Parte II

Aplicações em Roteamento de Veículos

5
Problema de Roteamento de Veículos com Restrição de Capacidade

O problema de roteamento de veículos com restrição de capacidade, mais conhecido pela sua sigla em inglês *CVRP* (*Capacitated Vehicle Routing Problem*), refere-se à determinação de rotas para os veículos de uma frota, para atender as demandas de um conjunto de clientes. Cada rota deve começar e terminar no mesmo ponto; cada cliente deve ser atendido por apenas uma rota e a soma das demandas de uma rota deve respeitar a capacidade do veículo que a percorre. O objetivo é minimizar a soma da distância total de cada uma das rotas. Este problema, proposto por Dantzig e Ramser em 1959 [35], pertence à classe $\mathcal{NP}$-difícil de problemas de otimização combinatória, já que generaliza o problema do caixeiro viajante (*TSP*), que por sua vez também pertence à classe $\mathcal{NP}$-difícil (veja [51]).

Existe uma extensa literatura disponível a respeito dos problemas de roteamento de veículos com restrição de capacidade e que cobre os mais diferentes aspectos dos mesmos. O caso particular do *CVRP* em que apenas um veículo está disponível no depósito e nenhuma restrição operacional é imposta, o chamado *TSP*, é extensivamente descrito em um livro editado por Lawler et al. [74]. Diversas outras variantes do problema foram classificadas por Desrochers, Lenstra e Savelsbergh e podem ser encontradas em [40]. Os métodos exatos disponíveis até 1980 foram coletados em um extenso estudo de Laporte e Nobert [73].

Christofides, Mingozzi e Toth [30] propuseram, em 1981, um algoritmo que usa um limite Lagrangeano, obtido resolvendo-se o problema das q-rotas mínimas. Uma q-rota é uma seqüência que inicia no depósito, percorre um conjunto de clientes com demanda total máxima de D e retorna ao depósito (a demanda é computada na ocupação do veículo a cada visita a um cliente). Como os vértices relativos aos clientes podem ser visitados mais de uma vez por uma q-rota, o conjunto de rotas válidas para o *CVRP* está estritamente contido no conjunto de q-rotas.

Este trabalho de Christofides, Mingozzi e Toth também descreve um limite inferior baseado em árvores geradoras em que o depósito tem grau $K \leq k \leq 2K$ e são utilizadas $2K-k$ arestas de menor custo possível, onde K é o número de veículos. Limites Lagrangeanos, calculados para o caso em que o depósito tem grau $2K$ e são utilizadas $n+K-1$ arestas, foram utilizados por Fisher [45] e Martinhon, Lucena, e Maculan [81], dentre outros. Formas de se melhorar limites Lagrangeanos foram propostas em [45, 82] e [81].

Uma segunda família de algoritmos exatos considera a formulação do *CVRP* como um problema de particionamento de conjuntos, dentre estes estão os algoritmos apresentados por Balinski e Quandt [14], Agarwal, Marthur e Salkin [3] e Hadjiconstantinou et al. [61]. Toth e Vigo apresentam em [103, 104] resultados comparativos entre vários desses algoritmos citados.

As pesquisas mais recentes têm se concentrado na descrição poliedral da envoltória convexa de conjuntos de K rotas viáveis (representadas por conjuntos de arestas) e no desenvolvimento de rotinas de separação [1, 7, 12, 13, 33, 76, 84]. Algoritmos *branch-and-cut* foram descritos por Araque et al. [8], Augerat et al. [13], Blasum and Hochstättler [24], Ralphs et al. [96], Achuthan, Caccetta e Hill [2] e Lysgaard, Letchford, e Eglese [80], os quais são os melhores métodos exatos atualmente disponíveis para o *CVRP*. Contudo, mesmo estes algoritmos podem falhar na obtenção de boas estimativas para o valor da solução ótima nos casos em que o valor de K é alto ($K \geq 7$). Assim, muitas instâncias com menos de 80 clientes, incluindo muitas propostas há mais de 30 anos por Christofides and Eilon [29], não puderam ser resolvidas por estes algoritmos.

