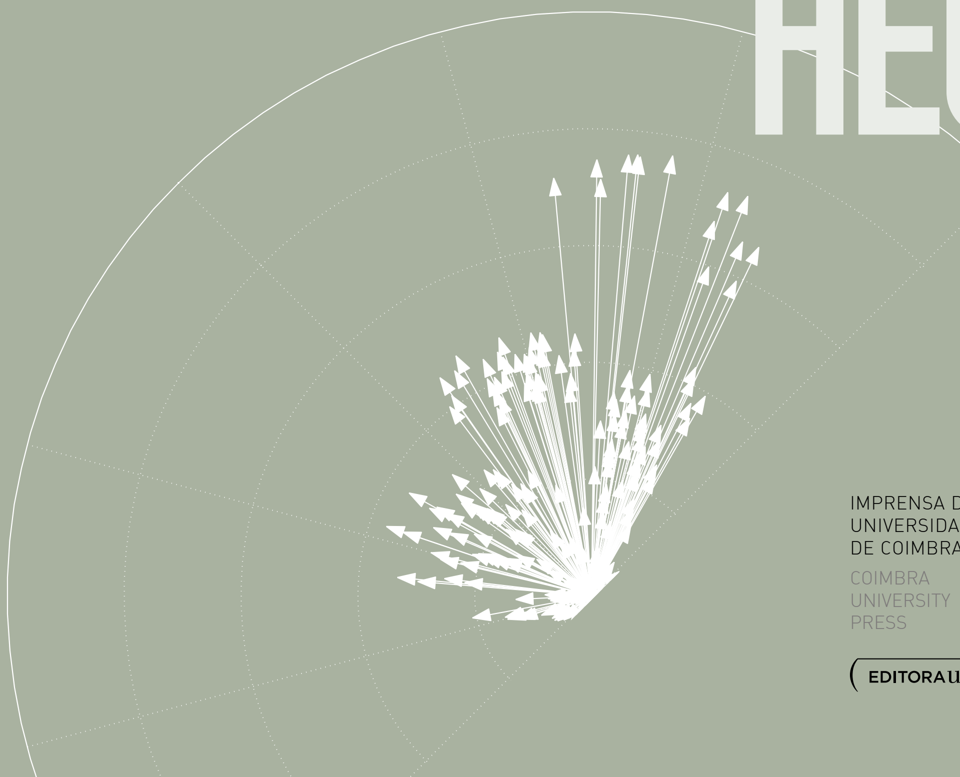
MANUAL DE

COMPUTAÇÃO EVOLUTIVA E METAHEURÍSTICA

ANTÓNIO GASPAR-CUNHA
RICARDO TAKAHASHI
CARLOS HENGGELER ANTUNES
COORDENADORES



IMPRENSA DA UNIVERSIDADE DE COIMBRA
COIMBRA UNIVERSITY PRESS

EDITORA UFMG

CAPÍTULO 16

Tomada de Decisão em Ambiente Multiobjectivo

Carlos Henggeler Antunes[a,c] *Maria João Alves*[b,c] *João Clímaco*[b,c]

[a]*Departamento de Engenharia Electrotécnica e de Computadores*
Universidade de Coimbra

[b]*Faculdade de Economia*
Universidade de Coimbra

[c]*Instituto de Engenharia de Sistemas e Computadores de Coimbra*

O mais forte argumento em favor da utilização de modelos e métodos para apoio à decisão em presença de critérios múltiplos reside na própria realidade: os problemas reais são intrinsecamente multidimensionais, sendo expectável que a consideração explícita de múltiplos eixos de avaliação do mérito das soluções potenciais concorra para a obtenção de soluções mais adequadas. A complexidade dos problemas reais que surgem nas sociedades tecnológicas modernas é essencialmente caracterizada pela pluralidade de perspectivas de análise, reflectindo aspectos económicos, sociais, políticos, físicos, de engenharia, administrativos, psicológicos, éticos, estéticos, etc, num dado contexto. Dado que, em geral, não existe uma solução admissível (face à tecnologia, aos recursos disponíveis e aos requisitos exigidos) que garanta o melhor valor em todos esses múltiplos aspectos de avaliação, o processo de tomada de decisões neste tipo de problemas será, na maioria das circunstâncias, redutor se o circunscrevermos à procura da solução óptima para um desses objectivos ou para algum objectivo "agregado", como seja a eficiência do ponto de vista económico medida por um qualquer indicador.

Os modelos para apoio à tomada de decisões, geralmente sob a forma de modelos de programação matemática, bem como a percepção dos problemas por parte dos decisores, tornam-se, em geral, mais realistas se os diferentes aspectos da realidade forem explicitamente considerados, construindo uma família coerente de critérios para avaliar as alternativas admissíveis. No entanto, note-se que os argumentos em favor da utilização de modelos que consideram explicitamente múltiplos aspectos de avaliação das soluções potenciais não se esgotam em argumentos de natureza realista. As abordagens multicritério comportam um valor acrescentado, quer na fase de estruturação do problema e construção dos modelos, quer na fase de análise crítica dos resultados, ao permitir descobrir um leque de soluções com diferentes características, e não apenas uma solução óptima, estabelecendo distintos compromissos entre os aspectos de avaliação. A agregação de preferências multicritério contribui ainda, pelo menos parcialmente, para a inclusão nos modelos de investigação operacional (IO) da decisão em grupo e da negociação, presentes em muitos casos reais mas historicamente ausentes dos modelos tradicionais de IO.

De uma forma geral, subdividem-se os problemas multicritério em problemas multiatributo e multiobjectivo. O problema multiatributo refere-se usualmente aos casos em que as acções potenciais, em número finito, são explicitamente conhecidas, bem como os respectivos índices de mérito avaliados segundo os vários critérios (definição enumerativa). Um exemplo típico é o problema de decisão que consiste na escolha de um automóvel, de entre as alternativas disponíveis num dado mercado, em que preço, consumo em cidade e em estrada, espaço, etc., são os critérios que são tidos em conta pelo decisor. Este problema pode ser estruturado através de uma matriz de impactos em que o elemento genérico d_{ij} representa o desempenho da alternativa i segundo o critério j.

Para além deste problema de escolha poderíamos considerar problemas de ordenação (*ranking*), cujo resultado seria a ordenação - eventualmente incompleta e/ou admitindo *ex-aequo* - das alternativas da melhor para a menos boa, ou problemas de categorização (*sorting*), cujo resultado seria a afectação de cada alternativa a uma categoria de um conjunto ordenado de categorias previamente definido. (Roy, 1996) vai um pouco mais longe no que se refere à flexibilidade do apoio multicritério à decisão, quando admite que em muitos casos a análise dos problemas pode conduzir apenas ao aclaramento do processo de decisão e/ou ao aconselhamento do decisor, não havendo lugar a propostas de natureza prescritiva.

