4
Uma formulação com número exponencial de variáveis e restrições para o CVRP

Algoritmos de geração de cortes e colunas, conhecidos como *branch-and-cut-and-price(BCP)*, são algoritmos de *branch-and-bound* onde os limites duais são obtidos resolvendo um programa linear com número exponencial de linhas e colunas. Um algoritmo de BCP é dito *robusto* quando a estrutura dos subproblemas de separação e de geração de colunas permanece inalterada durante a execução do algoritmo. O primeiro BCP é normalmente atribuído a Nemhauser e Park [58] para o problema de coloração de arestas. O algoritmo não era robusto, pois os novos cortes adicionados ao problema mestre complicavam o subproblema de geração de colunas, tendo esse que ser resolvido por técnicas complicadas de programação inteira. Algoritmos de BCP não-robustos não são eficientes na maioria dos casos.

No final dos anos 90, alguns pesquisadores, de forma independente, notaram que cortes expressos em termos de variáveis da formulação original podiam ser dinamicamente separados, transformados e adicionados ao problema mestre. Esses cortes não modificam a estrutura do subproblema de geração de colunas. Tal observação permitiu a construção de algoritmos de BCP verdadeiramente robustos [73, 43, 44, 68, 29, 10].

Nesse capítulo apresenta-se uma formulação com número exponencial de variáveis e restrições para o CVRP. Essa formulação permitiu a construção de um algoritmo de BCP robusto para o problema. Inicialmente apresenta-se a técnica geral descrita em [61] para a construção de algoritmos de BCP robustos e, depois, sua aplicação para o CVRP, dando origem à formulação que permitiu a construção do algoritmo de BCP que é o principal resultado dessa dissertação.

4.1 Técnicas para construção de Algoritmos de BCP Robustos

Duas contribuições essenciais para a construção de algoritmos de BCP robustos são dadas em [61]. Apresenta-se uma alternativa para o tradicional problema mestre de Dantzig-Wolfe para a geração de colunas [72], o *Mestre Explícito*, que torna o mecanismo para o BCP robusto bem evidente, além de trazer outros benefícios potenciais. Também é apresentada uma solução para o problema de simetria de variáveis que limita a construção de algoritmos de BCP robustos efetivos para muitos problemas importantes, como problemas de empacotamento, coloração de grafos, etc. Nessa dissertação, apresentamos apenas a primeira contribuição, já que a segunda não se aplica ao CVRP.

Outra contribuição importante de [61] é simplesmente apontar o potencial de impacto vasto dos algoritmos de BCP na prática de programação inteira. Até o momento, algoritmos de BCP foram aplicados apenas a poucos problemas onde um algoritmo de geração de colunas puro (*branch-and-price*), funciona bem. Nessa dissertação, mostra-se os excelentes resultados obtidos no CVRP. Porém, resultados muito bons foram obtidos para outros problemas de otimização combinatória como o problema da alocação generalizada (*GAP*, do inglês *generalized assignment problem*)[62] e o problema da arvóre geradora mínima com restrição de capacidade (*CAPMST*, do inglês *capacitated minimum spanning tree*)[31].

4.1.1 Mestre Dantzig-Wolfe x Mestre Explícito

Seja o seguinte politopo R_{10} com n variáveis

$$Ax = b, \tag{4-1}$$

$$Dx \leq d, \tag{4-2}$$

$$x \in Z_+^n. \tag{4-3}$$

Então, pode-se definir o seguinte programa inteiro

$$\min\{cx : x \in R_{10}\}. \tag{4-4}$$

Assume-se que $\mathcal{Q} = \{x \in Z^n_+ | Dx \leq d\}$ é um conjunto *finito* com elementos $x^1, \cdots, x^p$. Seja Q uma matriz $n \times p$ onde cada coluna corresponde a um elemento de $\mathcal{Q}$. Pode-se,então, definir o seguinte politopo, R_{11},

$$\mathbf{1}\lambda = 1, \tag{4-5}$$

$$\lambda \in \{0,1\}^p. \tag{4-6}$$

Existe uma correspondência um-para-um entre os elementos de $\mathcal{Q}$ e as soluções de