Mais recentemente, Fukasawa et al. [50] apresentaram um algoritmo *branch-and-cut-and-price* para o *CVRP* que utiliza a formulação mestre explícita (veja seção 3.3) apresentada na seção 5.1. Essa abordagem combina um algoritmo *branch-and-cut* com q-rotas, as quais são interpretadas como o subproblema de geração de colunas, para a obtenção de melhores estimativas para o valor da solução ótima. Os experimentos computacionais realizados com este algoritmo, com as principais instâncias da literatura, mostraram que o mesmo pode resolver de forma consistente instâncias de até 135 clientes.

5.1 Formulação do *CVRP*

O *CVRP* é definido em um grafo não orientado $G = (V, E)$, onde $V = \{v_0, v_1, \ldots, v_n\}$ e $E = \{(v_i, v_j) | v_i, v_j \in V, i < j\}$. O vértice v_0 representa o depósito e os demais representam os clientes ($V_+ = \{v_1, \ldots, v_n\}$). O depósito é a base de uma frota de m veículos de capacidade idêntica D (o número de veículos pode tanto ser conhecido a priori quanto ser expresso como uma variável de decisão). Custos c_{ij} são associados a cada aresta $(v_i, v_j) \in E$. Demandas d_i são associadas aos vértices v_i e o objetivo é encontrar um conjunto F de m caminhos fechados, com início e término no vértice v_0, cujo custo total seja mínimo. Cada demanda deve ser atendida por apenas um caminho e a demanda total a ser atendida por cada caminho de F não deve exceder a capacidade Q do veículo.

A formulação mais clássica para o *CVRP* [72] representa por x_{ij} o número de vezes que um veículo percorre a aresta $(v_i, v_j) \in E$. Seja $d(S)$ a soma das demandas de todos os vértices em S, onde $S \subseteq V_+$; $\delta(S)$ é o conjunto de arestas incidentes aos vértices do conjunto S que se removidas desconectam o grafo G (*cut-set* definido por S). Considere o seguinte politopo P_1 em $\mathbb{R}^{|E|}$:

$$P_1 = \begin{cases} \sum\limits_{e \in \delta(\{v_i\})} x_e = 2 & \forall\ v_i \in V_+ \quad (1) \\ \sum\limits_{e \in \delta(\{v_0\})} x_e = 2.K & \quad (2) \\ \sum\limits_{e \in \delta(S)} x_e \geq 2.k(S) & \forall\ S \subseteq V_+ \quad (3) \\ x_e \leq 1 & \forall\ e \in E \setminus \delta(\{v_0\}) \quad (4) \\ x_e \geq 0 & \forall\ e \in E\ . \end{cases}$$

As restrições (3) são as chamadas restrições de eliminação de subciclos, as quais especificam, para qualquer subconjunto S de clientes, que $k(S)$ veículos devem entrar e sair de S. Em resumo, tais restrições forçam todo subconjunto de clientes a ser atendido pela quantidade adequada de veículos. O calculo de $k(S)$ é um problema $\mathcal{NP}$-difícil, uma vez que corresponde a resolução de uma instância do *Bin-Packing Problem*. O valor $k(S)$ é o número mínimo de compartimentos (*bins*) de capacidade D necessário para guardar todos os elementos d_i, $\forall\ i \in S$. Contudo, a formulação continua válida se $k(S)$ é trocado por uma estimativa inferior deste valor (por exemplo $\lceil d(S)/D \rceil$).

Os demais blocos de restrições especificam que: (1) – cada cliente (vértice) deve ser atendido por apenas um veículo; (2) – K veículos devem sair e retornar ao depósito; e (4) – cada aresta não adjacente ao depósito está limitada a ser percorrida no máximo uma vez (arestas adjacentes ao depósito podem ser usadas duas vezes quando a rota é composta de apenas um cliente a ser atendido).