O problema de optimização multiobjectivo refere-se aos casos em que as acções potenciais são definidas implicitamente por um conjunto de restrições (definição analítica). O espaço das acções potenciais (espaço de decisão) é mapeado no espaço dos objectivos, no qual cada solução (i.e., uma instanciação das variáveis de decisão) tem como representação um vector, cujos componentes são os correspondentes valores de cada função objectivo. Daqui decorre a designação comum de problemas de optimização vectorial. Este capítulo é dedicado a este segundo tipo de problemas, focando sobretudo os problemas de programação linear multiobjectivo (PLMO) mas salientando também algumas características importantes de problemas lineares com variáveis inteiras e problemas não lineares.

Os termos critério e objectivo serão usados de forma intermutável, embora pressupondo a distinção anterior, geralmente adoptada na literatura.

Perante a existência de múltiplas funções objectivo, a noção de solução óptima (que conduz ao melhor valor admissível para uma única função objectivo) cede lugar à noção de solução não dominada. Uma solução não dominada (também designada por óptima de Pareto, eficiente ou não inferior; uma distinção mais formal entre estes termos será feita mais à frente) caracteriza-se por não existir outra solução admissível (i.e., satisfazendo um dado conjunto de restrições) que melhore simultaneamente todos os objectivos: a melhoria numa função objectivo só pode ser alcançada à custa da degradação do valor de pelo menos uma das outras funções objectivo do modelo matemático.

Sendo o conceito essencial em optimização multiobjectivo, o conceito de solução não dominada é "pobre" no sentido em que é pouco discriminante. Ou seja, as soluções não dominadas são não comparáveis entre si e, portanto, esta operação de comparação não fornece uma recomendação final de qual adoptar como solução final do problema.

Na resolução de um problema com uma única função objectivo o processo de pesquisa da melhor solução é puramente técnico, no sentido em que a solução óptima está implícita no modelo, cabendo ao algoritmo de optimização a sua determinação. Assim, não há lugar à tomada de decisões dado que esta solução é inequivocamente o plano a adoptar face ao modelo matemático instanciado com um conjunto de dados. Num problema multiobjectivo torna-se necessário fazer intervir no processo de pesquisa não apenas meios técnicos de calcular soluções não dominadas mas também informação sobre as preferências do decisor, que enriqueçam o conceito de solução não dominada de modo a permitir discriminar entre elas. A estrutura de preferências do decisor representa um conjunto de opiniões, valores, convicções e perspectivas da realidade em causa, que configuram um modelo pessoal da realidade sobre o qual aquele se apoia para avaliar diferentes possibilidades de acções potenciais. Ou seja, não é possível classificar uma solução como boa ou má apenas com referência ao modelo matemático e às técnicas de resolução: a qualidade de uma decisão é influenciada pelos aspectos organizacionais, políticos, culturais, etc., subjacentes ao processo de decisão (Roy, 1990).

O problema multiobjectivo pode então colocar-se como a escolha, de entre os elementos do conjunto de soluções não dominadas, de uma que constitua a uma solução de compromisso aceitável pelo decisor tendo em atenção as suas preferências, que podem evoluir ao longo do processo de apoio à tomada de decisões. Este processo é entendido como uma entidade dinâmica, constituída por ciclos interactivos (muitas vezes iterativos) de geração de acções potenciais, avaliação, interpretação de informação, alterações de valores, aprendizagem e adaptação de preferências. Em certos casos o modelo matemático multiobjectivo pode servir apenas para seleccionar um número limitado de soluções não dominadas, com vista a uma análise mais detalhada a posteriori, tendo em conta dimensões do problema, em muitos casos de natureza qualitativa, não contempladas no modelo. Note-se que por vezes não se seleccionam para análise a posteriori apenas soluções não dominadas do problema original, mas também soluções ligeiramente dominadas, consideradas indiferentes face àquelas. De facto, faz todo o sentido não excluir de um estudo mais detalhado, incluindo outros vectores de análise, soluções que são apenas matematicamente dominadas por alguma das seleccionadas para esse estudo, sendo de facto indiferentes do ponto de vista prático no que se refere aos valores das funções objectivo do problema original.

Na secção 2 são apresentadas as noções básicas num contexto de programação linear multiobjectivo. Na secção 3 são apresentadas algumas extensões a modelos de programação linear multiobjectivo com variáveis inteiras e programação não linear multiobjectivo. Os processos de cálculo de soluções eficientes usando funções escalares substitutas são descritos na secção 4. Na secção 5 são sucintamente descritos alguns métodos para tratar problemas de programação linear multiobjectivo, nalguns casos facilmente adaptáveis a problemas lineares com variáveis inteiras e problemas não lineares. Presta-se especial atenção às abordagens interactivas. Como ilustração do funcionamento de um método interactivo, para programação linear multiobjectivo, é descrito na secção 6 o funcionamento do método TRIMAP - programação linear TRI-objectivo - Método de Aprendizagem Progressiva das soluções não dominadas - recorrendo a um exemplo ilustrativo. Na secção 7 são brevemente apresentadas vias de integração entre métodos em sistemas de apoio à decisão, visando tirar partido das características particulares de cada método. Na secção 8 são mencionados alguns casos de aplicação com envolvimento dos autores.

1. Programação Linear Multiobjectivo - Formulação e Conceitos Fundamentais

O problema de programação linear com múltiplas funções objectivo consiste na optimização de p funções objectivo lineares sujeitas a um conjunto de restrições lineares:

$$\begin{aligned}
&\max z_1 = f_1(\mathbf{x}) = \mathbf{c}_1\,\mathbf{x} = \textstyle\sum_{j=1}^{n} c_{1,j}\,x_j \\
&\max z_2 = f_2(\mathbf{x}) = \mathbf{c}_2\,\mathbf{x} = \textstyle\sum_{j=1}^{n} c_{2,j}\,x_j \\
&\dots \\
&\max z_p = f_p(\mathbf{x}) = \mathbf{c}_p\,\mathbf{x} = \textstyle\sum_{j=1}^{n} c_{p,j}\,x_j \\
&\text{sujeito a: } \begin{cases} \sum_{j=1}^{n} a_{i,j}\,x_j = b_i & i = 1, \dots, m \\ x_j \geq 0 & j = 1, \dots, n \end{cases}
\end{aligned} \tag{16.1}$$

ou

$$\begin{aligned}
&\text{Max}\quad \mathbf{z} = \mathbf{f}(\mathbf{x}) = C\mathbf{x} \\
&\text{sujeito a: } \mathbf{x} \in \mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \geq 0, A\mathbf{x} = \mathbf{b}, \mathbf{b} \in \mathbb{R}^m\}
\end{aligned} \tag{16.2}$$

C é a matriz dos coeficientes das funções objectivo (dimensão $p \times n$), cujas linhas são os vectores $\mathbf{c}_k$ (coeficientes da função objectivo $\mathbf{f}_k(\mathbf{x})$). A é a matriz dos coeficientes tecnológicos $(m \times n)$, que se admite, sem perda de generalidade, tendo todas as restrições sido convertidas em igualdades através da introdução de variáveis desvio (*slack* e *surplus*) auxiliares. $\mathbf{b}$ é o vector dos termos independentes (genericamente recursos disponíveis para restrições do tipo $\leq$ ou requerimentos para restrições do tipo $\geq$). Assume-se que a região admissível $\mathcal{X}$ é não vazia e compacta. Sem perda de generalidade, e de modo a facilitar a notação, considera-se que as funções objectivo são todas a maximizar.