$$\{x = Q\lambda : \lambda \in R_{11}\}. \tag{4-7}$$

Seja o politopo, R_{12}, formado por (4-5) mais as inequações

$$(AQ)\lambda = b, \tag{4-8}$$

$$\lambda \geq 0. \tag{4-9}$$

A reformulação tradicional por programação inteira de (4-4) consiste em substituir x pela expressão (4-7). A partir do politopo R_{12} pode-se definir a formulação

$$\min\{(cQ)\lambda : \lambda \in R_{12}\}. \tag{4-10}$$

A formulação (4-10) é conhecida como programa Mestre de Dantzig-Wolfe Linear (na realidade todas as formulações descritas nessa dissertação são lineares, porém aqui quando é dito linear está-se referindo à relaxação de Programação Linear). A relaxação é feita substituindo-se (4-6) por (4-9).

A formulação (4-10) será chamada de (DWM). O valor da solução ótima de DWM, que será chamada de Z_{DWM}, é um limite inferior que pode ser melhor que o dado pela relaxação linear de (4-4), pois $Z_{DWM} = \{\min cx | Ax = b, x \in Conv(Dx \leq d, x \in Z_+^n)\}$. Entretanto, como p cresce exponencialmente, a técnica de geração de colunas se apresenta como a mais adequada para a resolução de (DWM). Seja o politopo, R_{13}, definido pelas inequações (4-2) e (4-3). O programa inteiro, abaixo, definido a partir de R_{13}, permite determinar a coluna (variável) de (DWM) que possui o menor custo reduzido em uma iteração do método Simplex Revisado. Nele os coeficientes das colunas de (DWM) são as variáveis, enquanto que seus parâmetros são as variáveis duais associadas à (DWM) juntamente com as restrições que definem as colunas,

$$\min \{(c - \mu A)x - \nu : x \in R_{13}\}, \tag{4-11}$$

onde μ e ν são as variáveis duais associadas às linhas de (DWM). O esquema de geração de colunas é prático somente quando as restrições (4-2) possuem uma "boa estrutura", que permite o problema de geração de colunas ser resolvido em tempo viável para este contexto. É essencial manter o subproblema de geração de colunas com a estrutura inalterada após a adição de cortes e após fazer fixar variáveis (*branching*).

Seja $\overline{\lambda}$ a solução fracionária de (DWM). Define-se $\overline{x} = Q\overline{\lambda}$. Considere separar um corte válido $a^i x \leq b_i$ tal que $a^i \overline{x} > b_i$. A restrição $(a^i Q)\lambda \leq b_i$ pode ser adicionada a (DWM), como se $a^i x \leq b_i$ pertencessem às restrições originais (4-1) em (4-4). O subproblema de geração de colunas continua sendo (4-11), a única diferença sendo a linha extra em A e o elemento extra de μ (nova variável dual) que tem que ser considerado no cálculo de $(c - \mu A)$, ou seja, haverá apenas modificações nos valores dos coeficientes de x na função objetivo.

A seguir, será mostrada a formulação para um programa linear que será chamado de *Mestre Explícito*. Considere novamente o problema original (4-4). Introduz-se variáveis adicionais x' a (4-4), definidas como sendo iguais a x. Tem-se, então, o politopo R_{14}, formado por (4-1) e (4-3) mais as inequações a seguir:

$$x' - x = 0, \tag{4-12}$$

$$Dx' \leq d, \tag{4-13}$$

$$x' \in Z_+^n. \tag{4-14}$$

A partir do politopo R_{14}, define-se a formulação abaixo:

$$\min\{cx : x \in R_{14}\}. \tag{4-15}$$

Considerando a formulação (4-15), substituindo x' por sua expressão equivalente dada por (4-7) e relaxando as restrições de integralidade (4-3), tem-se o politopo R_{16}, formado por (4-1), (4-5) e (4-9) mais as restrições a seguir:

$$Q\lambda - x = 0, \tag{4-16}$$

$$x \geq 0. \tag{4-17}$$

A partir do politopo R_{16} define-se a formulação

$$\min\{cx : x \in R_{16}\}. \tag{4-18}$$

A formulação (4-18) será chamada de *Mestre Explícito Linear* e será referenciada por (EM) (chama-se de Z_{EM} a solução ótima de (EM)). Sejam π, ν e μ as variáveis duais associadas aos três conjuntos de restrições em (EM), (4-16), (4-5) e (4-1), respectivamente. As colunas correspondentes às variáveis λ são obtidas resolvendo o seguinte programa inteiro (IP, do inglês *Integer Program*):

