PROJETO DE REDE COM CUSTOS CONVEXOS E BALANCEAMENTO DE FLUXOS

Pollyanna G. Faria Dias*
pollyanna@cpdee.ufmg.br

Rodney Rezende Saldanha*
rodney@cpdee.ufmg.br

Gilberto de Miranda Jr.†
miranda@dep.ufmg.br

Ricardo Saraiva de Camargo†
rcamargo@dep.ufmg.br

*Departamento de Engenharia Elétrica
UFMG - Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627, Pampulha, CEP 31.270-901 - Belo Horizonte, MG, Brasil

†Departamento de Engenharia de Produção
UFMG - Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627, Pampulha, CEP 31.270-901 - Belo Horizonte, MG, Brasil

ABSTRACT

Convex Flow Balancing Network Design

In this work, the single source tree network design problem under convex costs is addressed. This is a referential problem when designing networks for materials, energy or data transportation. The modeling effort yields a large scale mixed-integer nonlinear program which is very hard to solve. In order to overcome the solution difficulties, two distinct solution methods are deployed: The first of them is the Generalized Benders Decomposition method; the second technique combines the Outer Approximation method with the ideas of projection in Benders Decomposition in order to imply the technique *Hybrid-OA*. The Hybrid-OA method is very effective on solving instances up to 702 edges, under reasonable computational costs, and makes the further application of the proposed technique to even more sophisticated models promising.

KEYWORDS: Network Design Problems, Benders Decomposition, Outer Approximation

RESUMO

O problema de projeto de redes arborescentes de fonte única com custos convexos é abordado neste trabalho. Trata-se de problema referencial no projeto de redes de transporte de matéria, energia ou informação. Resulta do esforço de modelagem um programa não-linear inteiro misto de grande escala e de difícil resolução. Para superar tais dificuldades, dois métodos distintos são aplicados ao problema: O primeiro deles é a técnica de Decomposição de Benders Generalizada; o segundo método combina a técnica de Aproximação Externa com as idéias de projeção subjacentes à Decomposição de Benders para derivar a técnica *OA-Híbrido*. O método OA-Híbrido se mostra muito eficiente na solução de instâncias com até 702 arcos, obtendo custos de computação bastante razoáveis, tornando promissora a sua aplicação a problemas ainda mais sofisticados.

PALAVRAS-CHAVE: Problemas de Projeto de Rede, Decomposição de Benders, Aproximação Externa.

1 INTRODUÇÃO

O problema de projeto de rede faz parte das atividades de planejamento estratégico de empresas que operam sistemas

Artigo submetido em 03/02/2011 (Id.: 01261)

Revisado em 26/04/2011, 20/06/2011

Aceito sob recomendação do Editor Associado Prof. Luís Fernando C. Alberto

de transporte de passageiros, carga, energia ou informação. O objetivo é atender às demandas dos clientes, respeitando os níveis de serviço pré-estabelecidos e assegurando o lucro e a eficiência da operação.

Portanto devem ser considerados tanto o aspecto da adequação da infra-estrutura para manter o nível de serviço necessário quanto a necessidade de operação otimizada que racionalize o uso da capacidade instalada. Encontrar o melhor equilíbrio entre estas duas visões aparentemente conflitantes – investir mais em infra-estrutura e aumentar a capacidade instalada ou priorizar a otimização das práticas operacionais – é a raiz de uma estratégia de negócio bem sucedida.

Uma forma de buscar tal equilíbrio é estudar o problema de projeto de rede sob a perspectiva da *Programação Matemática*. Assim, o problema consiste basicamente em ligar um servidor aos seus diversos clientes dispersos geograficamente, satisfazendo a demanda a um custo total mínimo. Os custos envolvidos são os de transporte/transmissão, de instalação de infra-estrutura e de qualidade de serviço, ou congestionamento.

Do ponto de vista matemático, o problema pode ser visto como uma generalização do *Problema da Árvore de Steiner* em um grafo direcionado (Maculan, 1987). De fato, ao se desconsiderar custos de transmissão e os de serviço nas conexões, tem-se basicamente o problema de Steiner (Maculan, 1987; Hwang and Richards, 1992). Desprezando-se os custos fixos, tem-se um problema de transbordo de fonte única (Dantzig, 1962).

Problemas de projeto de rede constituem uma classe importante de problemas de otimização combinatória, com aplicações diretas nas áreas de distribuição de energia elétrica como em (Benchakroun et al., 1992) e (Ramirez-Rosado and Dominguez-Navarro, 2006), de transporte de cargas fracionadas (Cordeau et al., 2006), de transporte público (Contreras et al., 2009; Contreras et al., 2010), de redes de acesso local de telefonia (Randazzo and Luna, 2001), e de redes de computadores (Klincewicz, 1998; Altiparmak et al., 2003).

No contexto brasileiro, um grande exemplo de problema de projeto de rede é o programa *Luz para Todos*. O governo brasileiro lançou em 2003 o desafio de acabar com a exclusão elétrica no país. O programa é coordenado pelo Ministério de Minas e Energia, operacionalizado pela Eletrobrás e executado pelas concessionárias de energia elétrica e cooperativas de eletrificação rural. O objetivo do programa é garantir o acesso ao serviço público de energia elétrica a todos os domicílios e estabelecimentos do meio rural. É considerado o programa de inclusão elétrica mais ambicioso implementado no mundo.

Outro exemplo é o *Projeto Banda Larga nas Escolas Públicas Urbanas* que tem por objetivo levar a Internet, em banda larga, a mais de 64 mil escolas públicas urbanas do país. Todas as escolas públicas urbanas deverão ser conectadas à Internet e o serviço mantido, de forma gratuita, até 2025. De acordo com estimativas do Ministério da Educação, mais de 50 milhões de alunos - 86% dos estudantes brasileiros - serão beneficiados pelo projeto.

Apesar de se acreditar que as técnicas aqui propostas são de escopo geral e que podem ser aplicadas a problemas reais tanto de projeto de redes de distribuição de energia quanto de projeto de redes de dados, é fácil ver que o modelo corrente descreve melhor as redes de transporte de informação, estando distante dos modelos mais modernos utilizados no projeto de redes de distribuição de energia. Entretanto, com as devidas adequações, na linha proposta por (Benchakroun et al., 1992), é possível abordar problemas de projeto de redes de transmissão de energia que envolvam funções sofisticadas, desde que o esforço de convexificação garanta boa precisão na descrição dos efeitos não-lineares.