Enquanto em programação com uma única função objectivo as soluções admissíveis do espaço de decisão $\mathbf{x} \in \mathcal{X}$ são mapeadas em $\mathbb{R}$, no caso multiobjectivo o espaço de decisão é mapeado num espaço p-dimensional $\mathcal{F} = \{\mathbf{z} = \mathbf{f}(\mathbf{x}) \in \mathbb{R}^p : \mathbf{x} \in \mathcal{X}\}$ designado por espaço dos objectivos. Neste espaço cada alternativa potencial $\mathbf{x} \in \mathcal{X}$ tem como representação um vector $\mathbf{f}(\mathbf{x}) = (z_1, z_2, \dots, z_p) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_p(\mathbf{x}))$ cujos componentes são os valores de cada função objectivo para esse ponto da região admissível (fig. 16.1).

Em geral, não existe uma solução admissível $\mathbf{x} \in \mathcal{X}$ que optimize simultaneamente todas as funções objectivo. De um ponto de vista operacional, "Max"representa a operação de determinar soluções eficientes.

Num contexto multiobjectivo, a noção de solução óptima cede lugar ao conceito de solução eficiente. Uma solução admissível para um problema multiobjectivo diz-se eficiente se e só se (sse) não existir outra solução admissível que melhore o valor de uma função objectivo, sem piorar o valor de, pelo menos, outra função objectivo.

Definição 1 (solução eficiente): *Uma solução $\mathbf{x}' \in \mathcal{X}$ é eficiente sse não existe uma outra solução $\mathbf{x} \in \mathcal{X}$ tal que $f_k(\mathbf{x}) \geq f_k(\mathbf{x}')$ para todo o $k (k = 1, \dots, p)$, sendo a desigualdade estrita para pelo menos um k, $f_k(\mathbf{x}) > f_k(\mathbf{x}')$. $\mathcal{X}_E$ representa o conjunto das soluções eficientes.*

O ponto $\mathbf{z} = \mathbf{f}(\mathbf{x})$ no espaço das funções objectivo é não dominado (não inferior) sse $\mathbf{x} \in \mathcal{X}_E$, ou seja $\mathcal{F}_E = \{\mathbf{z} = \mathbf{f}(\mathbf{x}) \in \mathcal{F} : \mathbf{x} \in \mathcal{X}_E\}$.

Em geral, enquanto o conceito de não dominância se refere ao espaço das funções objectivo, o conceito de eficiência refere-se ao espaço das variáveis de decisão, isto é a imagem de uma solução

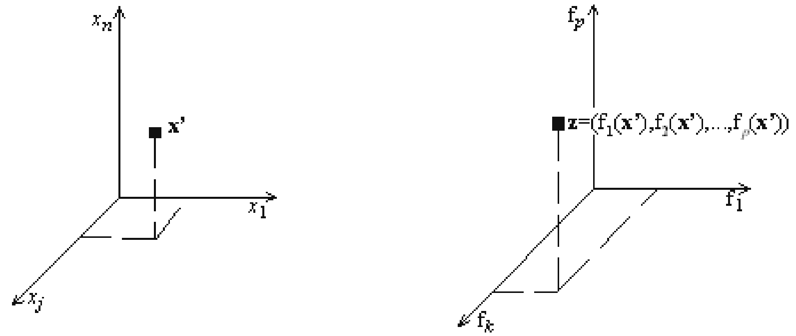


Figura 16.1: O espaço de decisão e o espaço dos objectivos

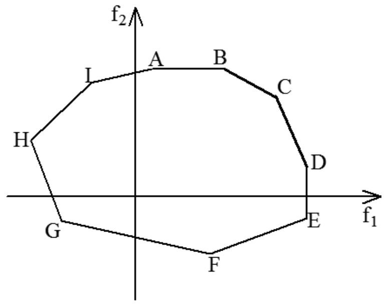


Figura 16.2: Ilustração dos diferentes tipos de soluções.

eficiente é uma solução não dominada. Em problemas de programação matemática multiobjectivo (optimização vectorial), os conjuntos de soluções eficientes e da respectiva imagem no espaço dos objectivos têm geralmente um número infinito de elementos.

Outra noção com interesse é a de solução fracamente eficiente, que pode considerar-se uma relaxação da anterior. Uma solução admissível para um problema multiobjectivo diz-se fracamente eficiente sse não existir outra solução admissível que melhore estritamente o valor de todas as funções objectivo.

Definição 2 (solução fracamente eficiente): *Uma solução* $\mathbf{x}' \in \mathcal{X}$ *é fracamente eficiente sse não existe uma outra solução* $\mathbf{x} \in \mathcal{X}$ *tal que* $f_k(\mathbf{x}) > f_k(\mathbf{x}')$ *para todo o* k $(k = 1, \ldots, p)$. $\mathcal{X}_{FE}$ *representa o conjunto das soluções fracamente eficientes.*

O ponto $\mathbf{z} = \mathbf{f}(\mathbf{x})$ *no espaço das funções objectivo é fracamente não dominado sse* $\mathbf{x} \in \mathcal{X}_{FE}$*, ou seja* $\mathcal{F}_{FE} = \{\mathbf{z} = \mathbf{f}(\mathbf{x}) \in \mathcal{F} : \mathbf{x} \in \mathcal{X}_{FE}\}$.

Note-se que, por definição, o conjunto de soluções fracamente eficientes inclui as soluções (estritamente) eficientes, antes definido. De uma forma geral, por razões de ordem prática, quando se mencionam as soluções fracamente eficientes não se estão a considerar as soluções estritamente eficientes.

A fig. 16.2 ilustra os conceitos de solução não dominada e fracamente não dominada, com duas funções a maximizar: as soluções sobre os segmentos AB e DE são fracamente não dominadas, excepto os pontos B e D, enquanto as soluções sobre os segmentos BC e CD são (estritamente) não dominadas.