$$\min\{-\pi x - \nu : x \in R_{13}\}. \tag{4-19}$$

As variáveis μ não aparecem no subproblema de geração de colunas. Fica claro, então, que separar cortes sobre x e adicioná-los à $Ax = b$ não altera a estrutura do subproblema, em particular, nem mesmos os coeficientes da função objetivo são modificados (diretamente como na decomposição em DWM).

Como esperado, $Z_{DWM} = Z_{EM}$. O tamanho de (DWM) é menor que o tamanho de (EM). Logo, para que seja vantajoso construir um BCP robusto sobre (EM) precisa-se ter algum benefício. Duas vantagens potenciais de se trabalhar sobre (EM) citadas em [61] são a facilidade de fazer fixação por custos reduzidos e de realizar geração de colunas múltiplas.

4.2 Limites inferiores para o CVRP

Aqui apresenta-se dois limites inferiores para o CVRP, baseados nas formulações descritas no capítulo 2 e que serão utilizados para a obtenção da formulação que permitiu a construção do algoritmo de BCP robusto para o problema. Primeiramente, mostra-se o limite inferior utilizando somente geração de cortes. Depois, apresenta-se o limite inferior utilizando somente geração de colunas.

4.2.1 Limite inferior por geração de cortes

Baseado na formulação por arestas (com número polinomial de variáveis e exponencial de restrições) apresentada na seção 2.1, pode-se definir o politopo P_1:

$$\sum_{e \in \delta(i)} x_e = 2, \quad i \in V_+, \tag{4-20}$$

$$\sum_{e \in \delta(0)} x_e = 2K, \tag{4-21}$$

$$\sum_{e \in E(S,\bar{S})} x_e \geq 2\left\lceil d(S)/C \right\rceil, \quad S \subseteq V_+, \ |S| \geq 2, \tag{4-22}$$

$$0 \leq x_e \leq 1, \ \ e \in E \setminus \delta(0), \tag{4-23}$$

$$0 \leq x_e \leq 2, \ \ e \in \delta(0). \tag{4-24}$$

Limitando os valores de x_e a valores inteiros, os vetores x em P_1 definem todas as soluções viáveis para o CVRP. Considerando P_1, tem-se o limite inferior, que será chamado de L_1, dado por

$$\min\left\{\sum_{e\in E}\ell_e x_e : x \in P_1\right\} \tag{4-25}$$

para a solução ótima do CVRP. Como existe um número exponencial de restrições do tipo (4-22), o limite inferior dado por L_1 deve ser computado através de um algoritmo de planos de cortes.

4.2.2

Limite inferior por geração de colunas

Um conceito importante para a formulação por geração de colunas para o CVRP é o de *q-rota*, que foi apresentado na seção 3.1. O conjunto de rotas válidas para o CVRP é um subconjunto do conjunto de q-rotas.

Assim a relaxação linear por geração de colunas para o CVRP, define variáveis (colunas) que correspondem a q-rotas sem 2-ciclos (subcaminhos $i \to j \to i, i \neq 0$). Restringir as q-rotas somente às que não possuem 2-ciclos fortalece a formulação e não muda a complexidade do sub-problema que precisa ser resolvido para encontrá-las [21]. Conforme já dito na seção 3.1, quanto maior o tamanho dos ciclos proibidos nas q-rotas, mais forte é a relaxação.