Neste trabalho, redes arborescentes de fonte única são tratadas considerando-se custos de instalação de infra-estrutura, de transporte e de congestionamento, sendo estes últimos responsáveis pela degradação dos níveis de serviço e das economias de escala em cada arco da rede. Resulta desse esforço um programa matemático inteiro misto não-linear de difícil resolução, uma vez que não-linearidades aumentam os *gaps* de integralidade em qualquer formulação. Uma abordagem capaz de resolver eficientemente a formulação é desenvolvida através de técnicas de decomposição, mais especificamente, usando o método de decomposição de Benders (Benders, 1962).

Este método se mostra muito eficiente em outros tipos de problemas de projeto de rede (Costa, 2005). Entretanto, em função da complexidade do problema aqui abordado e da estrutura especial da formulação, o método pode ser combinado com outra técnica clássica de decomposição: Aproximação Externa (Fletcher and Leyfer, 1994). Esta combinação inovadora permitiu reduzir o tempo computacional de resolução e aumentar o tamanho das redes usadas como teste.

O artigo está assim organizado: na Seção 2, as definições usadas e a formulação proposta são apresentadas. A decomposição de Benders é formalizada na Seção 3, enquanto a Aproximação Externa e a combinação com o primeiro método são mostradas na Seção 4. Finalmente, os experimentos computacionais e os comentários finais são feitos nas Seções 5 e 6, respectivamente.

2 DEFINIÇÕES E FORMULAÇÃO

Uma rede genérica é representada por um grafo direcionado $\Gamma(V, E)$, onde V e E são os conjuntos de nodos e de arcos que representam as conexões possíveis entre os nodos, respectivamente. Um arco ou ligação entre os nodos i e j é representado por $(i, j) \in E$, onde $i, j \in V$. Seja ainda s o nodo fonte, que deve enviar matéria energia ou informação a um conjunto K de nodos de demanda. Cada um deles possuindo uma demanda $d_k \geq 0$ distinta, onde $k \in K$ e $K \subseteq V$.

O custo unitário de transporte/transmissão da demanda do nó $k \in K$ no arco $(i, j) \in E$ é representado por c_{ijk} . Assume-se que $c_{ijk} = \theta^k \delta_{ij}$, $\forall k \in K$, onde θ^k é o custo específico de transporte da demanda do nó k e δ_{ij} é a distância entre os nodos i e j . Desta forma, o custo unitário de transporte fica dependente tanto do nó da destino e quanto do comprimento do arco. Caso isto não seja necessário, faz-se $\theta^k = \theta$, $\forall k \in K$. O custo de ativação do arco é dado por b_{ij} , onde $b_{ij} = \zeta_{ij} \delta_{ij}$, $\forall (i, j) \in E$ e ζ_{ij} é o custo da instalação da infraestrutura por unidade de distância.

A formulação proposta utiliza ainda as seguintes variáveis de decisão: $x_{ij} \in \{0, 1\}$, $\forall (i, j) \in E$, que indica se o arco (i, j) é ativado ($x_{ij} = 1$) ou não ($x_{ij} = 0$); $f_{ijk} \geq 0$, para representar o fluxo com destino a $k \in K$ passando pelo arco $(i, j) \in E$; e $g_{ij} \geq 0$, representando o fluxo global do arco $(i, j) \in E$.

A função convexa não-linear de degradação do nível de serviço é representada pelo custo de congestionamento no arcos, sendo separável por arco e representada pela seguinte lei de potência

$$\tau_{ij}(g_{ij}) = e g_{ij}^p, \forall (i, j) \in E \quad (1)$$

Em que $e > 0$ e $p \geq 1$ são escalares representando a participação dos custos de congestionamento na composição do custo total. Valores elevados para e e p implicam em um peso maior dos efeitos de congestionamento no projeto de rede.

O modelo proposto resulta em:

$$\min \sum_{(i,j) \in E} [b_{ij}x_{ij} + \tau_{ij}(g_{ij})] + \sum_{k \in K} c_{ijk}f_{ijk} \quad (2)$$

sujeito a:

$$\sum_i x_{ij} \leq 1 \quad \forall j \in V \quad (3)$$

$$\sum_{k \in K} f_{ijk} - g_{ij} = 0 \quad \forall (i, j) \in E \quad (4)$$

$$\sum_{(s,j) \in E} f_{sjk} = d_k \quad \forall k \in K \quad (5)$$

$$\sum_{(i,k) \in E} f_{ikk} = d_k \quad \forall k \in K \quad (6)$$

$$\sum_{(i,j) \in E} f_{ijk} = \sum_{(j,i) \in E} f_{jik} \quad \forall j \in V \setminus \{s\},$$

$$k \in K : k \neq j \quad (7)$$

$$f_{ijk} \leq d_k x_{ij} \quad \forall (i, j) \in E, k \in K \quad (8)$$

$$f_{ijk} \geq 0 \quad \forall (i, j) \in E, k \in K \quad (9)$$

$$g_{ij} \geq 0 \quad \forall (i, j) \in E \quad (10)$$

$$x_{ij} \in \{0, 1\} \quad \forall (i, j) \in E \quad (11)$$

A função objetivo (2) possui 3 termos: O primeiro termo contabiliza o custo total da instalação dos arcos; o segundo calcula os custos convexos de congestionamento; enquanto o terceiro totaliza o custo de transporte das diversas demandas. As restrições de configuração (3) obrigam a geração de arborescências ao impedir a incidência de mais de um arco em cada nodo. As restrições (4) calculam o fluxo global que atravessa o arco (i, j) . As restrições (5)–(7) garantem o balanço de fluxo para cada produto em cada nodo desde a origem s até cada nodo de destino k . As restrições de acoplamento (8) permitem que o fluxo passe através do arco (i, j) somente se o mesmo estiver instalado. Finalmente, (9) e (10) são as restrições de não negatividade para os fluxos f_{ijk} e g_{ij} respectivamente, enquanto as restrições (11) obrigam a integralidade das variáveis x_{ij} .

3 DECOMPOSIÇÃO DE BENDERS

A técnica de Decomposição de Benders é uma abordagem clássica de solução de problemas de otimização combinatória, baseada nas idéias de particionamento e geração de restrições. (Benders, 1962) propôs um método de particionamento para resolver problemas de programação inteira mista e não-linear. O método particiona o problema em dois níveis: no nível superior resolve-se um problema mestre inteiro e no nível inferior resolve-se um subproblema contínuo. O primeiro consiste em uma versão relaxada do problema original ao qual se adicionam *em vóo* novas restrições, enquanto o segundo consiste no problema original resolvido para um dado valor fixado das variáveis inteiras.