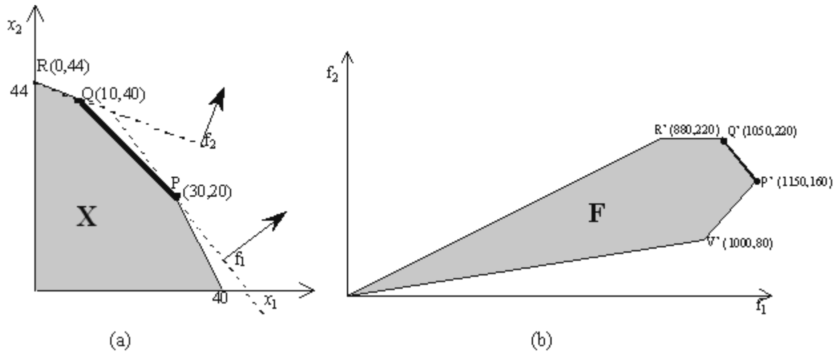


Figura 16.3: Exemplo 1: (a) espaço das variáveis de decisão; (b) espaço das funções objectivo

Para uma melhor ilustração dos conceitos de solução eficiente (não dominada) e fracamente eficiente (fracamente não dominada), considere-se o seguinte exemplo ilustrativo (Clímaco et al., 2003, p. 96).

Exemplo 1:

$$\max f_1(\mathbf{x}) = 25x_1 + 20x_2$$

$$\max f_2(\mathbf{x}) = 2x_1 + 5x_2$$

$$\text{sujeito a:} \begin{cases} x_1 + x_2 \leq 50 \\ 2x_1 + x_2 \leq 80 \\ 2x_1 + 5x_2 \leq 220 \\ x_1, x_2 \geq 0 \end{cases}$$

Na figura 16.3 está representada a região admissível do problema no espaço das variáveis de decisão (a) e no espaço das funções objectivo (b). Como se pode observar, todas as soluções do segmento QR (a que corresponde $Q'R'$ no espaço dos objectivos) optimizam a função objectivo $f_2(\mathbf{x})$, enquanto que apenas o ponto P (P' no espaço dos objectivos) optimiza $f_1(\mathbf{x})$. As soluções do segmento QR, com excepção do ponto Q, são fracamente eficientes. Estas soluções são dominadas por Q que tem valor igual em f_2 e superior em f_1, não havendo nenhuma solução admissível que melhore simultaneamente as duas funções objectivo relativamente a qualquer solução de $[RQ[$. As soluções (estritamente) eficientes deste problema são todos os pontos do segmento QP. No espaço das funções objectivo, os pontos não dominados são os do segmento $Q'P'$.

Note-se que apesar de neste exemplo, bem como no ilustrado na fig. 16.2, as soluções fracamente eficientes estarem sobre segmentos que optimizam uma das funções objectivo, isto não tem que acontecer necessariamente em problemas com mais de duas funções objectivo.

Solução ideal

Designa-se geralmente por solução ideal $\mathbf{z}^*$ (ou ponto utopia) a solução que optimizaria simultaneamente todas as funções objectivo, ou seja, cujas componentes são o óptimo de cada função objectivo na região admissível, quando optimizadas separadamente. Em geral a solução ideal não pertence à

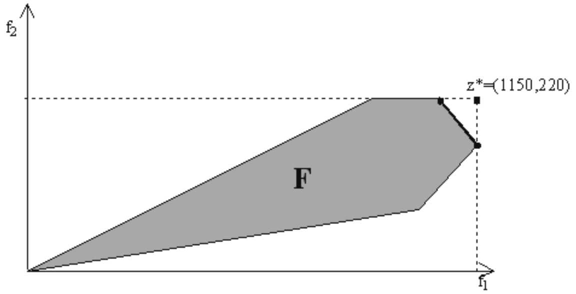


Figura 16.4: Solução ideal do Exemplo 1.

região admissível (caso contrário o problema seria trivial, pois todas as funções teriam o seu óptimo na solução ideal), embora cada uma das componentes z_k^* da solução ideal seja individualmente alcançável. Note-se que embora se possa definir sempre a solução ideal $\mathbf{z}^*$ no espaço dos objectivos, nem sempre existe a respectiva imagem no espaço de decisão (ou seja, pode não existir $\mathbf{x}^*$, mesmo que não admissível, tal que $\mathbf{z}^* = \mathbf{f}(\mathbf{x}^*)$).

A atitude de procurar uma solução o mais próxima possível da solução ideal está na base de processos de cálculos de soluções eficientes. A solução ideal é muitas vezes usada como o ponto de referência (inatingível) do decisor em funções escalares que representam uma distância a minimizar para determinar uma solução eficiente de compromisso.

Na fig. 16.4 está representada a solução ideal para o problema do Exemplo 1.

2. Programação Linear Inteira e Programação Não Linear Multiobjectivo

A consideração de variáveis inteiras, eventualmente apenas binárias, em modelos de programação linear multiobjectivo, bem como a existência de não linearidades nas funções objectivo e/ou nas restrições requer alguns conceitos adicionais.

A definição de soluções eficientes próprias comporta uma noção mais restrita de solução eficiente de modo a eliminar soluções eficientes que apresentem compromissos ilimitados entre objectivos, ou seja soluções em que a relação melhoria/degradação entre os valores das funções objectivo possa ser feita arbitrariamente grande (Geoffrion, 1968).

Definição 3 (solução eficiente própria): *Uma solução $\mathbf{x}' \in \mathcal{X}$ é eficiente própria se é eficiente e existe um número finito $M > 0$ tal que, para cada $\mathbf{x} \in \mathcal{X}$ e para cada função $f_k(\mathbf{x})$, $k = 1, \ldots, p$ com $f_k(\mathbf{x}) > f_k(\mathbf{x}')$, se verifica*

$$\frac{f_k(\mathbf{x}) - f_k(\mathbf{x}')}{f_j(\mathbf{x}') - f_j(\mathbf{x})} \leq M$$

para algum j em que $f_j(\mathbf{x}) < f_j(\mathbf{x}')$. $\mathcal{X}_{PE}$ representa o conjunto das soluções eficientes próprias.

Em programação linear multiobjectivo $\mathcal{X}_{PE} \equiv \mathcal{X}_E$. Também em programação linear inteira e inteira mista todas as soluções eficientes são eficientes próprias. No entanto, em problemas multiobjectivo não lineares podem existir soluções eficientes não próprias.

