UMA VERSÃO FINITA DO MÉTODO C.F.A.

Carlos Humes Jr.
Departamento de Ciência da Computação
Instituto de Matemática e Estatística
Universidade de São Paulo

SUMÁRIO

O resultado aqui apresentado é o de validade de um algoritmo "projeção - direções viáveis" que se caracteriza por substituir no C.F.A. clássico a solução do problema de designação de fluxos por um único desvio de fluxo, com passo máximo sem preocupar-se com a manutenção de viabilidaue deste fluxo, perante a capacidade fixada naquela iteração.

Para este algoritmo valem as seguintes observações:

(1) tal como o C.F.A.clássico (projeção - projeção) computa um ponto estacionário de Kuhn-Tucker.

(2) a convergência, em termos de número de iterações é finita.

(3) a cada iteração resolve-se uma designação de capacidades e um desvio de fluxo (parte de uma iteração para a solução de designação de fluxos) diminuindo sensivelmente o custo de cada iteração.

Este resultado é baseado na demonstração de que o valor ótimo do problema de désignação de capacidades, em função dos fluxos indicados, é uma função quasicôncova e que, portanto, toda direção de descida local é direção de descida global.

1. INTRODUÇÃO

A formulação tradicional do problema de projeto de redes de computadores considera que é dado o conjunto N de localizações dos nós comutadores de mensagens, a demanda média de transmissão de mensagens e suas características para cada par (origem, destino). De posse destes dados busca-se determinar tanto o rateamento das mensagens (qual o caminho que seguem ou, equivalentemente, quais os fluxos de mensagens nos vários canais), quanto a capacidade de transmissão (velocidade) de cada um dos canais a serem instalados. Nesta formulação, pode-se considerar que o conjunto de possíveis interconexões é variável de projeto (projeto topológico) ou que é dado (o que é a tônica desta apresenta; ou seja, o problema de designação de fluxos e capacidades).

Para avaliar o projeto temos que definir critérios, critérios estes que caracterizaram os pontos focais de abordagem. Evitando maiores discussões que podem ser encontradas em GERLA [1] e HUMES [2], consideraremos o implícito problema bicritério onde as funcões objetivo são custo e tempo médio de encaminhamento de mensagem entre origem e destino. Estes critérios serão modelados por funções $D(c)$ = custo dos m canais instalados com capacidades c_i, $i = 1, 2, \ldots, m$, e $T(f,c)$ = retardo com roteamentos associado a um fluxo de mensagens $f = (f_1 \ldots f_m)^t$ e capacidades instaladas $(c_1, c_2, \ldots, c_m)^t$.

A notação acima, pressupõe que existem m canais possíveis de interconexão, correspondendo ao projeto toplógico. O projeto topológico corresponde a definirmos o esquema de interconexão entre os nós comutadores de mensagens. Esta definição é naturalmente modelada através de um grafo $G = (N, \bar{A})$ onde $\bar{A}$, conjunto dos arcos do grafo G, representa a existência de canais de comunicação entre os nós (comutadores).

Como $\bar{A}$ é finito, não há perda de generalidade em utilizarmos $\bar{A} = \{1,2, \ldots, m\}$ caracterizando os extremos das arestas por

$$\bar{a} : \bar{A} \rightarrow \{\{i, j\} \subset 2^N \mid i > j\} \equiv Q$$

Note-se que a definição acima (Q), implica em não considerarmos canais cuja origem e destino coincidam (hipótese esta bastante natural). Na mesma linha, suporemos $\bar{a}(\cdot)$ injetora (canais em paralelo são tratados como canal de maior capacidade).

Na medida em que considerarmos os canais *full-duplex* e formos estudar fluxos é conveniente considerar o digrafo associado à interconexão. Digrafo este cujas arestas orientadas são dadas por:

$$a : \bar{A} \rightarrow \{(i, j) \in N \times N\}$$

caracterizadas por

$$a(i) = \{(k, l), (l, k)\} \iff \bar{a}(i) = \{k, l\}$$

O conjunto de arestas do digrafo será indicado por A, onde

$$\bar{A} = U_{i \in \bar{A}}\, a(i) = a(\bar{A}) .$$

Considerando como dados:

$N = \{1, 2, \ldots, n\}$ = conjunto de localizações de nós comutadores de mensagens, isto é, pontos onde um ou mais dos canais de comunicação tem extremos,

$q : N \times N \to R_+$ = demanda média de transmissão entre os nós comutadores de mensagem. Tipicamente esta demanda é medida em kilobits/seg.*,

$\bar{q} : N \times N \to R_+$ = demanda média de envio de mensagens entre os nós comutadores. Tipicamente, esta demanda é medida em milhares de mensagens/seg.*,

$\bar{l} : N \times N \to R_{++}$ = comprimento médio das mensagens entre os nós comutadores*,

---

* Por simplicidade de notação usaremos

$q_{ij} = q((i, j))$, $q = \Sigma_{i \in N} \Sigma_{j \in N} q_{ij}$,

$\bar{q}_{ij} = \bar{q}((i, j))$, $\bar{q} = \Sigma_{i \in N} \Sigma_{j \in N} \bar{q}_{ij}$,

$\bar{l}_{ij} = \bar{l}((i, j))$.

Mais ainda, faremos as seguintes hipóteses simplificadoras:

$\bar{l}_{ij} \doteq \frac{1}{\mu}$ (homogeneidade no tamanho de mensagens);

$q_{ij} = \frac{1}{\mu} \bar{q}_{ij}$ (independência de processos);

$q_{jj} = 0$, $\bar{q}_{jj} = 0$, para $j \in N$.

podemos então definir o problema de projeto de redes de computadores como:

"Dados N e q : $N \times N \to R_+$, encontre, se existir, a região de pontos eficientes (ou um subconjunto desta região) em relação ao critério $(T(f, c), d(c))^t$, obedecendo às restrições

$$(R1)\ f^{rs} \in F^{rs} = \{\bar{f} \in R_+^{2m} \mid \Sigma_{l:(k,l)\in A}\, \bar{f}_{kl} - \Sigma_{l:(l,k)\in A}\, \bar{f}_{lk} =$$

$$= (\delta_{kr} - \delta_{ks})\, q_{rs}\}$$

para $(r, s) \in N \times N$ (com $q_{rs} > 0$*;

$$(R2)\ f_{ij} = \Sigma_{(r,s)\in N\times N}\, f_{ij}^{rs}, \text{ para } (i, j) \in A;$$

$$(R3)\ f_i = f_{kl} + f_{lk}, \text{ para } \hat{a}(i) = \{k, l\};$$

$$(R4)\ \forall i \in \hat{A},\ (f_i, c_i) \in Y_i = \{(0, 0)\} \cup \{(x, y) \in R \times C_i \mid 0 \le x < y\},$$

onde C_i é dado;

(R5) $\hat{A}$ é um conjunto de arcos com extremos em N, gozando, conforme a formulação, de uma propriedade P".

As restrições (R1) e (R2) correspondem a modelar um fluxo multicomodidade (*multi-commodity flow*) onde cada tipo de comodidade (unidade de transmissão) é caracterizada por sua origem e destino.

---

* Caso contrário, $f_{ij}^{rs} = 0$, para todos os arcos $(i, j) \in A$.

A restrição (R3) corresponde a uma modelagem clássica para canais *full-duplex*, onde o sentido em que a unidade de transmissão percorre o canal não é determinante de seu efeito no fluxo.

Em alguns pontos, será conveniente notar que estamos trabalhando com vetores $\bar{f} \in R_+^{2m}$, tais que $\bar{f} \in \Sigma_{(r,s)\in N\times N} F^{rs}$ e com vetores $f \in R^m$, onde $f_i = \bar{f}_{kl} + \bar{f}_{lk}$ para $a(i) = \{k, l\}$. Porém, nos reservamos o direito de, a menos de necessidade para a clareza do texto, usar a expressão "$f = (f_1\ f_2\ \ldots\ f_m)^t$ é a soma de fluxos multicomodidade".

Note-se que nas condições acima não há perda de generalidade em assumir que:

$$r \le s \Rightarrow q_{rs} = 0 \Leftrightarrow F^{rs} = \{0 \in R^{2m}\}$$

Mais ainda, deve ser claro que em qualquer solução eficiente

$$\forall\ (k, l) \in A\ ,\ \forall.(r, s) \in N \times N\ ,\ f_{kl}^{rs} \cdot f_{lk}^{rs} = 0$$

A restrição (R4) indica que o canal inativo ($f_i = 0$) pode não existir fisicamente ($c_i = 0$) e que, exceto neste caso, o grau de congestionamento (f_i/c_i) é bem definido e pertence ao intervalo $[0,1)$.