3.1 Problema Mestre

O método de Decomposição de Benders trata problemas essencialmente via projeção, seguido por linearização externa, dualização e relaxação. Começa-se por uma projeção do problema (2) - (11) no espaço das variáveis topológicas x , resultando em

$$\min_{x \in X} \sum_{(i,j) \in E} b_{ij}(x_{ij}) + v(x) \quad (12)$$

onde $X = \{x \mid \text{Existem fluxos viáveis que satisfazem (3) - (10)}\}$ e onde $v(x)$ é calculado resolvendo o seguinte problema

$$v(x) = \min_{(f,g) \in G} \sum_{(i,j) \in E} [\tau_{ij}(g_{ij}) + \sum_{k \in K} c_{ijk} f_{ijk}] \quad (13)$$

sujeito a condição (8) para x fixado onde $G = \{(f, g) \mid f \geq 0 \text{ e } g \geq 0\}$ satisfazendo (3) - (7).

Associando um vetor de variáveis duais $\alpha \geq 0$ às restrições de acoplamento (8), lembrando que a hipótese de convexidade para $\tau_{ij}(g_{ij})$ garante *gap* de otimalidade nulo para qualquer $x \in X$, o valor ótimo do subproblema (13) é dado por

$$v(x) = \max_{\alpha \geq 0} [\min_{(f,g) \in G} \sum_{(i,j) \in E} [\tau_{ij}(g_{ij}) + \sum_{k \in K} c_{ijk} f_{ijk} + \sum_{k \in K} \alpha_{ijk} (f_{ijk} - d_k x_{ij})]] \quad (14)$$

ou

$$v(x) = \max_{\alpha \geq 0} [\sum_{(i,j) \in E} \sum_{k \in K} -\alpha_{ijk} d_k x_{ij} + \min_{(f,g) \in G} \sum_{(i,j) \in E} [\tau_{ij} g(ij) + \sum_{k \in K} (c_{ijk} + \alpha_{ijk}) f_{ijk}]] \quad (15)$$

O problema completo (12) fica então equivalente a

$$\min_{x \in X} \left\{ \sum_{(i,j) \in E} b_{ij} x_{ij} + \max_{\alpha \geq 0} \left[\sum_{(i,j) \in E} \sum_{k \in K} -\alpha_{ijk} d_k x_{ij} + \min_{(f,g) \in G} \sum_{(i,j) \in E} [\tau_{ij} g(ij) + \sum_{k \in K} (c_{ijk} + \alpha_{ijk}) f_{ijk}] \right] \right\} \quad (16)$$

Utilizando agora uma variável auxiliar t , o problema (2)-(11) é equivalente a

$$\min_{t, x \in X} \sum_{(i,j) \in E} b_{ij} x_{ij} + t \quad (17)$$

s.a:

$$t \geq \sum_{(i,j) \in E} \sum_{k \in K} -\alpha_{ijk} d_k x_{ij} + \min_{(f,g) \in G} \sum_{(i,j) \in E} [\tau_{ij} g(ij) + \sum_{k \in K} (c_{ijk} + \alpha_{ijk}) f_{ijk}], \forall \alpha \geq 0 \quad (18)$$

O problema reformulado (17) - (18) é de resolução extremamente difícil, uma vez que se trata de problema de otimização de *dimensão infinita*. Entretanto é possível superar tal dificuldade através de uma estratégia de relaxação, garantindo que o procedimento gere somente um número de restrições exponencial com o número de variáveis inteiras (da ordem de 2^n em programação 0 - 1, uma restrição por solução inteira viável) no pior caso, possuindo portanto número finito de passos, uma vez estabelecidas algumas propriedades de (2) - (11). Segundo (Geoffrion, 1972), tais propriedades decorrem da convexidade da função $\tau_{ij}(g_{ij})$ e da separabilidade linear do problema entre f e x , garantindo a convergência para o *ótimo global*. Para maiores detalhes de natureza teórica e provas de convergência, ver a definição da *Propriedade P* em (Geoffrion, 1972).

3.2 O Problema Mestre Relaxado

O método de Decomposição de Benders generalizado resolve o problema (17)-(18) baseado na estratégia de relaxação, ignorando todas exceto um pequeno grupo de restrições (18). Esta estratégia requer um procedimento que adicione iterativamente tais restrições ao problema mestre quando necessário.

Então, a cada iteração h , o valor ótimo de $v(x^h)$ é encontrado a partir de (15) após resolver o subproblema para um dado projeto de rede x^h e recuperar-se o vetor dos multiplicadores duais Lagrangeanos α^h . Portanto o valor ótimo de $v(x^h)$ é dado por

$$v(x^h) = \sum_{(i,j) \in E} \sum_{k \in K} -\alpha_{ijk}^h d_k x_{ij}^h + \min_{(f,g) \in G} \sum_{(i,j) \in E} [\tau_{ij} g(ij) + \sum_{k \in K} (c_{ijk} + \alpha_{ijk}^h) f_{ijk}] \quad (19)$$

A partir de (18), tem-se a seguinte restrição associada com α^h :

$$t \geq \sum_{(i,j) \in E} \sum_{k \in K} -\alpha_{ijk}^h d_k x_{ij}^h + \min_{(f,g) \in G} \sum_{(i,j) \in E} [\tau_{ij} g(ij) + \sum_{k \in K} (c_{ijk} + \alpha_{ijk}^h) f_{ijk}] \quad (20)$$

e isolando o valor do mínimo dado por (19), obtém-se o seguinte corte baseado em x^h e α^h , conhecido como o corte de Benders.

$$t \geq v(x^h) + \sum_{(i,j) \in E} \sum_{k \in K} \alpha_{ijk}^h d_k (x_{ij}^h - x_{ij}) \quad (21)$$

Observa-se que a manipulação fez uso da separabilidade do problema como apontado ao final da seção 3.1. Sem tal característica, que advém do acoplamento de natureza linear entre f e x , seria muito mais trabalhoso obter uma expressão fechada para o corte de Benders como a equação (21).

Além disso, utilizar apenas as restrições restantes escritas sobre x , adicionando-se (21), não garante que a solução obtida pelo problema mestre a cada iteração h constitui uma árvore conectando o nodo fonte a todos os nodos em K . Algumas vezes, a solução dada pelo problema mestre relaxado pode implicar em ciclos na topologia proposta ou em sub-árvores desconexas.

A fim de tratar este problema, explora-se a estratégia de trabalhar com ambos os conjuntos de variáveis de x e g no problema mestre relaxado. A idéia aqui é garantir o fechamento do balanço de fluxos no nível superior, evitando a geração de topologias inviáveis.