A fig. 16.5 ilustra o conceito de solução eficiente não própria a partir da representação no espaço dos objectivos de dois problemas não lineares, com ambas as funções a maximizar. Na fig.16.5(a) as soluções eficientes situam-se nos arcos AB e CD, excluindo o ponto D que é fracamente eficiente. As soluções A, B e C são eficientes não próprias. Na fig. 16.5(b), toda a fronteira de A a C é eficiente,

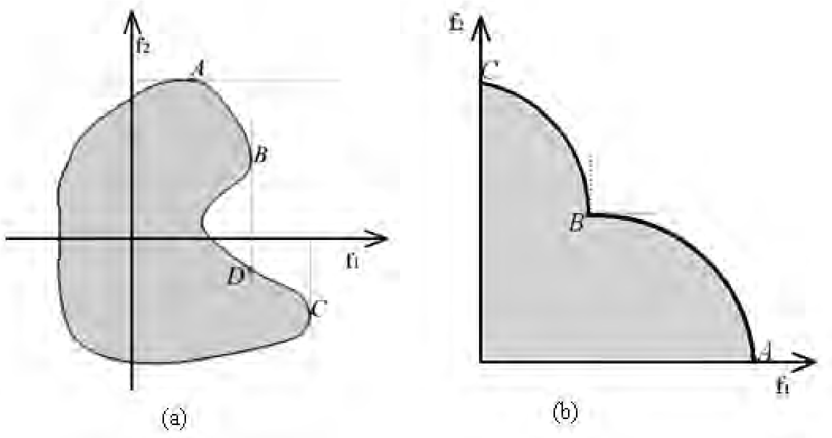


Figura 16.5: Ilustração do conceito de solução eficiente não própria

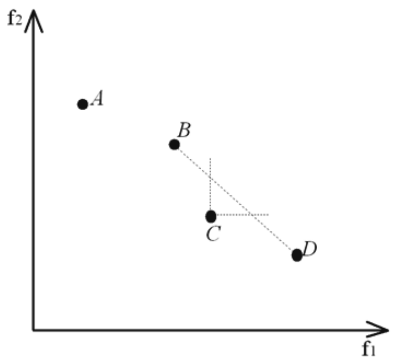


Figura 16.6: Ilustração do conceito de solução eficiente não suportada no caso discreto

passando pelo ponto B que é uma solução eficiente não própria.

Uma outra noção importante é a distinção entre soluções eficientes suportadas e não suportadas.

Definição 4 (solução não dominada/eficiente não suportada): *Um ponto não dominado* $\mathbf{z}' \in \mathcal{F}_E$ *é não suportado se for dominado por uma combinação convexa (não admissível) de pontos pertencentes a* $\mathcal{F}_E$. *A um ponto* $\mathbf{z}' = \mathbf{f}(\mathbf{x}')$ *não dominado não suportado corresponde uma solução* $\mathbf{x}$*' eficiente não suportada. As outras soluções eficientes dizem-se suportadas.*

Em problemas de programação linear multiobjectivo todas as soluções eficientes são suportadas, mas em problemas com $\mathcal{F}$ não convexo podem ocorrer soluções eficientes que sejam não suportadas. Ilustram-se em seguida os casos da programação linear inteira, inteira mista e não linear, multiobjectivo.

A fig. 16.6 ilustra a situação de existência de soluções eficientes não suportadas num problema de programação inteira, com ambas as funções a maximizar. As soluções A, B e D são eficientes suportadas, enquanto que C é uma solução eficiente não suportada porque é dominada por algumas combinações convexas - não admissíveis - de B e D (todas as definidas pela intersecção do cone que emana de C com o segmento que une B e D).

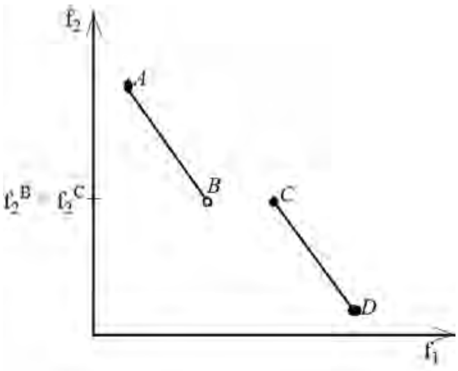


Figura 16.7: Soluções eficientes não suportadas num problema de programação inteira mista

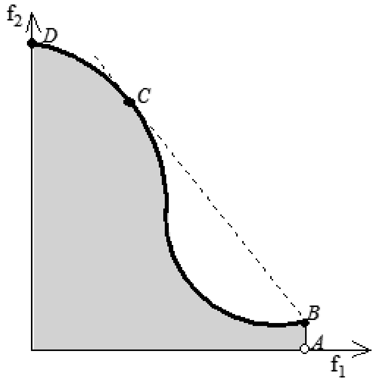


Figura 16.8: Soluções eficientes não suportadas num problema de programação não linear

A fig. 16.7 mostra a fronteira não dominada de um problema de programação linear inteira mista, com ambas as funções a maximizar. A solução A e todas as soluções do segmento CD são eficientes suportadas; as soluções do segmento AB, excluindo os pontos A e B, são eficientes não suportadas porque são dominadas por combinações convexas (não admissíveis) de A e C; a solução B é fracamente eficiente porque não existe nenhuma outra solução que seja estritamente melhor do que esta nas duas funções objectivo, mas é dominada pela solução C que tem valor igual em f_2 e valor superior em f_1.

A fig. 16.8 representa um problema não linear, em que a fronteira não dominada está representada a traço mais grosso (de B a D). As soluções no segmento AB, excluindo B, são soluções fracamente eficientes; as soluções desde B (exclusive) até a C (exclusive) são eficientes não suportadas; a solução B e as soluções do arco CD são eficientes suportadas.

3. Processos de Cálculo de Soluções Eficientes em Programação Linear Multiobjectivo

Nesta secção são abordadas três processos de cálculo de soluções eficientes através da optimização de funções escalares substitutas (funções escalarizantes), que agregam temporariamente numa única dimensão as p funções objectivo originais do modelo incluindo ainda parâmetros de informação das preferências do decisor. A solução óptima da função escalarizante deve ser uma solução eficiente (embora por vezes apenas possa ser garantida a obtenção de soluções fracamente eficientes) do problema multiobjectivo.

Existem diversos tipos de funções escalarizantes, às quais se requerem normalmente propriedades objectivas como sejam:

- Devem gerar apenas soluções eficientes;
- Devem poder gerar todas as soluções eficientes;
- Devem ser independentes de soluções não eficientes, ou seja, não devem ser sensíveis a alterações da região admissível não eficiente;

bem como propriedades subjectivas, tais como:

- O esforço computacional envolvido não deve ser muito grande (por exemplo, para um problema linear multiobjectivo a função substituta deve ser linear e a dimensão do problema não deve ser muito aumentada);
- Os parâmetros de preferência, quando existem, devem ter uma interpretação simples e não devem exigir do decisor grande esforço cognitivo.

Qual o significado das funções escalarizantes? Dependendo do contexto, isto é, dos princípios subjacentes aos métodos onde são usadas, podem distinguir-se duas aproximações básicas para o significado das funções escalarizantes:

- Constituirem apenas um mero artifício técnico para agregar temporariamente as múltiplas funções objectivo e gerar soluções eficientes a propor ao decisor (sem preocupação em reflectirem uma expressão das suas preferências);
- Serem encaradas como a representação analítica das preferências do decisor.