A questão de canais inativos poderem não existir fisicamente é complexa ao considerarmos que na restrição (R5) poderiamos impor condições tais como biconexidade sobre o grafo $(N,\bar{A})$. Uma possível solução para tal questão seria impor restrições do tipo $f_i \geq \bar{f}_i$ ou $c_i \geq \bar{c}_i$ para $i \in \bar{A}$. O segundo tipo de restrições pode fazer sentido no caso $C_i = \{0, c_{1i}, c_{2i}, \ldots, c_{pi}\}$.

Para evitar maiores problemas, consideraremos $\bar{A}$ como dado e seguindo a abordagem usual para problemas bicritério (ver, p. ex., LIN [3] e [4]), consideremos os problemas abaixo:

$$\text{PD }(T_{MAX}) \quad \begin{aligned} &\text{minimizar } \Sigma_{i=1}^{m} d_i\,(c_i) \\ &\Sigma_{i=1}\, t_i\,(f_i, c_i) \leq \bar{q}\, T_{MAX} \\ &(f_i, c_i) \in Y_i \\ &f \in F; \end{aligned}$$

$$\text{PT }(D_{MAX}) \quad \begin{aligned} &\text{minimizar } \Sigma_{i=1}^{m} t_i\,(f_i, c_i) \\ &\Sigma_{i=1}\, d_i\,(c_i) \leq D_{MAX} \\ &(f_i, c_i) \in Y_i \\ &f \in F; \end{aligned}$$

onde F é o poliedro das somas de fluxos multicomodidades descritas pelas restrições (R1) a (R3) de e estamos considerando $\bar{A} = \{1, 2, \ldots, m\}$ dado. Claramente supomos $F \neq \emptyset$. Estes problemas são denominados **problemas de designação de fluxo e capacidade** (CFA - Capacity and Flow Assignement).

Neste trabalho nos concentraremos no caso $C_i = [0, +\infty)$, chamado de **relaxação contínua** ou **caso contínuo**.

Por simplicidade, nós consideramos acima que o custo da rede é somente o custo dos canais de transmissão, ou mais preci samente, a soma dos custos individuais dos canais. Utilizando a notação $D(\cdot)$ para o custo total e $d_i(\cdot)$ para o custo dos canais individuais, é claro que

$$D(\cdot) = \Sigma_{i=1}^{m} d_i(\cdot) .$$

Os custos individuais dos canais são definidos nos valores possíveis de capacidades ($d_i : C_i \to R_+$), mas, não há perda em generalidade em supor:

$$\forall\ i \in \bar{A},\ d_i : R_+ \to R_+ .$$

Em termos simplistas, a hipótese acima conflita com, por exemplo, $- C_i = p \in N$. Neste caso, existiria uma infinidade de funções custo $d_i : R_+ \to R_+$ que retratariam a estrutura de custos restrita às capacidades disponíveis.

Mas, é natural supor que as funções $d_i(\cdot)$ procurem representar fenômenos econômicos e que desta forma sejam gerados seus valores. Assim sendo, suporemos ao longo deste trabalho:

(H1) $d_i(0) = 0$ (capacidade nula <=> canal não instalado <=> custo nulo);

(H2) $\lim_{x \to \infty} d_i(x) = +\infty$ (não há limite superior para os custos a menos que as capacidade sejam limitadas);

(H3) $x > y \geq 0 \Rightarrow d_i(x) > d_i(y) \geq 0$ (custos monotonicamente crescentes);

(H4) $d_i(\cdot)$ é C^∞ em $(0, +\infty)$ (com continuidade simples na fronteira);

Além destas hipóteses usaremos freqüentemente:

(H5) $d_i(\cdot)$ é côncava ou (H5A) $d_i(\cdot)$ é estritamente côncava;

(H6) $d_i(\cdot)$ é côncava na origem ou (H6A) $d_i(\cdot)$ é estritamente côncava na origem.

A não obrigatoriedade das hipóteses de concavidade está associada ao nosso interesse em obter resultados teóricos sem estas hipóteses. A hipótese que corresponde a uma tecnicalidade é a hipótese (H4). Defendemos fortemente a validade desta hipótese, no caso geral, pois acreditamos que os modelos de geração de preços pelos serviços de comunicação utilizem funções para as quais (H4) vale.

No caso brasileiro, temos

$$d_i(x) = l_i \, c^{\alpha},$$

onde

(i) $\alpha \in (0, 1)$, i.é., $\alpha \simeq 0{,}56$;

(ii) l_i depende da distância em quilômetros entre os extremos do canal de comunicação.

É interessante notar que este resultado foi obtido inicialmente por análise de tabelas da EMBRATEL e confirmada posteriormente em contatos com profissionais desta organização.

Em relação ao retardo, consideramos válida a hipótese de existência de uma forma produto para as probabilidades de mensagens nos *buffers* do nós chamadores de mensagens, ou seja, (p. ex., ver KLEINROCK [5]):

$$T\ (f,\ c) = \frac{1}{\bar{q}} \sum_{i=1}^{m} t_i\ (f_i,\ c_i) = T$$

No caso particular de retardo médio em redes de filas M/M/1 teremos

$$T = \sum_{i=1}^{m} \frac{\lambda_i}{\bar{q}}\ T_i = \sum_{i=1}^{m} \frac{1}{\bar{q}}\ \frac{\lambda_i}{\mu}\ \frac{1}{c_i - f_i} = \frac{1}{\bar{q}} \sum_{i=1}^{m} \frac{f_i}{c_i - f_i}$$

Isto é, para retardo médio em redes de filas M/M/1, teremos

$$t_i\ (f_i,\ c_i) = \frac{f_i}{c_i - f_i}$$

As funções t_i $(\cdot,\cdot)$ não apresentam um caráter qualquer por representarem parcelas do fenômeno físico de retardo. Assim sendo, ao longo do trabalho suporemos que alumas hipóteses são a elas aplicáveis. Estas hipóteses, para $i = 1, 2, \ldots, m$, são:

(H7)* $t_i : Y \rightarrow R_+$, onde $Y = \{(0,\ 0)\} \cup Y_0$ e $Y_0 = \{(f,\ c)\ |\ 0 \leq f < c\}$, $t_i\ (0,\cdot) = 0$;

(H8) $t_i : Y_0 \rightarrow R_+$ é C^∞ (com os devidos cuidados na fronteira);

(H9) $\forall\ M > 0$, $t_i\ (f,\ c) = t_i\ (M\ f,\ M\ c)$;

---

* As hipóteses H1 a H6 dizem respeito à estrutura de custos.

(H10) $\forall\, \bar{c} > 0,\ t_i\ (\cdot,\ \bar{c}) : Y \cap \{(a,\ \bar{c} \mid a \in R\} \to R_+$ é

(a) estritamente convexa

(b) monotonicamente crescente (no sentido estrito)

(c) $\lim_{f \to \bar{c}-} t_i\ (f,\ \bar{c}) = +\infty.$

(H11) $\forall\, \bar{f} > 0,\ t_i\ (\bar{f},\ \cdot) : Y \cap \{(\bar{f},\ a) \mid a \in R\} \to R_{++}$ é

(a) estritamente convexa

(b) monotonicamente decrescente (no sentido estrito)

(c) $\lim_{c \to \bar{f}_+} t_i\ (\bar{f},\ c) = +\infty.$

A hipótese (H7) especifica que o grau de congestionamento deve ser menor do que a unidade e que canais não ativos ($f_i = 0$) não contribuem para nossa medida de retardo (mais ainda, neste caso, a melhor opção é a não instalação do canal, isto é, $c_i = 0$).

A hipótese (H8) é meramente técnica e refletida nos modelos de filas usuais. A hipótese (H9) de homogeneidade de grau zero está intrinsicamente ligada à idéia de que os $t_i\ (\cdot,\cdot)$ são adimensionais retratando a parcela do canal i na composição de $\bar{q}$ T e, portanto, não podem ser influenciados por "mudanças de escala" na aferição de c_i e f_i.

As hipóteses (H10) e (H11) representam o comportamento natural de sistemas com congestão.

Uma observação é que deve ser notado que a homogeneidade de grau zero implica em descontinuidade em (0, 0). Existem imprecisões na literatura devido à ausência deste cuidado.

2. Resultados Preliminares

Para os problemas acima definidos, valem os seguintes resultados, para a relaxação contínua.

<u>Fato 2.1</u> PD (T_{MAX}) é viável qualquer que seja T_{MAX} 0 e PT (D_{MAX}) é viável qualquer que seja $D > D^0$ onde $D^0 = \inf \{ \sum_{i=1}^{m} d_i\ (f_i) \mid f \in F\} > 0$.