Através deste artifício, garante-se que todas as soluções obtidas pelo programa mestre serão arborescências. Esta abordagem explora também a informação de acoplamento das variáveis f_{ijk} e g_{ij} do modelo original (2) - (11). Logo, o novo problema mestre relaxado, adicionadas as variáveis g é agora dado por

$$\min \sum_{(i,j) \in E} b_{ij} x_{ij} + t \quad (22)$$

sujeito a:

$$\sum_i x_{ij} \leq 1 \quad \forall j \in V \quad (23)$$

$$\sum_{(i,k) \in E} g_{ik} - \sum_{(k,j) \in E} g_{kj} = d_k \quad \forall k \in K \quad (24)$$

$$\sum_{(i,j) \in E} g_{ij} - \sum_{(j,i) \in E} g_{ji} = 0 \quad \forall j \in V \setminus K \cup \{s\} \quad (25)$$

$$\sum_{(s,j) \in E} g_{sj} = \sum_{k \in K} d_k \quad (26)$$

$$g_{ij} \leq \sum_{k \in K} d_k x_{ij} \quad \forall (i,j) \in E \quad (27)$$

$$t \geq v(x^h) + \sum_{(i,j) \in E} \sum_{k \in K} \alpha_{ijk}^h d_k (x_{ij}^h - x_{ij}) \quad \forall h = 1, \dots, H \quad (28)$$

$$g_{ij} \geq 0 \quad \forall (i,j) \in E$$

$$x_{ij} \in \{0, 1\} \quad \forall (i,j) \in E$$

$$t \geq 0$$

As restrições (24)-(27) são restrições de balanço de fluxo no espaço de variáveis g . Resolver o problema mestre com estas restrições e variáveis adicionais é computacionalmente oneroso, mas evita o uso dos cortes de Benders do tipo II, isto é, cortes associados a raios extremos do poliedro original (2) - (11). Caso contrário, gastaria-se tempo para resolver programas inteiros apenas para descartar configurações de rede inviáveis. Isto decorre da diferença entre o espaço inteiro-viável do problema mestre e o espaço inteiro-viável do problema original (2) - (11), sugerindo o uso de restrições de configuração adicionais para amenizar tais dificuldades. Para maiores detalhes técnicos sobre a geração de cortes de Benders associados a raios-extremos do poliedro de um problema, veja o desenvolvimento envolvendo o *Lema de Farkas* em (Lasdon, 1970).

Na próxima seção, a solução do subproblema de Benders para um dado x^h é investigada.

3.3 Subproblema Não-linear

Desde que $\tau_{ij}(g_{ij})$ é uma função crescente e convexa, o subproblema não-linear obtido fazendo $x = x^h$ tem solução global. Dualizando-se (4) com o auxílio das variáveis $\beta \geq 0$ obtém-se uma decomposição Lagrangeana em dois subproblemas envolvendo os conjuntos de variáveis f e g respectivamente, como mostrado em (29) e (30).

$$r(\beta_{ij}) = \min_{g \geq 0, f \geq 0} \sum_{(i,j) \in E} [\tau_{ij}(g_{ij}) + \sum_{k \in K} c_{ijk} f_{ijk}] + \sum_{((i,j) \in E)} \beta_{ij} \left(\sum_{k \in K} f_{ijk} - g_{ij} \right) \quad (29)$$

logo

$$r(\beta_{ij}) = \min_{f_k \geq 0} \sum_{k \in K} \sum_{(i,j) \in E} (c_{ijk} + \beta_{ij}) f_{ijk} + \min_{g_{ij} \geq 0} \sum_{(i,j) \in E} (\tau_{ij}(g_{ij}) - \beta_{ij} g_{ij}) \quad (30)$$

Dada a hipótese de convexidade de $\tau_{ij}(g_{ij})$, o problema dual Lagrangeano associado a β , que também responde o valor ótimo do subproblema $v(x^h)$ para um dado x^h , é dado por

$$v(x^h) = \max_{\beta \geq 0} r(\beta) \quad (31)$$

Observando que os valores de x^h e g^h são conhecidos para h a partir da solução do problema mestre relaxado, resta determinar o valor ótimo de β associado com a solução corrente. Aplicando-se diretamente a condições de Karush-Kuhn-Tucker ao problema (31), e focando o mínimo mais à direita, $\tau_{ij}(g_{ij}) - \beta_{ij} g_{ij}$, $\forall (i, j) \in E$, determina-se o valor ótimo de cada componente β_{ij} associada com x^h como

$$\beta_{ij}^h = \tau'(g_{ij}^h), \forall (i, j) \in E \quad (32)$$

Portanto, a partir de uma solução ótima x^h do problema mestre relaxado, onde se tem como subproduto o cálculo dos valores de g^h , é trivial obter β^h através da equação (32) o que possibilita a determinação das demais variáveis duais associadas com o subproblema Lagrangeano em f via programação linear.

3.4 Subproblema Linear: Primal e Dual

Fixados os valores ótimos de β_{ij} , $\forall (i, j) \in E$ na iteração h , o subproblema Lagrangeano em f , isto é, o o primeiro termo da equação (30), é escrito para cada destino $k \in K$ como

$$\min \sum_{(i,j) \in E} (c_{ijk} + \beta_{ij}^h) f_{ijk}^h \quad (33)$$

sujeito a:

$$\sum_{(s,j) \in E} f_{sjk}^h = d_k \quad (34)$$

$$\sum_{(i,k) \in E} f_{ik}^h = d_k \quad (35)$$

$$\sum_{(i,j) \in E} f_{ijk}^h = \sum_{(j,i) \in E} f_{jik}^h \quad \forall j \in V \setminus \{s\}, k \neq j \quad (36)$$

$$-f_{ijk}^h \geq -d_k x_{ij}^h \quad \forall (i, j) \in E \quad (37)$$

$$f_{ijk}^h \geq 0 \quad \forall (i, j) \in E$$

O subproblema primal (33)-(37) possui uma única e trivial solução dada por (38)

$$f_{ijk}^h = \begin{cases} d_k & \text{se } (i, j) \in C_{ks}^h \subseteq A^h \\ 0 & \text{caso contrário} \end{cases} \quad (38)$$