A apresentação das formas de escalarização mais comuns procura colocar em relevo as relações existentes entre a formulação matemática da função escalar substituta e a intervenção de parâmetros de preferência.

Optimização de uma das funções objectivo considerando as outras $p-1$ como restrições

O processo mais simples de construir um problema escalar é escolher como função substituta a optimizar uma das funções objectivo (normalmente aquela a que o decisor atribui mais importância), enquanto se estabelecem limitações inferiores nas outras $p-1$ funções objectivo que são tratadas como restrições (representando os níveis mínimos que o decisor está disposto a aceitar):

$$\max z_i = f_i(\mathbf{x}) = \mathbf{c}_i\mathbf{x}$$
$$\text{sujeito a: } \begin{cases} f_k(\mathbf{x}) = \mathbf{c}_k\mathbf{x} \geq e_k \quad k = 1, \ldots, p, \; k \neq i \\ \\ \mathbf{x} \in \mathcal{X} \end{cases} \tag{16.3}$$

Resolvendo (16.3) para algum i e variando os valores das limitações e_k é possível obter todo o conjunto das soluções eficientes.

Demonstra-se que se $\mathbf{x}' \in \mathcal{X}$ for solução óptima única de (16.3) para algum i, com e_k qualquer, então x' é uma solução eficiente do problema multiobjectivo (16.1) (Steuer, 1986).

A optimização desta função escalar substituta garante a obtenção de uma solução eficiente do problema multiobjectivo original (16.1) desde que a região admissível reduzida seja não vazia (o

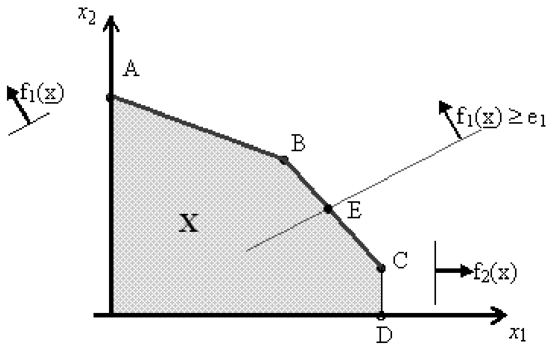


Figura 16.9: Optimização de uma função objectivo restringindo a outra gerando uma solução estritamente eficiente

que pode não acontecer se os níveis mínimos e_k forem demasiado exigentes) e não existam óptimos alternativos da função objectivo escolhida para ser optimizada (caso em que apenas há a garantia de obter pontos fracamente eficientes).

Note-se que, sem a condição de a solução ser única, pode existir um $\mathbf{x}" \in \mathcal{X}$ tal que $f_i(\mathbf{x}") = f_i(\mathbf{x}')$ e $f_k(\mathbf{x}") \geq f_k(\mathbf{x}')$, $k \neq i$, com desigualdade estrita para pelo menos um k, o que apenas garante que $\mathbf{x}'$ seja fracamente eficiente. Juntando um termo de perturbação à função objectivo do problema (16.3), isto é, substituindo $f_i(\mathbf{x}) = \mathbf{c}_i\mathbf{x}$ por

$$\mathbf{c}_i\mathbf{x} + \sum_{k=1,k\neq i}^{p} \rho_k\, \mathbf{c}_k\, \mathbf{x},$$

com $\rho_k > 0$ suficientemente pequenos, garante-se que a solução $\mathbf{x}'$ é (estritamente) eficiente para o problema (16.1).

A fig. 16.9 ilustra este processo de escalarização. A fronteira eficiente é constituída pelas soluções nos segmentos AB e BC. Limitando $f_1(\mathbf{x}) \geq e_1$ e optimizando $f_2(\mathbf{x})$ é obtida a solução eficiente E. Note-se que E é obviamente um vértice da região admissível reduzida, mas não é um vértice da região admissível original.

No caso de existirem soluções óptimas alternativas para (16.3) podem obter-se soluções que não são estritamente eficientes, mas apenas fracamente eficientes. Esta situação é ilustrada na fig. 16.10, em que impondo $f_1(\mathbf{x}) \geq e_1$ e optimizando $f_2(\mathbf{x})$ as soluções no segmento CF são óptimas alternativas para (16.3), mas apenas C é estritamente eficiente.

Um motivo de interesse adicional desta forma de escalarização é que a variável dual associada à restrição correspondente à função objectivo k pode ser interpretada como a taxa de compromisso local entre os objectivos $f_i(\mathbf{x})$ e $f_k(\mathbf{x})$ na solução óptima do problema escalar (16.3). No entanto, há que ter especiais cuidados na interpretação e uso desta informação em situações em que a solução seja degenerada, dado que neste caso estes valores não são únicos (óptimos alternativos do dual).

Embora esta forma de escalarização seja simples de compreender pelos decisores, captando a atitude de dar mais importância a uma função objectivo e aceitando limitações inferiores para as outras, a escolha da função objectivo a optimizar pode revelar-se difícil em muitos problemas. No quadro operacional de um dado método, a fixação da função a optimizar durante todo o processo torna o método pouco flexível e os resultados demasiado dependentes da função seleccionada.

Através do problema (16.3) é possível obter todos os pontos da fronteira não dominada, quer sejam ou não vértices da região admissível do problema original.

A informação de preferências associada a este tipo de escalarização consiste em:

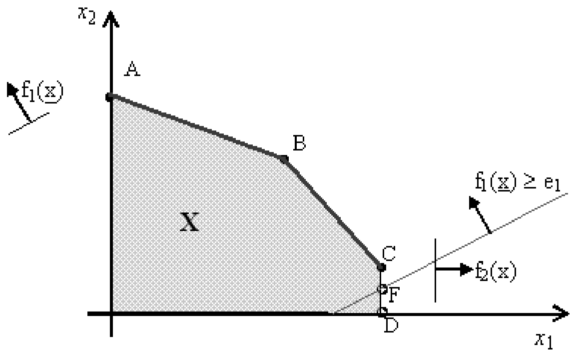


Figura 16.10: Optimização de uma função objectivo restringindo a outra gerando soluções fracamente eficientes

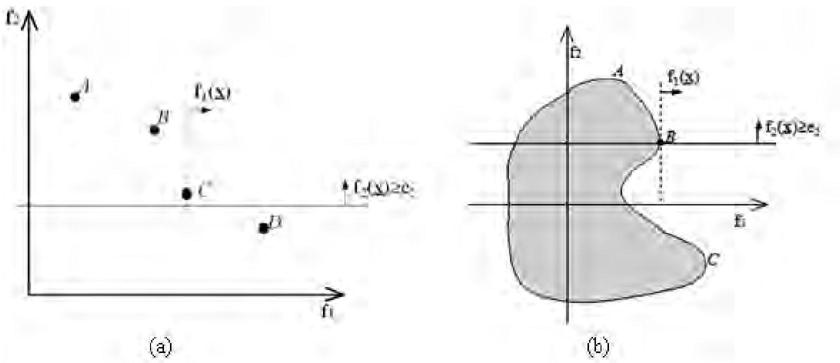


Figura 16.11: Optimização de uma função objectivo restringindo a outra: (a) num problema de programação inteira; (b) num problema não linear

- Informação inter-critérios: escolha da função objectivo a optimizar;
- Informação intra-critérios: imposição de níveis mínimos para as outras funções objectivo.