<u>Demonstração</u>: trivial a partir das hipóteses (H1) a (H4) e (H7), (H8), (H10) e (H11), notando que $D^0 = \min \{ \sum_{i=1}^{m} d_i\ (f_i) \mid f \in F\}$.

O mínimo acima indicado é bem definido, pois F é um poliedro na forma canônica $F \equiv \{[V\ (F)] + C_F\}$ e como $d_i(\cdot)$ é monotonica estritamente crescente,

$$D^0 = \inf \left\{ \sum_{i=1}^{m} d_i\ (f_i) \mid f \in F \right\} =$$

$$= \inf \left\{ \sum_{i=1}^{m} d_i\ (f_i) \mid f \in [V\ (F)] \right\} =$$

$$= \min \left\{ \sum_{i=1}^{m} d_i\ (f_i) \mid f \in [V\ (F)] \right\} \quad \square$$

Para o caso contínuo podemos ainda afirmar:

<u>Fato 2.2</u> Para todo $T_{MAX} > 0$ e $D_{MAX} > D_0$, os problemas

PT (D_{MAX}) e PD (T_{MAX}) têm solução ótima.

Demonstração

Apresentaremos a prova em duas partes distintas:

(i) $T_{MAX} > 0 \Rightarrow$ PD (T_{MAX}) tem solução ótima.

Antes de mais nada notemos que a restrição $0 < f_i < c_i$ pode ser substituída por $f_i \leq \bar{\rho}_i\, c_i$ onde $\bar{\rho}_i$ é definido por $t_i\,(\bar{\rho}_i, 1) \leq T_{MAX}$. Claramente, $\bar{\rho}_i \in [0, 1)$.

Tal fato é possível pois para $c_i > 0$, por (H9)

$$t_i\,(f_i, c_i) = t_i\,(f_i/c_i, 1).$$

Mais ainda, por (H10), $t_i\,(\cdot, 1)$ é estritamente crescente e por (H7), $t_k\,(\cdot,\cdot) \geq 0$.

Portanto, a restrição $(f_i, c_i) \in Y_i$ pode ser substituída por

$$(f_i, c_i) \in \{(a, b) \in R^2_+ \mid a \leq \bar{\rho}_i\, b\}\ .$$

Mais ainda, se $(\bar{f}, \bar{c})$ é um ponto viável de PD (T_{MAX}) podemos, sem perda de generalidade, impor a hipótese adicional $c_i \leq c_i^+$ onde c_i^+ é definido por $d_i\,(c_i^+) = \sum_{j=1}^{m} d_j\,(\bar{c}_j)$.

A existência de c_i^+ é garantida por (H1) a (H4).

Assim sendo, PD (T_{MAX}) pode ser reescrito como

$$\text{minimizar} \sum_{i=1}^{m} d_i (c_i)$$

$$(f_i, c_i) \in K_i = \{(a, b) \mid 0 \le a \le b\, \bar{\mu}_i \le c_i^+$$

$$\sum_{i=1}^{m} t_i (f_i, c_i) \le \bar{q}\, T_{MAX}$$

$$f \in F$$

Claramente K_i é compacto e portanto o conjunto de pontos viáveis está contido no compacto $K \equiv K_1 \times K_2 \times \ldots \times K_m$.

Mais ainda, por (H8) $t_i (\cdot,\cdot)$ é contínua em todos os pontos de K_i, exceto (0, 0). Mas, neste ponto, qualquer que seja a seqüência $\{(f_i^k, c_i^k)\}_{k \in N}$, $(f_i^k, c_i^k) \in K_i$, convergindo para (0, 0), temos

$$t_i (f_i^k, c_i^k) \ge 0 = t_i (0, 0)$$

portanto, $t_i (\cdot,\cdot)$ é semicontínua inferior em K_i.

Podemos então, afirmar que $\sum_{i=1}^{m} t_i (\cdot,\cdot)$ é semicontínua inferior em K e, portanto, $\{(f, c) \in K \mid \sum_{i=1}^{m} t_i (f_i, c_i) \le \bar{q}\, T_{MAX}\}$ é fechado em K e, portanto, é compacto.

Como F é fechado e $\sum_{i=1}^{m} d_i(\cdot)$ é contínua (H4), segue que estamos minimizando uma função contínua em um compacto não vazio (o ponto $(\bar{f}, \bar{c})$ viável, utilizado via axioma da escolha pertence ao compacto construído). Portanto, existe solução ótima do problema. $\square$

(ii) D_{MAX} tem solução ótima.

A argumentação é similar à apresentada acima, sendo que a restrição de custo impõe o limite superior nas capaicades, um ponto viável gera um retardo utilizado para definir os $\bar{\nu}_i$, i = = 1 ... m, e a compacidade dos pontos viáveis segue da continuidade das restrições de custo e de fluxo (lineares).

Como o retardo é limitado inferiormente (t_i $(\cdot,\cdot) \geq 0$) e semicontínuo inferior, com a compacidade obtida segue a existência de mínimo. □

A importância deste resultado fica clara ao enunciarmos o seguinte fato, para o caso contínuo:

<u>Fato 2.3</u> Para todo $T > 0$, o ponto $(T, D^*(T))$, onde

$D^*(T)$ = valor ótimo de PD (T)

é um projeto eficiente.

Antes de demonstrar este fato, provemos o seguinte lema válido para o caso contínuo:

<u>Lema 2.4</u> Para todo $T > 0$, se (f^*, c^*) é solução ótima de PD (T) então $T(f^*, c^*) = T$.

<u>Demonstração</u>: suponhamos, por contradição, que $T(f^*, c^*) < T$. Então podemos diminuir c_1^* para $\bar{c}_1 = c_1^* - \Delta_1$, de modo que

$$t_1(\bar{c}_1, f_1^*) = t_1(c_1^*, f^1) + T - T(f^*, c^*) .$$

Isto é possível por (H11) e (H8).

Então por (H3) obtemos um ponto viável com custo inferior ao da solução ótima. O que é uma contradição. □

Com este lema em mente, segue a demonstração de

Demonstração de (2.3): suponhamos, por contradição, que $(T, D^*(T))$ não é eficiente, isto é, é possível encontrar um vetor $(\bar{f}, \bar{c})^t$ tal que $(F(\bar{f}, \bar{c}), D(\bar{c}) \leq (T, D^*(T))$.

Se $T(\bar{f}, \bar{c}) < T$ então $(\bar{f}, \bar{c})$ é viável em PD (T) e $D(\bar{c}) = D^*(T)$. Portanto $(\bar{f}, \bar{c})$ solve PD (T) e por $T(\bar{f}, \bar{c}) = T$, o que é uma contradição.

Se $T(\bar{f}, \bar{c}) = T$, então $D(\bar{c}) < D^*(T)$, o que negaria a definição de $D^*(T)$. □

De forma completamente análoga provam-se os seguintes fatos válidos para o caso contínuo.

Fato 2.5 Para todo $D > D_0$, o ponto $(T^*(D), D)$, onde

$$T^*(D) = \text{valor ótimo de PT } (D)$$

é um projeto eficiente.

Lema 2.6 Para todo $D > D_0$, se (f^*, c^*) é solução ótima de PT (D), então $D(c^*) = \sum_{i=1}^{m} d_i(c_i^*) = D$.

As demonstra ções são omitidas por serem variações simples das provas anteriores.

Estes resultados iniciais correspondem a um aproveitamento parcial de peculiar estrutura matemática do problema em estudo. Sob o ponto de vista de Programação Matemática de Grande

Porte, devemos ainda destacar as estruturas de separabilidade, monotonicidade e conversidades parciais.

As funções custo e retardo são somas de parcelas atribuíveis a cada canal específico. As funções custo por canal são monotonicas crescentes e, usualmente, côncavas. Os retardos por canal são funções que com um argumento fixo são estritamente convexas e extritamente monotônicas. Mais ainda, se pudéssemos desprezar a restrição de retardo, teríamos dois grupos de variáveis com completa independência. Neste caso, a aproximação contínua teria valor ótimo obviamente igual a ($\sum_{i=1}^{m} d_i (f_i)$) caso o problema fosse viável. Em relação ao retardo, existe a estrutura adicional de que t_i $(\cdot,\cdot)$ é função homogênea de grau zero.