Seja Q uma matriz $m \times p$ onde as colunas são vetores de incidência das arestas em cada uma das q-rotas que estão no LP (p colunas). Seja q_j^e o coeficiente associado a aresta e na j-ésima coluna de Q. Este coeficiente corresponde ao número de vezes em que a q-rota passa pela aresta e. Considere o seguinte politopo (P_2) definido a seguir:

$$\sum_{j=1}^{p} q_j^e \cdot \lambda_j - x_e = 0, \quad \forall e \in E, \tag{4-26}$$

$$\sum_{j=1}^{p} \lambda_j = K, \tag{4-27}$$

$$\sum_{e\in\delta(\{i\})} x_e = 2, \quad \forall i \in V_+, \tag{4-28}$$

$$x_e \geq 0 \quad ,\forall e \in E, \tag{4-29}$$

$$\lambda_j \geq 0, \quad \forall j \in \{1, \cdots, p\}. \tag{4-30}$$

As restrições (4-26) definem a relação entre as variáveis x e λ. A restricão (4-27) define o número de veículos a ser usado. Se os valores de x em P_2 forem restritos a valores inteiros, pode ser mostrado que o conjunto de vetores de P_2 também define todas as soluções viáveis para o CVRP. Por causa do número exponencial de variáveis λ, o limite inferior, que será chamado de L_2, dado por

$$\min \left\{ \sum_{e \in E} \ell_e x_e : x \in P_2 \right\} \tag{4-31}$$

deve ser calculado utilizando geração de colunas ou relaxação Lagrangeana.

A descrição do poliedro P_2 associado com a geração de colunas ou com a relaxação Lagrangeana foi feita em termos de dois conjuntos de variáveis (λ e x) o que é chamado na seção 4.1 de Mestre Explícito. O politopo P_2 descrito nessa seção difere um pouco do politopo descrito na seção 2.3 onde se apresenta uma formulação com número exponencial de variáveis e polinomial de restrições porém, em ambos os casos, a técnica de geração de colunas deve ser usada para a obtenção do limite inferior correspondente.

4.3
A formulação com número exponencial de variáveis e restrições

A formulação com número exponencial de variáveis e restrições para o CVRP é obtida através da intersecção entre os dois politopos descritos anteriormente. O politopo P_3 correspondente à intersecção dos politopos P_1 e P_2, no formato de Mestre Explícito, é dado pelas inequações (4-20), (4-21), (4-22), (4-23), (4-26), (4-27), (4-29) e (4-30).

A restrição (4-27) pode ser descartada, pois ela é implicada por (4-21) e (4-26). Calcular o limite inferior, que será chamado de L_3,

$$\min \left\{ \sum_{e \in E} \ell_e x_e : x \in P_3 \right\} \tag{4-32}$$

requer resolver um programa linear (LP) com número exponencial tanto de restrições quanto de variáveis. Um LP mais compacto é obtido se cada ocorrência de x_e em (4-20)-(4-23) por sua forma equivalente dada por (4-26). O LP resultante será referenciado como o problema Dantzig-Wolfe Mestre (DWM), que corresponde a calcular L_3 utilizando

$$\min \left\{ \sum_{j=1}^{p} \sum_{e \in E} \ell_e \cdot q_j^e \cdot \lambda_j : \lambda_j \in P_4 \right\} \tag{4-33}$$

onde o politopo P_4 referenciado em (4-33) é definido por (4-30) mais as inequações abaixo:

$$\sum_{j=1}^{p} \sum_{e \in \delta(\{i\})} q_j^e \cdot \lambda_j = 2, \quad \forall\, i \in V_+ \tag{4-34}$$

$$\sum_{j=1}^{p} \sum_{e \in \delta(\{0\})} q_j^e \cdot \lambda_j = 2 \cdot K, \tag{4-35}$$

$$\sum_{j=1}^{p} \sum_{e \in \delta(S)} q_j^e \cdot \lambda_j \geq 2 \lceil d(S)/C \rceil, \quad \forall\, S \subseteq V_+, \tag{4-36}$$

$$\sum_{j=1}^{p} q_j^e \cdot \lambda_j \leq 1, \quad \forall\, e \in E \setminus \delta(\{0\}). \tag{4-37}$$

As desigualdades de capacidade (capacity cuts) não são as únicas que podem aparecer em (DWM). Um corte genérico $\sum_{e \in E} a_e x_e \geq b$ pode ser incluído como $\sum_{j=1}^{p} (\sum_{e \in E} a_e q_j^e) \cdot \lambda_j \geq b$.

Utilizando P_3 ou P_4 consegue-se construir um algoritmo de BCP robusto para o CVRP. Os detalhes para a construção e implementação do mesmo são descritos no próximo capítulo.