Em que $C_{ks}^h \subseteq A^h$ é o subconjunto de arcos da arborescência A^h que pertence ao caminho do produto k desde a origem s . Com a solução do subproblema primal, pode-se calcular o valor de $v(x^h)$ facilmente a cada iteração de Benders. Infelizmente, não se produz valores ótimos das variáveis duais α_{ijk}^h . Portanto é preferível trabalhar a versão dual de (33)-(37). Associando os variáveis duais $p_{sk}^h, p_{kk}^h, p_{jk}^h$ e α_{ijk}^h às restrições (34), (35), (36) e (37), respectivamente, e lembrando que x_{ij}^h está fixado pelo o problema mestre relaxado, escreve-se o problema dual de (33)-(37) para cada destino $k \in K$ como

$$\max_{p^h, \alpha^h \geq 0} d_k (p_{kk}^h - p_{sk}^h - \sum_{(i,j) \in E} x_{ij}^h \alpha_{ijk}^h) \quad (39)$$

s.a:

$$p_{jk}^h - p_{ik}^h - \alpha_{ijk}^h \leq c_{ijk} + \beta_{ij}^h, \forall (i, j) \in E \quad (40)$$

Ressalta-se que enquanto o subproblem linear primal tem solução única, sua contraparte dual pode ter múltiplas soluções ótimas tornando a escolha adequada de variáveis duais ótimas um tópico à parte. No presente trabalho, optou-se por descartar técnicas de geração de cortes *Pareto-ótimos*. As variáveis duais ótimas foram determinadas através da solução de (39)-(40) via algoritmo padrão de programação linear. Doravante, nos referimos à técnica apresentada nesta seção como *Generalized Benders Decomposition* ou mais simplesmente *GBD*.

4 APROXIMAÇÃO EXTERNA

A técnica conhecida como Aproximação Externa(OA) (do inglês *Outer Approximation*) foi desenvolvida em trabalhos

pioneiros de (Duran and Grossman, 1986), (Fletcher and Leyfer, 1994) e (Yuan et al., 1988). É um método simples mas eficiente para resolução de problemas inteiros não-lineares, utilizando uma abordagem de planos cortantes. A técnica de Aproximação Externa tem sido aplicada em otimização de síntese de processos (Grossmann and Kravanja, 1995), (Karupiah et al., 2008), e em várias outras aplicações de projeto em engenharia. Mais recentemente, o método tem sido aplicado a sistemas logísticos como em (Huang et al., 2005) e em problemas de otimização em geral (Grossmann and Kravanja, 1995).

O método também possui uma técnica de coordenação entre o problema mestre e o subproblema, trabalhando de forma similar à técnica GBD, mas os problemas mestres de OA são escritos no espaço de todas as variáveis do problema, dispensando a projeção no espaço das variáveis inteiras. Este recurso melhora o poder dos planos cortantes gerados e assegura a convergência em um número de iterações menor que o obtido ao se aplicar a técnica de GBD. Como desvantagem, a solução do problema mestre é mais cara, geralmente aumentando o tempo de solução e o esforço computacional à medida que a o tamanho da instância aumenta. Para entender o desenvolvimento da técnica de OA, uma revisão geral do método é necessária. Dado um problema não-linear inteiro misto em sua representação algébrica mais básica, onde f e x representam o conjunto das variáveis contínuas e discretas, respectivamente, $\phi : \mathbb{R}^{qxs} \mapsto \mathbb{R}$ e $\gamma : \mathbb{R}^{qxs} \mapsto \mathbb{R}^m$ são duas funções continuamente diferenciáveis, e F e X são conjuntos poliedrais

$$\min \quad \phi(x, f) \quad (41)$$

s.a:

$$\gamma_j(x, f) \leq 0 \quad \forall j = 1, \dots, m \quad (42)$$

$$x \in X, X \in \mathbb{Z}^q \quad (43)$$

$$f \in F \quad (44)$$

É possível reduzir este problema a um problema não-linear ao escolher um vetor fixo $x = x^h, f \in \mathbb{F}$, para cada iteração h , produzindo o seguinte subproblema não-linear

$$\min \quad \phi(x^h, f) \quad (45)$$

s.a:

$$\gamma_j(x^h, f) \leq 0 \quad \forall j = 1, \dots, m \quad (46)$$

$$f \in F$$

Quando o problema acima é resolvido (45)-(46), este permite deduzir os gradientes das funções $\phi(x, f)$ e $\gamma_j(x, f)$, $\forall j$ em

(x^h, f^h) . Então é possível reescrever o problema (41)-(44) como sendo equivalente ao problema linear inteiro misto

$$\min \quad \xi \quad (47)$$

s.a:

$$\xi \geq \phi(x^h, f^h) + \nabla \phi(x^h, f^h) \begin{pmatrix} x - x^h \\ f - f^h \end{pmatrix} \quad \forall h = 1, \dots, H \quad (48)$$

$$0 \geq \gamma(x^h, f^h) + \nabla \gamma(x^h, f^h) \begin{pmatrix} x - x^h \\ f - f^h \end{pmatrix} \quad \forall h = 1, \dots, H \quad (49)$$

$$\xi \geq 0 \quad (50)$$

$$x \in X, X \in \mathbb{Z}^q$$

$$f \in F$$

onde H representa o número máximo de iterações e ξ representa uma variável de estimativa para função objetivo.

O problema (47)-(50) é conhecido como problema mestre OA. As restrições (48)-(49) são responsáveis pela aproximação externa da função objetivo e de sua região viável, respectivamente. Quando as funções $\gamma(x, f)$ são convexas e as restrição de qualificação são atendidas para cada solução de (45) - (46), então as restrições (49) são necessárias e suficientes para o método OA aproximar externamente a região viável. Ressalta-se que as restrições (49) tem dupla função no método OA: além de assegurar a viabilidade, são responsáveis por realizar a aproximação externa da região viável, se $\gamma(x, f)$ for não-linear.

No caso da formulação (2)-(11), a função objetivo é separável em termos lineares e um único termo não-linear, $(\sum_{(i,j) \in E} \tau_{ij}(g_{ij}))$. Logo, este é o único termo que precisa sofrer a aproximação externa. Portanto pode-se substituir somente $\tau_k(g_k)$ por ξ_k para cada k na função objetivo e adicionar a restrição (48) no formato de (52).