Este conjunto de consideracões despertaram o nosso interesse e, em particular, induziram-nos a aceitar a abordagem usual da literatura, projeção. Note-se que a projeção natural corresponde a trabalhar alternadamente nos subespaços de fluxos e de capacidades, resolvendo alternadamente problemas projetados que são denominados problema de designação de capacidades (CA) (fluxo fixo) e problema de designação de fluxo (FA) (capacidades fixas). Estes problemas são definidos como:

O problema de designação de capacidades é aquele obtido através da particularização do problema de designação de fluxos e capacidades ao prefixarmos um fluxo multicomodidade. Isto é, (CA) é obtido a partir de PD (T_{MAX}) ou PT (D_{MAX}), impondo $F \equiv \{\bar{f}\}$.

Nestas condições, obtemos os problemas CD (T_{MAX}) e

CT (D_{MAX}) definidos por

$$\min \sum_{i=1}^{m} d_i\ (c_i)$$

$$CD\ (T_{MAX}) = \sum_{i=1}^{m} t_i\ (\bar{f}_i,\ c_i) \le \bar{q}\ T_{MAX}$$

$$c_i \in C_i \text{ e } c_i > \bar{f}_i \quad i = 1, 2, \ldots, m$$

$$\min \sum_{i=1}^{m} t_i\ (\bar{f}_i,\ c_i)$$

$$CT\ (D_{MAX}) = \sum_{i=1}^{m} d_i\ (c_i) \le D_{MAX}$$

$$c_i \in C_i \text{ e } c_i > \bar{f}_i \quad i = 1, 2, \ldots, m$$

Note-se que estamos assumindo, sem perda de generalidade, $\bar{f} > 0$.

O problema de designação de fluxos (FA) é aquele obrido através da particularização do problema de designação de fluxos e capacidades ao prefixarmos as capacidades instaladas. Isto é, (FA) é obtido a partir de PT (D) impondo $C_i = \{\bar{c}_i\}$, $i = 1, 2, \ldots, m$ e $D\ (c) = \sum_{i=1}^{m} d_i\ (\bar{c}_i)$.

Assim sendo, o problema (PA), para $c = \bar{c}$, é definido por

$$\min \sum_{i=1}^{m} t_i\ (f_i,\ \bar{c}_i)$$

$$\text{(FA)} \quad \begin{cases} f \in F \\ c_i > f_i, \text{ para } i \in I\ (\bar{c}) \\ f_i = 0, \text{ para } \bar{c}_i = 0. \end{cases}$$

onde F é o poliedro de fluxos multicomodidade descrito pelas restrições (R1) a (R3) de

Recordando o já exposto:

$$(R1)\ f^{rs} \in F^{rs} = \{\bar{f} \in R_+^{2m} \mid \Sigma_{l:(k,l)\in A}\ \bar{f}_{kl} -$$

$$- \Sigma_{l:(l,k)\in A}\ \bar{f}_{lk} = (\delta_{kr} - \delta_{ks})\ q_{rs}\}$$

$$(R2)\ f_{ij} = \Sigma_{(r,s)\in N\times N}\ f_{ij}^{rs}, \quad \text{para} \quad (i,\ j) \in A,$$

$$(R3)\ f_i = f_{kl} + f_{lk}, \quad \text{para} \quad \bar{a}(i) = \{k,\ l\}.$$

Note-se que não há perda de generalidade em assumir $I\ (\bar{c}) = \{1,\ 2,\ \ldots,\ m\}$.

O problema de designação de fluxos é um problema não-linear convexo de fluxos multicomodidade. Apesar de dificuldades naturais associadas a porte, este problema é bem comportado no sentido de que para ele teoremas fortes de dualidade (tanto minimax como Wolfe) são válidos. Mais ainda, todo mínimo local é global.

O problema de designação de capacidades tem duas versões: contínua e discreta. No caso discreto, é um problema de mochila não-linear. No caso contínuo, o (CA) aparente ter a dificuldade de apresentar mínimos locais que não são globais (devido à concavidade de $d_i(\cdot)$), exceto no caso simples de custo lineares. Tais dificuldades são aparentes e em HUMES [2] elas são eliminadas com a introdução de uma hipótese adicional. Dizemos que vale a hipótese adicional para a estrutura de custos e retardos (HA) para o problema de projeto, se

$$\forall\, i = 1, 2, \ldots, m, \ \forall\, \bar{f}_i > 0, \ \forall\, \bar{c}_i > \bar{f}_i;$$

$$\frac{d^2}{dc_i^2} d_i\,(c_i) \quad \frac{d}{dc_i} t_i\,(\bar{f}_i, c_i) -$$

$$- \frac{d^2}{dc_i^2} t_i\,(\bar{f}_i, c_i) \quad \frac{d}{dc_i} d_i\,(c_i) \quad \Big|_{c_i = \bar{c}_i} < 0$$

ou, mais sumariamente,

$$d''\,(\bar{c}_i)\; t_i'\,(\bar{c}_i) - t''\,(\bar{c}_i)\; d'\,(\bar{c}_i) < 0 \qquad \square$$

Note-se que se o custo for linear, a convexidade estrita do retardo e a monotonicidade crescente do custo implica na validade da hipótese adicional.

É interessante notar que para o caso mais estudado na literatura, que é o de filas M/M/1 com custo dado por *power law*, a hipótese adicional (HA) é válida.

Mas, um ponto que devemos ter em mente é que um ponto (f^*, c^*) tal que f^* resolve o (FA) com $C_i = \{c_i^*\}$ e c^* resolve o (CA) com $F = \{f^*\}$ não é necessariamente sequer um mínimo local, como é afirmado erroneamente na literatura.

3. DESIGNAÇÃO DE FLUXOS E CAPACIDADES USANDO PROJEÇÃO

No seu clássico artigo de 1970, GEOFFRION [6] introduziu as idéias manipulação e estratégias, para o tratamento de problemas de grande porte em programação matemática. Uma das manipulações destacadas é a chamada projeção, com a qual já tratamos implicitamente ao estudar a utilização de programação dinâmica no (CA) discreto (ver, por exemplo, HUMES [4]).

Tipicamente, dado o problema (PTP)

$$\text{(PTP)} \quad \begin{array}{l} \min f(x, y) \\ \text{sujeito a } g(x, y) \leq 0 \\ x \in X,\ y \in Y \end{array}$$

a manipulação de projeção deste problema, sobre o espaço das variáveis y, como:

$$(PAP_y) = \begin{array}{l} \min v(y) \\ \text{sujeito a } y \in Y \cap V, \end{array}$$

onde:

$$V = \{y \in Y \mid \exists\, x - X : g(x, y) \leq 0\},$$

$v : V \to R\ \{-\infty\}$ é definida por

$$v(\bar{y}) = \inf \{f(x, \bar{y}) \mid \in X \text{ e } g(x, \bar{y}) \leq 0\}.$$

A técnica de projeção é naturalmente aplicável ao problema de designação de fluxos e capacidades, desde que seja tomado o cuidado técnico do tratamento da restrição $c > f$.

Basicamente a idéia é que para $c = \bar{c}$ (ou $f = \bar{f}$) para o qual o problema seja viável, isto é, $\bar{c} \in V$ ($\bar{f} \in V$), existe um limitante superior do retardo e pode-se substituir $c > f$ por $\bar{c} \geq kf$ (ou $c \geq k\bar{f}$).

Os principais resultados necessários a uma fácil aplicação do método de projeção estão associados aos seguintes fatos:

Fato 3.1 Os problemas PD (T) e PT (D), se viáveis, têm solução ótima tanto no caso $C_i = [0, +\infty)$ como no caso $C_i = 0, c_{i1}, \ldots, c_{ip}$, $i = 1, 2, \ldots, m$.

Demonstração: no caso contínuo é repetição de 2.2. No caso discreto, é óbvio pela finitude de capacidades disponíveis. □

Fato 3.2 Se o ponto (f^*, c^*) é solução ótima de PD (T) (PT (D)), então

(a) c^* é solução ótima de CD (T) (CT (D)) com $f = f^*$;

(b) f^* é solução ótima de (FA), com $c = c^*$.

Demonstração: óbvia. □

Os fatos acima indicam a existência de solução ótima para PD (T) e PT (D) e que, caso usemos projeção sobre as capacidades ou sobre os fluxos, o ínfimo presente na definição de $V(\cdot)$ é um mínimo.

A abordagem tradicional para a solução (ou tentativa de solução) do problema de designação de capacidades e fluxos é o método (CFA) que a partir de um fluxo viável f^0 gera a seqüência $\{(c^i, f^i)\}_{i \in N}$, definida por

$$c^i = \text{solução ótima do (CA), com } f = f^{i-1};$$

$$f^i = \text{solução ótima do (FA), com } c = c^i.$$

Este método pode ser visto como um de **dupla projeção** e é o encontrado na literatura, como p. ex., KLEI;RPCK [5] e GERLA [1].