Este processo de cálculo de soluções eficientes pode também ser aplicado a problemas de programação inteira, inteira mista ou não linear multiobjectivo, permitindo obter qualquer tipo de solução eficiente para estes problemas.

A fig. 16.11 mostra exemplos de programação inteira (a) e de programação não linear (b) bi-objectivo, em que se restringe $f_2(\mathbf{x})$ e se optimiza $f_1(\mathbf{x})$; em (a) obtém-se a solução eficiente C (não suportada) e em (b) obtém-se a solução eficiente B (não própria).

Soma ponderada das funções objectivo

Um dos processos de cálculo de soluções eficientes mais utilizado em programação linear multiobjectivo consiste na resolução de um problema escalar cuja função objectivo é uma soma ponderada das p funções objectivo originais considerando pesos λ_k positivos:

$$\max z_\lambda = \lambda_1 f_1(\mathbf{x}) + \lambda_2 f_2(\mathbf{x}) + \ldots + \lambda_p f_p(\mathbf{x}) = \sum_{k=1}^{p} \lambda_k \, \mathbf{c}_k \, \mathbf{x} \tag{16.4}$$

$$\text{sujeito a: } \mathbf{x} \in \mathcal{X}$$

O conjunto de pesos admissível é definido por

$$\Lambda^0 = \left\{ \boldsymbol{\lambda} \in \mathbb{R}^p : \lambda_k \geq 0, \; k = 1, 2, \ldots, p, \; \sum_{k=1}^{p} \lambda_k = 1 \right\}$$

e o seu interior por

$$\Lambda = \left\{ \boldsymbol{\lambda} \in \mathbb{R}^p : \lambda_k > 0, \; k = 1, 2, \ldots, p, \; \sum_{k=1}^{p} \lambda_k = 1 \right\}.$$

Fixando um vector de pesos obtém-se uma função linear escalar ponderada das p funções objectivo, a optimizar em $\mathcal{X}$.

Demonstra-se que, em programação linear multiobjectivo, $\mathbf{x}' \in \mathcal{X}$ é uma solução eficiente para (16.1) sse for uma solução óptima para (16.4), para um vector de pesos $\boldsymbol{\lambda} \in \Lambda$ (Steuer, 1986).

Note-se que se podem obter soluções eficientes considerando $\boldsymbol{\lambda} \in \Lambda^0$, com algum $\lambda_k = 0$ (em particular, as soluções que optimizam individualmente cada uma das funções objectivo). Contudo, nesta situação, e caso existam óptimos alternativos, podem existir soluções óptimas de (16.4) que são apenas soluções fracamente eficientes do problema multiobjectivo. Ou seja, $\mathbf{x}'$ é uma solução eficiente para (16.1) se for uma solução óptima para (16.4), com $\boldsymbol{\lambda} \in \Lambda^0$, e uma das condições se verificar:

1. $\boldsymbol{\lambda} \in \Lambda$, ou
2. $\mathbf{x}'$ é solução única de (16.4) com $\boldsymbol{\lambda} \in \Lambda^0$.

A fig. 16.12 (correspondente ao Exemplo 1 apresentado atrás) ilustra este processo de escalarização. Optimizando a soma ponderada das funções objectivo cujo vector gradiente está representado na figura (z_λ) obtém-se a solução Q. Repare-se que, neste exemplo, o uso de peso nulo para $f_1(\mathbf{x})$, isto é, $\lambda_1 = 0$ e $\lambda_2 = 1$, conduz a soluções óptimas alternativas para o problema (16.4) entre as quais existem soluções fracamente eficientes. Esta situação corresponde à optimização de apenas $f_2(\mathbf{x})$. Como se observou atrás, todas as soluções do segmento QR optimizam esta função mas apenas o ponto Q é estritamente eficiente. As outras soluções deste segmento são fracamente eficientes.

O cálculo de soluções fracamente eficientes pode ser evitado substituindo o(s) peso(s) nulo(s) por um valor positivo ϵ bastante pequeno (por exemplo, 10^{-3} ou 10^{-4}). No exemplo representado na fig. 16.12, se for optimizada a soma ponderada $0.001\, f_1(\mathbf{x}) + 0.999\, f_2(\mathbf{x})$ obtém-se apenas a solução eficiente Q.

Da resolução do problema escalar (16.4) para um dado vector de pesos obtém-se sempre pelo menos uma solução básica (vértice) do problema multiobjectivo original, existindo vários vectores de pesos que conduzem à mesma solução básica. No caso do exemplo anterior, todos os vectores de pesos $\boldsymbol{\lambda}$ de $(0, 1)$ a $(\frac{3}{8}, \frac{5}{8})$ conduzem ao vértice Q e os vectores de pesos de $(\frac{3}{8}, \frac{5}{8})$ a $(1, 0)$ conduzem ao vértice P. Para os pesos $\lambda_1 = \frac{3}{8}$ e $\lambda_2 = \frac{5}{8}$ obtêm-se simultaneamente os pontos P e Q, visto que a optimização da soma ponderada das funções objectivo com este vector de pesos particular conduz ao segmento PQ (óptimos alternativos).

É útil analisar este processo de escalarização usando o quadro simplex multiobjectivo associado à função escalar soma ponderada que é optimizada na região admissível original. O quadro simplex multiobjectivo inicial (obtido acrescentando ao quadro simplex normal uma linha por cada função objectivo) tem a forma:

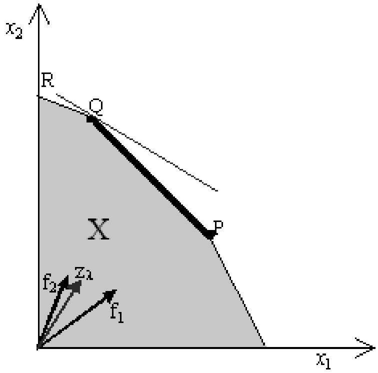


Figura 16.12: Optimização de uma soma ponderada das funções objectivo

$$\begin{array}{c|c} A & \mathbf{b} \\ \hline -C & 0 \end{array}$$

Relativamente à base B este quadro pode ser transformado em:

$$\begin{array}{c|c|c} \mathbf{x}_N & \mathbf{x}_B & \\ \hline B^{-1}N & I & B^{-1}\mathbf{b} \\ \hline C_B B^{-1}N - C_N & \mathbf{0} & C_B B^{-1}\mathbf{b} \end{array}$$

onde B e N, C_B e C_N são as sub-matrizes de A e de C correspondentes às variáveis básicas ($\mathbf{x}_B$) e não básicas ($\mathbf{x}_N$), respectivamente.