$$\min \sum_{(i,j) \in E} [b_{ij}x_{ij} + \xi_{ij} + \sum_k c_{ijk}f_{ijk}] \quad (51)$$

s.a: (5) - (11), (23) - (27) e

$$\xi_{ij} \geq \tau_{ij}(g_{ij}^h) + \beta_{ij}^h(g_{ij} - g_{ij}^h) \quad \forall (i,j) \in E, \forall h = 1, \dots, H \quad (52)$$

$$\xi_{ij} \geq 0 \quad \forall (i,j) \in E \quad (53)$$

A restrição (49) não está presente na formulação (51)-(53), porque as restrições de acoplamento são lineares (poliedrais), tornando desnecessária a inserção destes cortes. Além do mais, como esperado, esta formulação é muito grande para ser resolvida eficientemente. A grande dimensão do conjunto de variáveis f representa grande sobrecarga durante a resolução do problema mestre OA. Uma maneira de lidar com esta desvantagem é projetar essas variáveis para fora da formulação, desde que algumas manipulações sejam realizadas. A idéia aqui é permitir a solução do problema mestre OA, por meio de algoritmo de decomposição de Benders. Criase então um método híbrido: lida-se com a não-linearidade via OA e com a parte de grande escala via Decomposição de Benders. Isto implica que as famílias de planos cortantes (55) - (56) podem ser adicionadas ao problema mestre em qualquer ordem.

$$\min \sum_{(i,j) \in E} [b_{ij}x_{ij} + \xi_{ij} + \sum_k t_k] \quad (54)$$

s.a: (10) - (11), (23) - (27) e

$$\xi_{ij} \geq \tau(g_{ij}^h) + \beta_{ij}^h(g_{ij} - g_{ij}^h) \quad \forall (i,j) \in E, \\ \forall h = 1, \dots, H \quad (55)$$

$$t_k \geq v_k(x^l) + \sum_{(i,j)} \alpha_{ijk}^l d_k(x_{ij}^l - x_{ij}) \quad \forall k \in K, \\ \forall l = 1, \dots, L \quad (56)$$

$$\xi_{ij} \geq 0 \quad \forall (i,j) \in E \\ t_k \geq 0 \quad \forall k \in K$$

Destaca-se que a estratégia de solução proposta tanto pode ser vista como a solução do programa mestre OA via decomposição de Benders, quanto como a solução do mestre (não-linear inteiro) de Benders via OA. Não há entrelaçamento entre as duas famílias de cortes: dada uma solução do mestre (x, g) , tanto cortes OA quanto cortes de Benders podem ser deduzidos sem ordem pré-imposta. De agora em diante, esta estratégia de solução recebe o nome de *OA-Híbrido*.

É importante ressaltar que, assim como a técnica GBD, o método de aproximação externa tem prova de convergência global. Tal prova depende da convexidade das funções não-lineares envolvidas e do espaço de soluções viáveis ser igualmente convexo, desde que as *Condições de Slater* sejam verificáveis para cada solução de (45)-(46). Para questões adicionais associadas a tais condições de convergência, verificar (Fletcher and Leyfer, 1994).

5 EXPERIMENTOS COMPUTACIONAIS

Para avaliar a performance dos algoritmos propostos, dois conjuntos de experimentos computacionais foram executa-

dos. Para estes dois conjuntos, foram utilizadas o conjunto de instâncias disponíveis na biblioteca online QAPLIB, disponível em <http://www.seas.upenn.edu/qaplib/inst.html>. As instâncias de teste foram retiradas entre aquelas propostas por Nugent et al., por serem consideradas as de maior dificuldade, uma vez que todas correspondem a grafo completo, possuindo $(n^2 - n)$ arcos e todos os nodos com grau $(n - 1)$. As instâncias selecionadas também possuem matrizes de distâncias baseadas na *Distância Manhattan*, o que eleva o número de níveis de simetria entre soluções e em geral torna mais difícil obter prova de otimalidade.

Os nomes das instâncias foram codificadas como n , onde n é o número de nodos da rede. As instâncias tem tamanho de 12, 14, 15, 16, 17, 18, 20, 21, 22, 24, 25, 27 respectivamente. Para cada instância, os valores do fator de congestionamento aplicados foram 0,001, 0,01 e 0,1. O número de variáveis inteiras e contínuas, e os tamanhos dos conjuntos E e K são mostrados na Tabela 1.

Tabela 1: Espaço das Variáveis das 12 instâncias

Instância	E	K	Num Var Int	Num Var Cont
12	132	5	132	729
14	182	3	182	728
15	225	3	225	900
16	256	5	256	1536
17	289	5	289	1734
18	306	6	306	2134
20	380	5	380	2280
21	420	5	420	2520
22	462	5	462	2772
24	552	4	552	2760
25	600	4	600	3000
27	702	5	702	4212

Os dois conjuntos de testes demonstram como a infraestrutura da rede se altera quando o fator de congestionamento é incrementado. Os testes computacionais foram escritos na linguagem AMPL CPLEX utilizando um sistema operacional Linux 64 bits. Os experimentos foram testados em uma estação de trabalho com um processador Intel Core 2 Quad 2.4 GHz e 8Gb de RAM, com tempo limite de 86.400 segundos (24 horas). Em todos os testes conduzidos a lei de potência que modela os custos de congestionamento (1) foi escrita com $p = 2$.

A Figura (1) ilustra um exemplo do impacto do congestionamento em uma instância de 18 nodos. A Figura 1-(a) apresenta os nodos e as conexões possíveis, na qual os vértices

em negrito e em branco são os nodos de demanda unitária e os de transbordo, respectivamente; e os valores ao lado das arestas representam o comprimento das mesmas. O sorvedouro é representado pelo círculo 1. Na Figura 1-(b), a configuração ótima sem considerar os impactos de congestionamento é mostrada, enquanto na Figura 1-(c) tem-se a estrutura ótima ao se incluir os efeitos de degradação de nível de serviço ($e = 0,1$). Nestas duas últimas figuras, os valores ao lado das arestas são o fluxo total passando pelas mesmas. Pode-se ver nas Figuras 1-(b) e 1-(c), que as arestas (2, 7), (7, 6), (14, 13) e (13, 18) deixam de ser ativadas ao se considerar os efeitos de congestionamento.

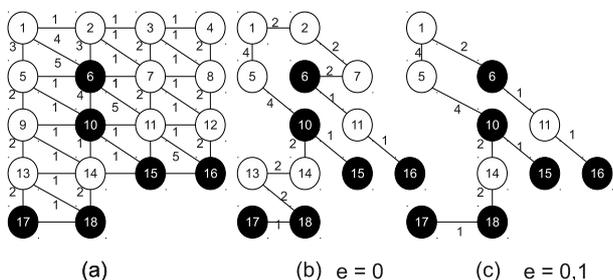


Figura 1: Exemplo de um caso com 18 nodos.

A comparação entre os algoritmos de GBD e OA-Híbrido para cada instância são apresentadas na Tabela 2. Para cada conjunto de testes, foram analisados o valor do custo de congestionamento (e), o número de iterações (It), os limites inferior (LB) e superior (UB), o *gap* de otimalidade, onde $gap = (UB - LB)/UB$, e o tempo computacional gasto, medido em segundos ($T[s]$).