A consideração pragmática de que o custo da solução do (CA) é de ordem de grandeza inferior ao da solução do (FA) nos leva a sugerir que utilizemos um método acoplando a estratégia de direcões viáveis com a manipulação de projeções sobre o espaço das variáveis fluxo. Em termos imprecisos, a partir de um fluxo viável f^0, geraríamos uma seqüência $\{(c^i, f^i)\}_{i \in N}$, tal que

$$c^i = \text{solução ótima do (CA), com } f = f^{i-1} \in F$$

$$f^i = (f^i + \lambda_i h^i) \in F,$$

com a propriedade de melhora do critério para cada par (f^i, c^i). Isto é, para PD (T), D (c^{i+1}) < D (c^i) e para PT (D), T $(f^{i+1}$, ci + 1) < T (f^i, c^i).

Proporemos a seguir o método desta família e analisaremos seu comportamento para o caso contínuo. Por facilidade de expressão usaremos só a expressão projeção para significar projeção sobre o espaço das variáveis fluxo.

Para podermos estudar esta abordagem temos que analisar a existência de soluções ótimas e o comportamento do valor ótimo do problema de designação de capacidades, para vários valores de fluxo multicomodidade f.

Para tal indiquemos os seguintes fatos:

Fato 3.3 Os problemas CD (T_{MAX}) e CT (D_{MAX}) são viáveis para

$$T_{MAX} > T_0 = \inf \left\{ \sum_{i=1}^{m} t_i (f_i, c_i) \mid c_i > \overline{f}_i \text{ e } c_i \in C_i, \ i = 1, 2, \ldots, m \right\}$$

$$D_{MAX} > D = \inf \left\{ \sum_{i=1}^{m} d_i (c_i) \mid c_i > \overline{f}_i \text{ e } c_i \in C_i, \ i = 1, 2, \ldots, m \right\}$$

sendo que para o caso discreto $T_{MAX} = T_0$ e $D_{MAX} = D_0$ também são condições de viabilidade. Para o caso contínuo ($C_i \equiv R_+$), temos $T_0 = 0$ e $D_0 = \sum_{i=1}^{m} d_i (f_i)$.

Demonstração: omitida por ser trivial no caso discreto, numa particularização de (2.1) no caso contínuo e uma simples constatacão no caso geral. □

Fato 3.4 Se o problema CD (T_{MAX}) (CT (D_{MAX})) é viável, então possui solução ótima, tanto no caso discreto como no contínuo.

Demonstração: omitida por ser trivial no caso discreto e uma simples particularização de (2.2) no caso contínuo. □

Na realidade, o resultado (3.4) é válido para o caso ge-método (CFA) que a partir de um fluxo viável f^0 gera a seqüência $\{(c^i, f^i)\}_{i \in N}$, definida por

$$c^i = \text{solução ótima do (CA), com } f = f^{i-1};$$

$$f^i = \text{solução ótima do (FA), com } c = c^i.$$

Este método pode ser visto como um de **dupla projeção** e é o encontrado na literatura, como p. ex., KLEI;RPCK [5] e GERLA [1].

A consideração pragmática de que o custo da solução do (CA) é de ordem de grandeza inferior ao da solução do (FA) nos leva a sugerir que utilizemos um método acoplando a estratégia de direções viáveis com a manipulação de projeções sobre o espaço das variáveis fluxo. Em termos imprecisos, a partir de um fluxo viável f^0, geraríamos uma seqüência $\{(c^i, f^i)\}_{i \in N}$, tal que

$$c^i = \text{solução ótima do (CA), com } f = f^{i-1} \in F$$

$$f^i = (f^i + \lambda_i h^i) \in F,$$

com a propriedade de melhora do critério para cada par (f^i, c^i). Isto é, para PD (T), $D(c^{i+1}) < D(c^i)$ e para PT (D), $T(f^{i+1}, ci + 1) < T(f^1, c^i)$.

Proporemos a seguir o método desta família e analisaremos seu comportamento para o caso contínuo. Por facilidade de expressão usaremos só a expressão projeção para significar projeção sobre o espaço das variáveis fluxo.

Para podermos estudar esta abordagem temos que analisar a existência de soluções ótimas e o comportamento do valor ótiral onde C_i é fechado, o que pode ser demonstrado com pequena variação nos argumentos da demonstração de (2.2).

Entre os dois problemas CD(·) e CT(·) há fortes relações no sentido de que ambos são condições necessárias de otimalidade para PD(·) e PT(·), portanto ambos são condições necessárias para Pareto eficiência de projeto e valem os análogos triviais de (2.5) e (2.6).

O caso contínuo nos permite um conjunto de resultados mais potentes. Portanto, ao longo desta seção, faremos a hipótese $C_i \equiv R_+$. Por exemplo, nestas condições

Fato 3.5 Para uma solução ótima de CD (T_{MAX}) (CT (D_{MAX})) a restrição de retardo (custo) é obedecida com igualdade.

Demonstração: omitida por ser particularização de

Fato 3.6 Sob a hipótese adicional HA, as funções v_D e v_T abaixo definidas são convexas, subdiferenciáveis e semicontínuas inferiormente, onde:

v_D : $\{T \in R \mid T > 0\} \to R$ é definida por:

$$v_D (T) = \inf \left\{ \sum_{i=1}^{m} d_i (c_i) \mid \sum_{i=1}^{m} t_i (\bar{f}_i, c_i) \leq \bar{q}\, T \text{ e } c_i > \bar{f}_i \right\},$$

e

v_T : $\{D \in R \mid D > D_0\} \to R$ é definida por:

$$v_T (D) = \inf \left\{ \sum_{i=1}^{m} t_i (\bar{f}_i, c_i) \mid \sum_{i=1}^{m} d_i (c_i) \leq D \text{ e } c_i > \bar{f}_i \right\}.$$

Demonstração: Vide HUMES [2].

Esta propriedade foi verificada empiricamente no pioneiro trabalho de GERLA [1] para redes de filas M/M/1 com custos do tipo ($\sum_{i=1}^{m} l_i \; c_i^{\alpha}$), onde $\alpha \in (0, 1)$.

Esta forte propriedade de convexidade e subdiferenciabilidade, levou-nos a esforços grandes (e mal-direcionados, em nossa presente opinião) para obter soluções do problema por esquemas próximos às idéias de decomposição de Benders.

Estas idéias parecem-nos mal direcionadas, pelo menos se acopladas à projeção usual, pois Benders está ligado à presença de suportes convexos e, portanto, à subdiferenciabilidade (se impusermos suportes diferenciáveis ou subdiferenciáveis) e, portanto, a convexidade. Convexidade esta, cuja principal característica é a "globalidade" de mínimos locais.

Estes comentários torna-se-ão mais claros perante o próximo fato e seu uso posterior. Antes, porém, de podermos enunciar o fato significante, há que introduzir a definição imediatamente abaixo:

Definição 3.7 As funções custo ótimo D* (·,·) e retardo ótimo T* (·,·) em função do fluxo e, respectivamente, do retardo máximo e do orçamento máximo são definidas por:

$$D^* (f, T) = \inf \{\Sigma_{i \in I(f)} \; d_i (c_i) \mid \Sigma_{i \in I(f)} \; t_i \; (f_i, c_i) \lesssim \bar{q} \; T \text{ e}$$

$$c_i > f_i \quad \text{para} \quad i \in I(f) \, ,$$

$$T^* (f, T) = \inf \{\Sigma_{i \in I(f)} t_i (f_i, c_i) \mid \Sigma_{i \in I(f)} d_i (c_i) \leq D \text{ e}$$

$$c_i > f_i \text{ para } i \in I(f) ,$$

onde

$$I(f) = \{i \in \bar{A} \mid f_i > 0\}.$$

Por conveniência, nós trabalharemos somente com pares $(f, T) \in F \times R_{++}$ e com pares $(f, D) - F \times \{x \in R \mid x\ \Sigma_{i \in I(f)} d_i (f_i)\}$. Nestas condições, o *inf* presente em ambas definições torna-se um mínimo.

Com estas definições em mente, podemos afirmar:

<u>Fato 3.8</u> Para todo real estritamente positivo T, a função $D^* (\cdot, T) : F \to R$ é côncava. Mais ainda, se vale (H5A) (custos estritamente côncavos), então $D^* (\cdot, T)$ é estritamente côncava em F.

<u>Demonstração</u>:. há várias formas de provar este resultado, mas a que consideramos mais simples e interessante é aquela em que tratamos o problema CD (T) como sendo um de designação de graus de congestionamento $\rho_i = f_i/c_i$.