O conjunto $\mathcal{F}_1 = \{x \in P_1\}$ define todas as soluções factíveis para o *CVRP*. Como o número de desigualdades do tipo (3) é exponencial, é necessário um algoritmo que gere iterativamente tais desigualdades para calcular a estimativa a seguir:

$$L_1 = \min_{x \in P_1} \sum_{e \in E} c_e . x_e \, . \tag{5-1}$$

Usando-se o conceito de q-rotas, uma formulação alternativa, com um número exponencial de variáveis, pode ser obtida ao se definir variáveis λ correspondentes a q-rotas. Uma q-rota é um caminho que começa no depósito, passa por uma seqüência de clientes (com demanda total máxima de D unidades) e retorna ao depósito. Cada cliente pode aparecer mais de uma vez em uma q-rota e sua demanda é computada para cada ocorrência. Assim, as variáveis λ são associadas ao conjunto de q-rotas que satisfazem a restrição de capacidade dos veículos. Observe que o problema de geração de colunas, usando-se as q-rotas, pode ser resolvido por um algoritmo de complexidade pseudo-polinomial. Idealmente essas variáveis deveriam ser associadas a rotas válidas. Contudo, isto implicaria na resolução de um problema $\mathcal{NP}$-difícil na geração de colunas.

Assim, defina que cada variável λ_j é associada a uma de p possíveis q-rotas. Seja Q uma matriz $m \times p$, onde as colunas são os vetores de incidência dessas p possíveis q-rotas. Seja, ainda, q_j^e o coeficiente associado à aresta e na j-ésima coluna de Q, ou seja, o número de vezes que a aresta e ocorre na l-ésima q-rota. Considere agora o seguinte politopo P_2 em $\mathbb{R}^{|E|}$, definido como a projeção de um politopo em $\mathbb{R}^{p+|E|}$:

$$P_2 = \text{proj}_x \left\{ \begin{array}{rrcll} \sum_{j=1}^{p} q_j^e . \lambda_j \; - & x_e & = & 0 & \forall \; e \in E \qquad (5) \\ \sum_{j=1}^{p} \lambda_j & & = & K & \qquad (6) \\ & \sum_{e \in \delta(\{v_i\})} x_e & = & 2 & \forall \; v_i \in V_+ \qquad (1) \\ & x_e & \geq & 0 & \forall \; e \in E \\ \lambda_j & & \geq & 0 & \forall \; j \in \{1, \ldots, p\} \; . \end{array} \right.$$

As restrições (5) definem as variáveis x em termos das variáveis λ; (6) define o número de veículos a ser utilizado.

Pode-se demonstrar que o conjunto $\mathcal{F}_2 = \{x \in P_2\}$ também define todas as soluções factíveis para o *CVRP*. Como o número de variáveis λ é exponencial, a estimativa L_2, definida a seguir, deve ser calculada usando-se a técnica de geração de colunas ou relaxação lagrangeana:

$$L_2 = \min_{x \in P_2} \sum_{e \in E} c_e . x_e \,. \tag{5-2}$$

Essa descrição de poliedros em termos de dois conjuntos de variáveis, λ e x, associado a geração de colunas ou relaxação Lagrangeana é chamada de *Mestre Explícita* (veja seção 3.3 e [94]). A formulação mestre explícita para o *CVRP*, como apresentada por Fukasawa et al. [50], combina a formulação por geração de colunas com a formulação clássica com variáveis por arestas. Essa formulação utiliza o politopo $P_3 = P_1 \cap P_2$, cuja descrição no formato mestre explícito, é a seguinte:

$$\text{proj}_x \left\{ \begin{array}{rcrcll}
 & & \sum_{e \in \delta(\{v_i\})} x_e & = & 2 & \forall\, v_i \in V_+ \quad (1) \\
 & & \sum_{e \in \delta(\{v_0\})} x_e & = & 2.K & \quad (2) \\
 & & \sum_{e \in \delta(S)} x_e & \geq & 2.k(S) & \forall\ S \subseteq V_+ \quad (3) \\
 & & x_e & \leq & 1 & \forall\ e \in E \setminus \delta(\{v_0\}) \quad (4) \\
\sum_{j=1}^{p} q_j^e . \lambda_j & - & x_e & = & 0 & \forall\ e \in E \quad (5) \\
\sum_{j=1}^{p} \lambda_j & & & = & K & \quad (6) \\
 & & x_e & \geq & 0 & \forall\ e \in E \\
\lambda_j & & & \geq & 0 & \forall\ j \in \{1, \ldots, p\}\ .
\end{array} \right.$$