A Tabela 2 demonstra que o método OA-Híbrido apresenta os melhores desempenhos para níveis maiores de congestionamento. Em média, o método é 406,2 vezes mais rápido que o GBD. Para níveis médios ($e = 0,01$), o método OA-Híbrido é apenas 37,3 vezes melhor. Observando o número de iterações, o método OA-Híbrido requer em média 158,2 menos iterações do que o GBD.

Isso pode ser melhor entendido ao se observar a redução do efeito *tail-off* na convergência dos métodos. A Figura 2 apresenta a convergência dos dois métodos ao resolver a instância 17. Nesta figura os limites LB e UB são plotados em função do número de iterações. Como se pode ver, o método GBD requer um número elevado de iterações até a convergência dos dois limites. Esse comportamento de convergência é conhecido como *tail-off*: é relativamente baixo o custo computacional para calcular limites próximos do ótimo, mas muito oneroso zerar efetivamente o *gap* de otimalidade. No caso do OA-Híbrido, os limites superior e inferior convergem muito rapidamente para o ótimo.

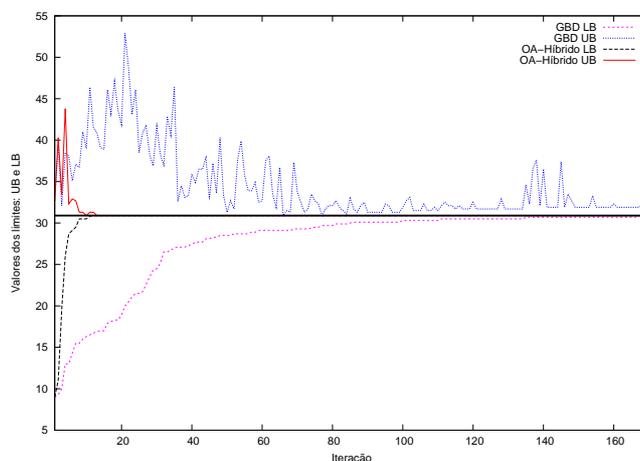


Figura 2: Efeito *tail-off* da instância 17 com custo de congestionamento igual a 0,1

Além disso, o tempo computacional em função do número de nodos é apresentado na Figura 3. O aumento do número de nodos e consequentemente o aumento do número de arcos não influencia tanto o desempenho do método OA-Híbrido quanto o desempenho do método GBD. Este último possui um desempenho que piora sensivelmente ao se aumentar o tamanho da rede.

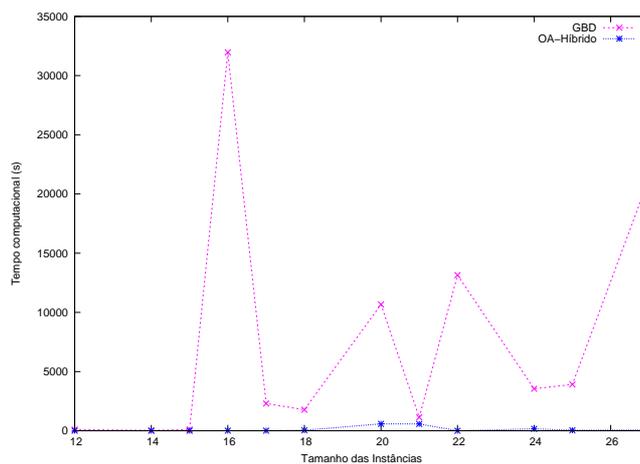


Figura 3: Comparação de Tempo Computacional para todas as instâncias com custo de congestionamento igual a 0,1

Finalmente, ainda na Tabela 2, há algumas instâncias de menor número de nodos com tempos de computação ligeiramente superiores aos de outras instâncias maiores, como por exemplo ao se comparar os tempos de solução das instâncias 20 e 21 com 25 e 27. Esse resultado não é inesperado, uma vez que a dificuldade de resolução de um problema de pro-

Tabela 2: Tabela comparativa entre os métodos GDB e OA-Híbrido.

Inst	e	GDB					OA-Híbrido					
		It	LB	UB	gap	T[s]	e	It	LB	UB	gap	T[s]
12	0,001	8	22,0350	22,0350	0,000	1	0,001	6	22,0350	22,0350	0,000	1
	0,01	10	22,3500	22,3500	0,000	1	0,01	6	22,3500	22,3500	0,000	1
	0,1	89	24,9000	24,9000	0,000	100	0,1	15	24,9000	24,9000	0,000	3
14	0,001	13	21,0150	21,0150	0,000	1	0,001	7	21,0150	21,0150	0,000	0
	0,01	14	21,1500	21,1500	0,000	2	0,01	7	21,1500	21,1500	0,000	0
	0,1	32	22,5000	22,5000	0,000	20	0,1	8	22,5000	22,5000	0,000	1
15	0,001	10	22,0300	22,0300	0,000	3	0,001	5	22,0300	22,0300	0,000	1
	0,01	7	22,3000	22,3000	0,000	2	0,01	5	22,3000	22,3000	0,000	0
	0,1	50	24,5000	24,5000	0,000	104	0,1	10	24,5000	24,5000	0,000	2
16	0,001	9	26,0560	26,0560	0,000	1	0,001	6	26,0560	26,0560	0,000	0
	0,01	10	26,5600	26,5600	0,000	3	0,01	6	26,5600	26,5600	0,000	1
	0,1	301	30,0000	30,0000	0,000	31.965	0,1	13	30,0000	30,0000	0,000	10
17	0,001	20	28,0350	28,0350	0,000	7	0,001	8	28,0350	28,0350	0,000	5
	0,01	18	28,3500	28,3500	0,000	9	0,01	8	28,3500	28,3500	0,000	4
	0,1	170	30,9000	30,9000	0,000	2.297	0,1	13	30,9000	30,9000	0,000	10
18	0,001	11	25,0310	25,0310	0,000	4	0,001	10	34,0550	34,0550	0,000	7
	0,01	32	25,3100	25,3100	0,000	4	0,001	10	34,0550	34,0550	0,000	7
	0,1	126	27,6000	27,6000	0,000	1.775	0,1	14	37,8000	37,8000	0,000	37
20	0,001	54	25,0140	25,0140	0,000	45	0,001	10	32,0350	32,0350	0,000	120
	0,01	37	25,1400	25,1400	0,000	10	0,01	12	32,3500	32,3500	0,000	78
	0,1	156	26,4000	26,4000	0,000	10.665	0,1	21	34,9000	34,9000	0,000	581
21	0,001	38	28,0140	28,0140	0,000	19	0,001	8	43,0570	43,0570	0,000	2
	0,01	42	28,1400	28,1400	0,000	136	0,01	10	43,5700	43,5700	0,000	9
	0,1	92	29,6000	29,6000	0,000	1.129	0,1	22	48,4000	48,4000	0,000	571
22	0,001	120	32,0050	32,0050	0,000	80	0,001	12	33,0050	32,0050	0,000	1
	0,01	117	32,0500	32,0500	0,000	67	0,01	12	32,0500	32,0500	0,000	2
	0,1	145	32,0500	32,0500	0,000	841	0,1	10	32,0500	32,0500	0,000	1
24	0,001	79	29,0140	29,0140	0,000	67	0,001	9	33,0310	33,0310	0,000	105
	0,01	207	23,0500	23,0500	0,000	268	0,01	12	33,3100	33,3100	0,000	67
	0,1	249	23,5000	23,5000	0,000	3.548	0,1	12	36,1000	36,1000	0,000	157
25	0,001	186	22,0050	22,0050	0,000	263	0,001	7	29,0300	29,0300	0,000	3
	0,01	169	22,0500	22,0500	0,000	383	0,01	10	29,3000	29,3000	0,000	4
	0,1	211	22,5000	22,5000	0,000	3.908	0,1	12	32,0000	32,0000	0,000	40
27	0,001	333	29,0050	29,0050	0,000	2.603	0,001	8	44,0310	44,0310	0,000	4
	0,01	351	29,0500	29,0500	0,000	2.017	0,01	12	44,3100	44,3100	0,000	8
	0,1	438	29,5000	29,5000	0,000	21.304	0,1	11	47,1000	47,1000	0,000	68