É trivial verificar, usando (3.4) que

$\forall f \in F, \forall T > 0,$

$$D^* (f, T) = \min \{\Sigma_{i \in I(f)} d_i (c_i) \mid \Sigma_{i \in I(f)} t_i (f_i, c_i) \leq \bar{q}\, T \text{ e}$$

$$f_i > c_i \text{ para } i \in I(f)\}$$

$$= \min \{\Sigma_{i \in I(f)} d_i (f_i/\rho_i)^{*} \mid \Sigma_{i \in I(f)} t_i (\rho_i, 1) \leq \bar{q}\, T_{MAX} \text{ e}$$

$$\rho_i - (0, 1) \text{ para } i \in I(f)\}$$

$$= \min \{\Sigma_{i \in I(f)} d_i (f_i/\rho_i) \mid \rho \in RO (f)\}$$

onde

$$RO\ (f) = \{\rho \in R^m \mid \Sigma_{i \in I(f)}\ t_i\ (\rho_i,\ 1) \leq T_{MAX}\ \ e$$

$$\rho_i \in (0,\ 1)\ \ para\ \ i = 1,\ 2,\ \ldots,\ m\}$$

Para tal constatação basta lembrar que

$\forall\ x - (0,1),\ d_i\ (0/x) = d_i\ (0) = 0,\ \ i = 1,\ 2,\ \ldots,\ m$ por (H).

Portanto, $\forall\ \lambda \in (0,\ 1),\ \forall\ (f^1,\ f^2) \in F \times F$ convexo; $f^1 \neq f^2$, se utilizarmos a notação $f\ (\lambda) = \lambda\ f^1 + (1 - \lambda)\ f^2$, temos

$$D^*\ (f\ (\lambda),\ T) = \min \left\{ \sum_{i=1}^{m} d_i\ (f_i\ (\lambda)/\rho_i)\ \middle|\ \rho \in RO\ (f(\lambda)) \right\}.$$

Mais ainda, $\forall\ i \in I\ (f\ (\lambda)) = I\ (f^1) \cup I\ (f^2)$, e $\forall\ z \in (0,\ 1)$

$$d_i\ (f_i\ (\lambda)/z) \geq \lambda\ d_i\ (f_i^1/z) + (1 - \lambda)\ d_i\ (f_i^2/z) \quad \text{(por (H5))}$$

Note-se que, se (H5A) vale, a desigualdade estrita vale e que para $i \notin I\ (f_\lambda),\ d_i\ (f_i\ (\lambda)/z) = 0$.

Portanto, como

$$(I\ (f\ (\lambda)) \equiv I\ (f^1) \cup I\ (f^2) \Rightarrow RO\ (f^i) \subset RO\ (f\ (\lambda)),$$
$$i = 1,\ 2,$$

segue que

$$D^* (f(\lambda), T) \geq \min \left\{ \sum_{i=1}^{m} \lambda\, d_i (f_i^1/\rho_i) + (1-\lambda)\, d (f_i^2/\rho_i) \mid \rho \in RO (f(\lambda)) \right\}$$

$$\geq \lambda \min \left\{ \sum_{i=1}^{m} d_i (f_i^1/\rho_i) \mid \rho \in RO (f(\lambda)) \right\}$$

$$+ (1-\lambda) \left\{ \sum_{i=1}^{m} d_i (f_i^2/\rho_i) \mid \rho \in RO (f(\lambda)) \right\}$$

$$\geq \lambda D^* (f^1, T) + (1-\lambda) D^* (f^2, T)$$

Sendo que a primeira desigualdade é estrita caso (H5A) valha. □

É interessante notar que as únicas propriedades de F utilizadas foram convexidade de F e que $f \in F \Rightarrow f \geq 0$. Portanto, o resultado acima enunciado para $D^* (\cdot, T) : F \to R$ é válido para $D^* (\cdot, T) : R_+^m \to R$.

Esta constatacão é a base para a seguinte afirmação:

<u>Fato 3.9</u> Para todo real estritamente positivo T, para toda partição (I, J) de {1, 2, ..., m},

$$D^* (\cdot, T) = \{f \in R^m \mid f_i > 0 \text{ para } i \in I, \text{ e } f_j = 0 \text{ para } i \in J\} \to R$$

é contínua e superdiferenciável.

<u>Demonstração</u>: no caso I = {1, 2, ..., m} e J ≡ ∅, o resultado segue trivialmente do fato de que funções côncavas definidas em abertos são contínuas e superdiferenciáveis.

Para o caso J não vazio, o raciocínio é o mesmo utilizando-se como espaço o subespaço linear $\{f \in R^m \mid f_j = 0$ para

$j \in J\}$ e o interior relativo a este subespaço, sendo que as coordenadas j ($j \in J$) do supergradiente são indeterminadas. □

A restrição de trabalharmos com fluxos f com exatamente as mesmas componentes não nulas nos permite enunciar um resultado de concavidade estrita no caso de custos lineares e redes de filas M/M/1:

<u>Fato 3.10</u> No caso de custos lineares e redes de filas M/M/1, a função

$$D^* (\cdot, T) : \{f \in F \mid f_i = 0 \Longleftrightarrow j \in J\} \to R$$

é C^∞ e estritamente côncava, para todo real estritamente positivo T.

<u>Demonstração</u>: vide HUMES [2].

O Fato 3.10 permite-nos reforçar parcialmente (3.8), enunciando

<u>Fato 3.11</u> No caso de redes de filas M/M/1, a função

$$D^* (\cdot, T) : \{f \in F \mid f_j = 0 \Longleftrightarrow j \in J\}$$

pe estritamente côncava.

<u>Demonstração</u>: sem perda de generalidade, suponhamos $I (f) = \{1, 2, \ldots, m\}$ para $f \in \bar{F} = \{f \in F \mid f_j = 0 \Longleftrightarrow j \in J\}$.

Sejam $(f^1, f^2, \lambda) \in \bar{F} \times \bar{F} \times (0, 1)$ e c^* a solução ótima de CD (T) para $f = f (\lambda) = \lambda f^1 + (1 - \lambda) f^2$.

Claramente $c^* \in R^m_{++}$ e, portanto, D (c) é superdiferenciável em c*, pois, por (H5), D(·) é côncava. Nestas condições,

$$\forall\ c \in R^m_{++}\ D\ (c) \leq D_L\ (c) = D\ (c^*) + \langle \gamma,\ c - c^* \rangle.$$

Por (3.10),

$$D_L{}^*\ (\lambda\ f^1 + (1 - \lambda)\ f^2,\ T) > \lambda\ D_L{}^*\ (f^1,\ T) +$$

$$+ (1 - \lambda)\ D_L{}^*\ (f^2,\ T)$$

$$\geq \lambda\ D^*\ (f^1,\ T) + (1 - \lambda)\ D^*\ (f^2,\ T)$$

onde

$$D_L{}^*\ (f,\ T) = \min\ \{D_L\ (c)\ |\ \sum_{i=1}^{m} t_i\ (f_i,\ c_i) \leq \bar{q}\ T\ \text{ e }$$

$$c_i > f_i, \quad i = 1,\ 2,\ \ldots,\ m\}$$

e, claramente,

$$D_L{}^*\ (\lambda\ f^1 + (1 - \lambda)\ f^2,\ T) = D^*\ (\lambda\ f^1 + (1 - \lambda)\ f^2,\ T)$$

e, portanto, segue a tese. $\square$

No presente ponto, duas questões tornam-se naturais: o que podemos afirmar sobre T* (·, D) e qual o comportamento de D* (·, T) T* (·, D)) quando alguma componente do fluxo anula-se.

Quanto ao comportamento de T^* $(\cdot, D))$ os resultados obtidos não são tão fortes quanto os obtidos para D^* $(\cdot, T))$. Esta afirmação é de certo modo frustrante, pois as funcoes T^* $(\cdot, D)$ e D^* $(\cdot, T)$ estão intimamente ligadas por:

<u>Fato 3.12</u> Para todo $f \in F$, valem as seguintes relações

(a) $\forall\ T > 0,\ T^*(f, D^*(f, T)) = T$

(b) $\forall\ D > D(f),\ D^*(f, T^*(f, D)) = D$

<u>Demonstração</u>: trivial a partir de (3.3) e (3.4). □

A única aparente assimetria em (a) e (b) acima é a imposição de $D > D(f)$. Esta assimetria torna-se presente ao considerarmos combinações convexas de fluxos, pois, excessão feita ao caso linear (custos lineares), não podemos afirmar que $\{f \in F \mid D(f) > D\}$ é convexo. Intruitivamente, afirmaríamos o oposto, isto é, que se os custos são estritamente côncavos, existiriam sempre $(f^1, f^2) - F \times F$ tais que $D(f^i) > D$, $i = 1, 2$ e $D(0{,}5\ f^1 + 0{,}5\ f^2) < D$. Em termos mais precisos, podemos enunciar:

<u>Fato 3.13</u> Para todo real (positivo) D, o conjunto $\{f \in F \mid D(f) \geq D\}$ é convexo.