As restrições (6) podem ser descartadas, já que são definidas por (2) e (5). O cálculo da estimativa L_3, definida a seguir,

$$L_3 = \min_{x \in P_3} \sum_{e \in E} c_e . x_e \,, \tag{5-3}$$

requer a resolução de um problema de programação linear *ME-CVRP* com número exponencial de restrições e de variáveis:

$$
\text{ME-CVRP}\left\{
\begin{array}{llclll}
\text{Min } \sum\limits_{e \in E} c_e . x_e & & & & & (0)\\
\text{sujeito a} & & & & & \\
 & \sum\limits_{e \in \delta(\{v_i\})} x_e & = & 2 & \forall\ v_i \in V \setminus \{0\} & (1)\\
 & \sum\limits_{e \in \delta(\{v_0\})} x_e & \geq & 2.K^* & & (2)\\
 & \sum\limits_{e \in \delta(S)} x_e & \geq & 2.k(S) & \forall\ S \subseteq V \setminus \{v_0\} & (3)\\
 & x_e & \leq & 1 & \forall\ e \in E \setminus \delta(\{v_0\}) & (4)\\
\sum\limits_{j=1}^{p} q_j^e . \lambda_j \quad - & x_e & = & 0 & \forall\ e \in E & (5)\\
 & x_e & \in & \{0,1,2\} & \forall\ e \in E & \\
\lambda_j & & \geq & 0 & \forall\ j \in \{1,\ldots,p\}\ . &
\end{array}
\right.
$$

Uma formulação mais compacta é obtida se toda ocorrência da variável x_e, nas restrições (0)–(4), é substituída pela expressão equivalente dada pela restrição (5). Relaxadas as restrições de integralidade, a formulação resultante, que será chamada de *DWM-CVRP* (*Dantzig-Wolfe Master CVRP)*, é apresentada a seguir:

$$
\text{DWM-CVRP}\left\{
\begin{array}{lclll}
L_3 = \min \sum\limits_{j=1}^{p} \sum\limits_{e \in E} c_e . q_j^e . \lambda_j & & & & \\
\text{sujeito a} & & & & \\
\sum\limits_{j=1}^{p} \sum\limits_{e \in \delta(\{v_i\})} q_j^e . \lambda_j & = & 2 & \forall\ v_i \in V_+ & (1)\\
\sum\limits_{j=1}^{p} \sum\limits_{e \in \delta(\{v_0\})} q_j^e . \lambda_j & = & 2.K & & (2)\\
\sum\limits_{j=1}^{p} \sum\limits_{e \in \delta(S)} q_j^e . \lambda_j & \geq & 2.k(S) & \forall\ S \subseteq V_+ & (3)\\
\sum\limits_{j=1}^{p} q_j^e . \lambda_j & \leq & 1 & \forall\ e \in E \setminus \delta(\{v_0\}) & (4)\\
\lambda_j & \geq & 0 & \forall\ j \in \{1,\ldots,p\}\ . &
\end{array}
\right.
$$

5.2
Geração de Colunas Projetada

A pesquisa desenvolvida nesta tese, com relação ao *CVRP*, consiste em estudar técnicas para suplantar uma dificuldade que vem junto com o uso de um algoritmo *branch-and-cut-and-price*: a necessidade de se aplicar a técnica de geração de colunas, que é computacionalmente cara, não apenas uma vez, mas em cada um dos nós da árvore de enumeração. Seria desejável que a geração de colunas e de cortes pudesse ser aplicada somente no nó raiz, para se obter um bom limitante inferior. A partir deste ponto o algoritmo deveria proceder como num *branch-and-cut* convencional, apenas gerando novos cortes no restante da árvore. Esta abordagem permitiria reduzir muito o tempo gasto em cada nó.

Nesse contexto, o objetivo proposto é substituir as restrições (5) da formulação *ME-CVRP* por um conjunto de desigualdades descritas somente sobre as variáveis x. Em outras palavras, deseja-se obter as desigualdades $\pi^j.x \leq \pi_0^j$, para $j = 1, \ldots, n_\pi$, que descrevem o poliedro definido por $\{x \mid x = \sum q_l.\lambda_l,\ \sum \lambda_l = K,\ \lambda_l \geq 0\}$.