gramação não-linear inteira mista não é dada exclusivamente pelo seu tamanho. Tal esforço de computação é resultado da interferência de diversos fatores, tipicamente dos *trade-offs* e das heterogeneidades que a estrutura de custos da instância propicia, e até mesmo de questões relativas a instabilidades de natureza numérica.

6 CONCLUSÕES

Este trabalho apresentou duas técnicas diferentes para a resolução de problemas de projeto de rede de grande escala sob congestionamento: A técnica GDB e o método de OA-Híbrido. Ambas foram capazes de resolver instâncias de até 27 nodos (702 arcos) utilizando a linguagem AMPL. Os resultados foram obtidos em tempo razoável, chegando a um tempo computacional de 68 segundos para o método OA-Híbrido com 27 nodos (702 arcos) contra 21304 segundos para o método GDB.

De modo geral a abordagem proposta para a solução de programas não-lineares inteiros mistos como o modelo de projeto de rede com custos não-lineares se comporta bem, e o esquema de decomposição de Benders generalizado associ-

ado a técnica de Aproximação Externa é bem sucedido ao tratar um problema de tais características. Isso acontece devido à grande redução do efeito *tail-off* e também à grande utilidade dos cortes de Benders para melhorar a relaxação de programação linear do problema mestre.

Pesquisas futuras devem agregar cortes Pareto-ótimos implementados via Branch-and-Cut e a medição do *trade-off* entre o uso de restrições lineares de configuração e de cortes de viabilidade.

REFERÊNCIAS

- Altıparmak, F., Dengiz, B. and Smith, A. E. (2003). Optimal design of reliable computer networks: A comparison of metaheuristics, *Journal of Heuristics* **9**: 471–487.
- Benchakroun, A., Ferland, J. and Cleroux, R. (1992). Distribution-system planning through a generalized Benders decomposition approach, *European Journal of Operational Research* **62**(2): 149–162.
- Benders, J. (1962). Partitioning procedures for solving mixed integer variables programming problems, *Numerische*

Mathematik **4**: 238–252.

- Contreras, I., Fernández, E. and Marín, A. (2009). Tight bounds from a path based formulation for the tree of hubs location problem, *Computers and Operations Research* **36**(12): 3117–3127.
- Contreras, I., Fernández, E. and Marín, A. (2010). The tree of hubs location problem, *European Journal of Operational Research* **202**: 390–400.
- Cordeau, J., Pasin, F. and Solomon, M. (2006). An integrated model for logistics network design, *Annals of Operations Research* **144**: 59–82.
- Costa, A. (2005). A survey on benders decomposition applied to fixed-charge network design problems, *Computers and Operations Research* **32**: 1429–1450.
- Dantzig, G. (1962). *Linear Programming and Extensions*, Princeton University Press.
- Duran, M. and Grossman, I. (1986). An outer-approximation algorithm for a class of mixed-integer nonlinear programs, *Mathematical Programming* **36**: 307.
- Fletcher, R. and Leyfer, S. (1994). Solving mixed integer nonlinear programs by outer approximation, *Mathematical Programming* **66**: 327.
- Geoffrion, A. (1972). Generalized benders decomposition, *Journal of Optimization Theory and Applications* **10**: 237–260.
- Grossmann, I. and Kravanja, Z. (1995). Mixed-integer nonlinear programming techniques for process systems engineering., *Computers and Chemical Engineering* **19**: 189–204.
- Huang, S., Batta, R. and Nagi, R. (2005). Distribution network design: Selection and sizing of congested connections, *Naval Research Logistics* **52**(8): 701–712.
- Hwang, F. and Richards, D. (1992). Steiner tree problems, *Networks* **22**: 55–89.
- Karuppiah, R., Furman, K. C. and Grossmann, I. E. (2008). Global optimization for scheduling refinery crude oil operations, *Computers and Chemical Engineering* **32**(11): 2745–2766.
- Klinecicz, J. G. (1998). Hub location in backbone/tributary network design: a review, *Location Science* **6**: 307–335.
- Lasdon, L. (1970). *Optimization Theory for Large Systems*, Macmillan, London.
- Maculan, N. (1987). The Steiner problem in graphs, *Annals of Discrete Mathematics* **31**: 185–212.
- Ramirez-Rosado, I. J. and Dominguez-Navarro, J. A. (2006). New multiobjective tabu-search algorithm for fuzzy optimal planning of power distribution systems, *Transactions on Power Systems* **21**: 224–233.
- Randazzo, C. and Luna, H. (2001). A comparison of optimal methods for local access uncapacitated network design, *The Annals of Operations Research* **106**: 263–286.
- Yuan, X., Zhang, S., Piboleau, S. and Domenech, S. (1988). Une methode d'optimisation nonlineare an variables pour la conception de procedes., *Operations Research* **33**: 331.