<u>Demonstração</u>: trivial pela concavidade do custo (H5). □

Com estas considerações, podemos apresentar uma versão mais fraca de (3.8), a qual é:

<u>Fato 3.14</u> Para todo real $D > \{\inf D(f) \mid f \in F\}$ a função

$$T^* (\cdot, D) : \{f \in F \mid D\} \to \mathbb{R}$$

é quasicôncava, isto é, sendo

$$A = \{f \in F \mid D(f) < D\},$$

$$\forall (f^1, f^2, \lambda) \in A \times A \times (0, 1) : (\lambda f^1 + (1 - \lambda) f^2) \in A$$

$$T^* (\lambda f^1 + (1 - \lambda) f^2, D) \geq \min_{i=1,2} \{T^* (f^i, D)\}.$$

Demonstração: por simplicidade, utilizamos

$$T^i = T^* (f^i, D), \quad i = 1, 2;$$

$$T_\lambda = T^* (f(\lambda), D),$$

onde

$$f(\lambda) + \lambda f^1 + (1 - \lambda) f^2$$

Suponhamos, sem perda de generalidade, $T^1 \leq T^2$ e que, por contradição, existe λ, tal que $T_\lambda < T^1 \leq T^2$. Então

$$D = D^* (\lambda f^1 + (1 - \lambda) f^2, T_\lambda) \quad \text{(por 3.12)}$$

$$\geq \lambda D^* (f^1, T_\lambda) + (1 - \lambda) D^* (f^2, T_\lambda) \quad \text{(por 3.8)}$$

$$> \lambda D^* (f^1, T^1) + (1 - \lambda) D^* (f^2, T^2) \quad (\text{pois } T_\lambda < T^1 \leq T^2)$$

$$= D \quad \text{(por 3.12)} \quad \square$$

Note-se que com o mesmo raciocínio de (3.14) podemos enunciar:

<u>Fato 3.15</u> Para todo real $D > \inf \{D(f) \mid f \in F\}$, a função

$$T^*(\cdot, D) : \{f \in F : D(f) < D\} \to R$$

é estritamente quasicôncava.

Demonstração: omitida por ser idêntica a (3.14), substituindo $(T_\lambda \leq T^1 < T^2)$. $\square$

A questão natural que surge é sobre a quasiconcavidade de $T^*(\cdot, D)$ em F.

Claramente, se (f^1, f^2) goza da propriedade $D(f^i) \geq D$, por (3.13) $\forall \lambda \in (0, 1)$, $f(\lambda) = \lambda f^1 + (1 - \lambda) f^2$ goza da propriedade $D(f(\lambda)) \geq D$.

Neste caso, com a convenção $(+\infty \geq +\infty)$, a caracterização de quasiconcavidade mantem-se.

Portanto, nos interessa o caso onde $D(f^1) < D$ e $D(f^2) \geq D$. Neste caso vale:

<u>Fato 3.16</u> Sejam $(f^1, f^2) \in F$, tais que $D(f^1) < D$ e $D(f^2) \geq D$, então: existe $\bar{\lambda} \in [0, 1)$, tal que

(i) $D(\lambda f^1 + (1 - \lambda) f^2) < D \iff \lambda > \bar{\lambda}$,

(ii) $D(\bar{\lambda} f^1 + (1 - \bar{\lambda}) f^2) = D$.

<u>Demonstração</u>: trivialmente, pela continuidade de $D(\cdot)$.

$$A = \{\lambda \in [0, 1) \mid D(\lambda f^1 + (1 - \lambda) f^2) \geq D\}$$

+e fechado.

Pela continuidade de $D(\cdot)$ e por $D(1 f^1 + (1 - 1) f^2) < D$, A possui limite superior menor que 1 e, portanto, é bem definido

$$\bar{\lambda} = \sup \{\lambda \in A\} < 1 \quad \text{e} \quad D(\bar{\lambda} f^1 + (1 - \bar{\lambda}) f^2) = D \quad \square$$

<u>Fato 3.17</u> Nas condições de (3.16), $\exists\ \varepsilon > 0$, tal que

$$\forall\ \lambda \in (\bar{\lambda}, \bar{\lambda} + \varepsilon],\ T^* (f(\lambda), D) > T^* (f^1, D).$$

<u>Demonstração</u>: como $D(f^2) > D(f^1)$, podemos assumir, sem perda de generalidade, que $f_1^2 > f_1^1$ e, portanto, $f_1(\lambda)$ é decrescente com λ.

Consideremos $\bar{c}$ definido por

$$t_1 (f_1 (0{,}5 (1 + \bar{\lambda})), \bar{c}) = T^* (f^1, D),$$

cuja existência é garantida por (H8) a (H11), e seja

$$d_1 = d_1 (\bar{c}) > 0, \quad \text{pois} \quad f_1^2 > f_1^1 \geq 0$$

Seja $D' = D - d_1$.

Pela continuidade de $D(\cdot)$ e pela definição de A e $\bar{\lambda}$, como em (3.16) segue que $\exists\ \varepsilon' > 0$, tal que

$\forall \lambda \in (\bar{\lambda}, \bar{\lambda} + \varepsilon], \quad D > D(f(\lambda)) \geq D'$.

Tomando $\varepsilon = \min \{\varepsilon', 0{,}5 (1 - \bar{\lambda}))\}$, temos que, sendo $c^* (\lambda)$ a solução ótima de $C\bar{T} (D)$,

$$T^* (f (\lambda), D) \geq t_1 (f_1 (\lambda), c_1^* (\lambda)),$$

$$d_1 (c_1^* (\lambda)) < D - D (f (\lambda)) \quad D - D' = d_1.$$

Portanto, $c^* (\lambda) < \bar{c}$.

Como $f_1 (\lambda)$ é decrescente com λ, $\forall \lambda \in (\bar{\lambda}, \bar{\lambda} + \varepsilon]$

$$\begin{aligned} T^* (f (\lambda), D) &\geq t_1 (f_1 (\lambda), c_1^* (\lambda)) \\ &> t_1 (f_1 (0{,}5 (1 + \bar{\lambda}), c_1^* (\lambda)) \\ &> t_1 (f_1 (0{,}5 (1 + \bar{\lambda}), \bar{c})) \\ &= T^* (f^1, D) \quad \square \end{aligned}$$

Utilizando (3.16) e (3.17) é fácil provar que

<u>Fato 3.18</u> Para todo real (positivo) D, a função

$$T (\cdot, D) : F \to R \cup \{ \infty\}$$

é estritamente quasicôncava.

<u>Demonstração</u>: omitida por ser trivial. $\square$

O resultado aparentemente desejável de que $T^* (\cdot, D)$ fosse côncava é falso em geral, como pode ser verificado com um exemplo simples no caso de custos lineares e redes de filas M/M/1.

Estes fatos aparentemente bizontinos, nos permitem apresentar um algoritmo mais simples que o de dupla projeção e que possui a propriedade de convergência em um número finito de passos.

A base deste algoritmo é apresentada pelos fatos que se seguem.

Fato 3.19 Seja $\bar{c}$ a solução do CT (D) (CD (T) com $f = \bar{f}$). Se para o (FA), com $c = \bar{c}$, existe uma direção de descida no ponto $\bar{f}$, então esta é uma direção de descida para $T^* (\cdot, D)$ $(D^* (\cdot, T))$ no ponto $f = \bar{f}$.

Demonstracão: a demonstração para $T^* (\cdot, D)$ é calcada no fato de que uma queda de retardo para direções de descida.

A demonstracão para $D^* (\cdot, T)$ é consequência do fato de que se existe uma direção de descida, para o ponto $f = \bar{f}$, no (FA) com $c = \bar{c}$

$$D^* (\bar{f} + \lambda h, T) \leq D (\bar{c}) = D^* (\bar{f}. T)$$

e

$$T (\bar{f} + \lambda h, \bar{c}) < T,$$

para λ em uma vizinhança da origem e $\lambda > 0$. □

Fato 3.20 Seja $w : F \to R$ uma função quasicôncava em F e seja h uma direção de descida para $w(\cdot)$ em $f \in F$. Isto é, $\exists\ (\bar{\lambda}, \bar{h}) \in R_{++} \times (F - F)$, tal que

(i) $\forall\ \lambda \in [0, \bar{\lambda}],\ (\bar{f} + \lambda\ \bar{h}) \in F$

(ii) $\forall\ \lambda \in [0, \bar{\lambda}],\ w\ (\bar{f} + \lambda\ \bar{h}) < w\ (\bar{f})$,

então $\forall\ \lambda \geq 0$, tal que $(\bar{f} + \lambda\ \bar{h}) \in F$

$w\ (\bar{f} + \lambda\ \bar{h}) < w\ (\bar{f})$.