Observe que esta proposta consiste em gerar desigualdades a partir de soluções parciais do problema, no sentido de que as colunas da matriz Q não descrevem uma solução completa do *CVRP*, *ME-CVRP* no caso, mas apenas cada uma de suas possíveis rotas (na verdade, as colunas correspondem às q-rotas, conjuntos de caminhos que incluem as rotas). Esta é uma generalização do caso explorado por Boyd em [26, 27], onde é descrita uma técnica para obter desigualdades a partir de soluções completas de um problema.

Seja $\overline{x}$ uma solução fracionária pertencente a P_1. Pode-se afirmar que ou $\overline{x}$ também pertence a P_2 (e a P_3), ou existem desigualdades $\pi.x \leq \alpha$ separando $\overline{x}$ de P_2. Note-se que $\overline{x} \in P_2$ se e somente se o seguinte sistema linear tem solução viável:

$$\left\{ \begin{array}{rcl} Q.\lambda & = & \overline{x} \\ \mathtt{1}.\lambda & = & K \\ \lambda & \geq & 0\,, \end{array} \right.$$

ou, equivalentemente, se o valor da solução do LP a seguir for 0:

$$\left\{ \begin{array}{llcccll} \min z = & & & \mathtt{1}.a & & & (1) \\ \text{sujeito a} & & & & & & \\ & Q.\lambda & + & I.a & = & \overline{x} & (2) \\ & \mathtt{1}.\lambda & & & = & K & (3) \\ & \lambda & & & \geq & 0 & (4) \\ & & & a & \geq & 0 & (5)\,, \end{array} \right.$$

onde I é uma matriz identidade e a é um vetor coluna com m variáveis artificiais. Note-se que a restrição (3) pode ser relaxada, uma vez que a coluna correspondente a q-rota vazia pertence à matriz Q. Relaxada essa restrição, obtém-se então o seguinte LP:

$$\begin{cases} \min z = & & & 1.a & & & (1) \\ \text{sujeito a} & & & & & & \\ & Q.\lambda & + & I.a & = & \overline{x} & (2) \\ & 1.\lambda & & & \leq & K & (6) \\ & \lambda & & & \geq & 0 & (4) \\ & & & a & \geq & 0 & (5) \end{cases}$$

Sejam π e ν os vetores de multiplicadores duais associados às restrições (2) e (6), respectivamente. O dual desse LP pode ser escrito como

$$D = \begin{cases} \max z = & \pi.\overline{x} & - & \nu.K & & & (7) \\ \text{sujeito a} & & & & & & \\ & \pi.Q & - & \nu.1 & \leq & 0 & (8) \\ & \pi.I & & & \leq & 1 & (9) \\ & & & \nu & \geq & 0 & (10) \end{cases}$$

Teorema 5.1 *Seja (π', ν') uma solução do problema dual D. A desigualdade $\pi'.x \leq \nu'.K$ é válida para todos os pontos em P_2 (logo para todas as soluções do CVRP). Além disso, se $z > 0$ ela é violada por $\overline{x}$.*

Prova. Se x é um ponto qualquer de P_2, então $x = Q.\overline{\lambda}$ para algum $\overline{\lambda} \geq 0$ tal que $1.\overline{\lambda} = K$. Fazendo-se uma combinação linear das desigualdades (8) ponderadas pelo vetor $\overline{\lambda}$, obtém-se $\pi'.Q.\overline{\lambda} \leq \nu'.1.\overline{\lambda}$, ou seja, $\pi'.x \leq \nu'.K$. Se $z > 0$, (7) implica que $\pi'\overline{x} > \nu.K$. □

5.2.1 Experiência Computacional

As experiências computacionais com a geração de colunas projetada até agora se limitaram a uma única instância do *CVRP* com 20 clientes e 4 veículos. Aplicando-se a geração de colunas e cortes de eliminação de subciclos, a solução ótima de valor 472 é encontrada no nó raiz, ou seja, o limite $L_3 = 472$. Aplicando-se apenas os cortes de eliminação de subciclos, obtemos o limite $L_1 = 471$ e uma solução fracionária $\bar{x}$. Gerando-se cortes válidos para P_2, resolvendo-se (7)–(10), obtém-se novamente o limite $L_3 = 472$ em apenas três iterações, ou seja, três cortes foram gerados.