<u>Demonstração</u>: supondo, por contradição, que $\exists\ \alpha > 0$:

$(\bar{f} + \alpha\ \bar{h}) \in F$

$w\ (\bar{f} + \alpha\ \bar{h}) \geq w\ (\bar{f})$,

então existe $p \in [0, 1]$, tal que

$w\ (p\ (\bar{f} + \alpha\ \bar{h}) + (1 - p)\ \bar{f}) \geq w\ (\bar{f})$,

e

$0 < p\ \alpha < \bar{\lambda}$

o que gera a contradição. □

A relevância do Fato 3.20 é clara se lembrarmos que $T^*\ (\cdot, D)$ é quasicôncava em F e $D^*\ (\cdot, T)$ é (estritamente) côncava e portanto quasicôncava em F.

Nestas condições propomos o seguinte método:

<u>Método 3.21</u> Seja f^{0} um fluxo viável no problema de designação de capaicades (isto é, $D(f^{0}) > D$ para PT (D) ou $T > 0$ para o PD (T)).

Passo 1) Normalização

Encontre c^{0} solução do problema de designação de capacidades.

Para $c = c^{0}$, para todo par (r, s), com $q_{rs} > 0$, encontre uma árvore de caminho mais custo, com distâncias

$$\alpha_{kl} = \alpha_{lk} = \alpha_i = \frac{(\partial t_i)\,(f_i^0,\, c_i^0)}{\partial f_i} \quad \text{para } \bar{a}(i) =$$

$$= \{k,\ l\}, \quad i = 1, 2, \ldots, m$$

Envie todas as mensagens de k a l pelo caminho mais curto encontrado, para todos os pares (k, l). Caso esta seja a situacão com o fluxo f^{0}, pare. Caso contrário, chame o fluxo obtido de f^{1}.

Passo 2) Melhora .

Para o fluxo f^{i}, $i \geq 1$, encontre c^{i} uma solução ótima do CA, com $f = f^{i}$.

Para $c = c^{i}$, encontre, se existir, uma árvore de caminhos mais curtos para $q_{rs} > 0$, que não seja o roteamento associado a f^{i} (a métrica usada é):

$$\alpha_{kl} = \alpha_{lk} = \alpha_i = \frac{\partial t_j\,(f_j^i,\, c_j^i)}{\partial t_j} \quad \text{para } \bar{a}(j) = \{k,\ l\}$$

e cujo caminho mais curto indicado seja estritamente melhor que o associado ao atual roteamento.

Se não existir tal árvore, pare. Caso contrário, altere o fluxo para que os caminhos mais curtos sejam obedecidos. O novo fluxo é f^{i+1}.

Retorne parao passo 2. □

Fato 3.22 O método proposto em (3.21) converge em um número finito de passos para o ponto $(\bar{f}, \bar{c})$, tal que:

(P1) $\bar{c}$ é a solução ótima do (CA), com $f = \bar{f}$;

(P2) $\bar{f}$ é a solução ótima do (FA), com $c = \bar{c}$.

Demonstração: após a normalização, os fluxos f^{o}, $i \geq 1$, estão associados biunivocamente a arborescência do grafo G.

Como as arborescências são em número finito e a cada passo a função objetivo do problema é diminuída estritamente, o método para em um número finito de passos.

Analisando as regras de parada, segue a tese. □

O método (3.21) é essencialmente um método similar ao simplex. Além da garantia de convergência e a aparente não necessidade de (HA), o método apresenta uma grande vantagem sobre o método de dupla projeção, que é a de não implicar na solução do (FA).

Esta vantagem é a substituição da solução do (FA), que exige uma série de iterações, onde, em cada uma, devem ser construidas $\binom{n}{2}$ árvores de Dijkstra e uma busca unidimensional, pelo conjunto de uma direção de descida. Usando a idéia de desvio de fluxo, este último computo exige, no pior caso, $\binom{n}{2}$ árvores de Dijkstra.

É importante notar que em Kleinrock [5], a semente desta idéia está presente, quando ele recomenda desvio total de fluxo. Infelizmente esta idéia aplicada à solução do (FA) pode levar a perda de viabilidade, além de não reduzir o número de desvios de fluxo a um só, como aqui é feito.

Um ponto de aceleração ao algoritmo de desvio de fluxo seria a utilização de árvores prévias para auxiliar a construção das árvores seguintes. Tais idéias foram implementadas por Bezerra [07], mas ainda há uma esperança de melhoria sobre este aspecto.

A facilidade do processo de encontrar a direção de máxima descida sugere naturalmente o caso do algoritmo "steepest decent", apesar de que em geral este método seja criticável quando comparado a métodos de segunda ordem, para a minimização correspondente a (FA). Para especificar completamente o algoritmo, só resta especificar a busca unidimensional. O usual na tradição de Zoutendijk é buscar o mínimo a logo da semireta $(\bar{f}+\lambda\bar{h})$, com a restrição de viabilidade. Como sabemos que $\lim_{f_i \to \bar{c}_i} t_i(f_i, \bar{c}_i) = +\infty$, a viabilidade praticamente restringe-se a manter $f_i \geq 0$, $i=1,\ldots,m$, o que é automaticamente garantido, com $\lambda \leq 1$. É interessante notar que nos casos rodados sobre os exemplos da rede LARC, obtivemos consistentemente $\lambda=1$. Tal não ocorreu em exemplos gerados aleatoriamente.

Em geral, recomendaríamos o uso da regra de Armijo, baseadas nas experiências relatadas por Polak [8].

Apesar de sua simplicidade teórica, a solução do (FA) tende a consumir ordens de grandeza de tempo a mais que a solução do (CA).

IV - Comentários finais

A utilização do método proposto simplifica a solução do problema de designação de fluxos e capacidades, no sentido de encontrar um ponto estacionário de Kuhn-Tucker, mas a constatação de "concavidade" nas funções T*(-,D) e D*(-,T) é clara indicação da existência de mínimos locais. Em particular, quando o fluxo físico restringe-se a uma árvore (m=n-1 e grafo conexo), pode-se provar [2] que estamos em um mínimo local.

Assim sendo, maiores resultados orientados para encontrar a solução de PD(T) ou PT(D) ainda não existem afora a recomendação de Gerla[1] de utilizar vários pontos iniciais viáveis e a esperança de obtermos resultados positivos seguindo as idéias de Tuy et alli [9], em particular, no caso de custos lineares e filas M/M/1.

Consideremos entretanto interessante neste trabalho a junção de intuição encontrada nos escritos de Gerla [1] e Kleinrock [5], com o tratamento formal dado a funções valor ótimo dependendo de argumentos do problema, levando a algoritmo mais simples e eficiente, nas linhas de projeção - direções viáveis, como sugerido em Geoffrion.

BIBLIOGRAFIA

(1) Gerla, M., The design of store-and-forward networks for computer networks. Los Angeles, 1973, 300p, Tese (Doutorado) - University of California.

(2) Humes Jr., C. Tópico de otimização e redes de computadores. São Paulo, 1988, 109p, Tese (Livre-Docência)-IME-Universidade de São Paulo.

(3) Lin, J. G. Maximal vectors and multiobjective optimization. Journal of Optimization Theory and Applications, 18(1): 41-64,1976.

(4) Lin, J. G. Multiobjective problems: Pareto-optimal solutions by method of proper equality constraits. IEEE Transactions Automatic Control, 21(5): 641-650,1976.

(5) Kleinrock, L. Queuing Systems. New York, John Wiley, 1975-76,2v.

(6) Geoffrion, A. M. Elements of large-scale mathmatical programming. Management Science, 16(11): 652-691,1970.

(7) Bezerra, J.R. M. Sobre problemas de otimização de redes de computadores. São Paulo, 1984, 223p. Dissertação (mestrado)-IME-USP.

(8) Polak, E. Computational methods im optimization. New York, Academic Press, 1971, 329p.

(9) Tuy, H.; Thieu, T. V.; Thai, NG Q. A conical algorithm for globally minimizing a concave function over a closed convex set. Mathematics of Operations Reserarch, 10(3): 498-514,1985.