Definição 5 (bases adjacentes): *As bases B_1 e B_2 são adjacentes sse podem ser obtidas uma a partir da outra por uma operação de pivotação.*

Definição 6 (matriz dos custos reduzidos): *A matriz dos custos reduzidos relativa à base B é dada por $W = C_B B^{-1}N - C_N$.*

Definição 7 (base eficiente): *B é uma base eficiente sse for uma base óptima do problema escalar ponderado (16.4), para algum vector de pesos $\boldsymbol{\lambda} \in \Lambda$, isto é, B é uma base eficiente sse o sistema $\{\boldsymbol{\lambda}^T W \geq 0,\ \boldsymbol{\lambda} \in \Lambda\}$ for coerente.*

Definição 8 (variável não básica eficiente): *A variável não básica x_j é eficiente em relação à base B sse existir $\boldsymbol{\lambda} \in \Lambda$ tal que*

$$\boldsymbol{\lambda}^T W \geq \mathbf{0}$$
$$\boldsymbol{\lambda}^T W_{\cdot,j} = 0$$

onde $W_{\cdot,j}$ é a coluna de W correspondente a x_j (ou seja, o custo reduzido de x_j pode ser feito nulo).

A definição 8 significa que, para uma dada base eficiente, se x_j for uma variável não básica eficiente, qualquer pivotação admissível (cujo elemento pivot seja positivo, ou negativo correspondente a uma variável básica degenerada) associada a x_j conduz a uma base eficiente adjacente (ou seja, obtida a partir da anterior por essa operação de pivotação). Se a pivotação que conduz de uma base a outra for não degenerada, os pontos extremos correspondentes a essas bases são diferentes e a aresta que os liga é constituída por soluções eficientes.

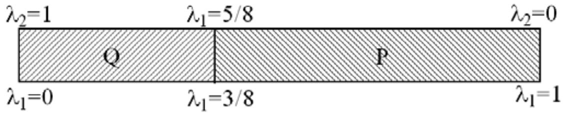


Figura 16.13: O diagrama paramétrico para um problema com 2 funções objectivo (exemplo 1)

Como as bases (vértices) eficientes são conexas é possível construir um método simplex multiobjectivo como extensão do método simplex (ver, por exemplo, (Steuer, 1986)), com o auxílio de sub-problemas de teste de eficiência de variáveis não básicas, que calcula todas as bases (vértices) eficientes, podendo a partir desta informação caracterizar também arestas eficientes não limitadas e faces eficientes (de diferentes dimensões).

A representação gráfica do conjunto de pesos $\boldsymbol{\lambda}$ que conduz a uma solução básica eficiente pode ser obtida através da decomposição do diagrama paramétrico (espaço dos pesos) Λ. Note-se que, devido à condição $\lambda_1+\lambda_2+\ldots+\lambda_p=1$, Λ pode ser representado num diagrama de dimensão $p-1$. Na fig. 16.13 é representado o diagrama paramétrico (unidimensional) de um problema com duas funções objectivo – o Exemplo 1 apresentado atrás (ver fig. 16.12) – na situação em que são conhecidas todas as bases eficientes (o diagrama paramétrico está totalmente preenchido). Na fig. 16.14, o diagrama paramétrico (bidimensional) é relativo a um problema com três funções objectivo e apenas são conhecidas as soluções básicas eficientes que optimizam individualmente cada função objectivo.

A partir do quadro simplex multiobjectivo correspondente a uma solução básica eficiente para (16.1), o conjunto de pesos correspondente é definido por $\left\{\boldsymbol{\lambda}^T W \geq \mathbf{0}, \boldsymbol{\lambda} \in \Lambda\right\}$. O elemento w_{kj} da matriz dos custos reduzidos W representa a variação da função objectivo $f_k(\mathbf{x})$ devido ao incremento unitário da variável não básica x_j que se torna variável básica. Cada coluna de W correspondente a uma variável não básica eficiente representa uma tendência de variação unitária das funções objectivo ao longo da respectiva aresta eficiente.

A região que engloba o conjunto de pesos correspondente a uma solução básica eficiente (região onde $\left\{\boldsymbol{\lambda}^T W \geq \mathbf{0},\ \boldsymbol{\lambda} \in \Lambda\right\}$) designa-se por *região de indiferença*. O decisor pode ser indiferente a todas as combinações de pesos nessa região, porque elas conduzem à mesma solução eficiente (figs. 16.13 e 16.14). Uma fronteira comum a duas regiões de indiferença significa que as respectivas soluções básicas eficientes estão ligadas por uma aresta eficiente, correspondente a uma variável não básica eficiente ser tornada variável básica. Se um ponto $\boldsymbol{\lambda} \in \Lambda$ pertence a várias regiões de indiferença, significa que estas correspondem a soluções eficientes localizadas na mesma face eficiente (esta face é apenas fracamente eficiente se esse ponto $\boldsymbol{\lambda}$ se situa na fronteira do diagrama paramétrico, ou seja $\boldsymbol{\lambda} \in \Lambda^0 \backslash \Lambda$).

A análise do diagrama paramétrico pode ser usada como um valioso utensílio na aprendizagem da "geometria" da região não dominada. As regiões de indiferença dependem das ordens de grandeza relativas dos valores das funções objectivo (ou seja, do comprimento relativo dos gradientes) e da geometria da região admissível. O tamanho da região de indiferença é de algum modo uma medida da robustez da solução face à variação dos pesos. Sempre que as diversas funções objectivo forem expressas em unidades com ordens de grandeza muito diferentes é aconselhável proceder à normalização das funções objectivo (o que altera a decomposição do diagrama paramétrico em termos dos espaços nele ocupados pelas regiões de indiferença).

Esta forma de escalarização é atractiva pela aparente facilidade de ser explicada aos decisores, captando a atitude de expressar o grau de importância que atribui a cada função objectivo. No entanto, esta informação, embora aparentemente simples, revela-se de grande dificuldade para o decisor e não há garantia de que as soluções encontradas resolvendo o problema (16.4) estejam de acordo com as preferências subjacentes à especificação dos pesos. Por exemplo, é possível obter a solução eficiente que optimiza $f_k(\mathbf{x})$ com $\lambda_k = 0$.

Através da resolução do problema (16.4) usando o método simplex só é possível obter vértices

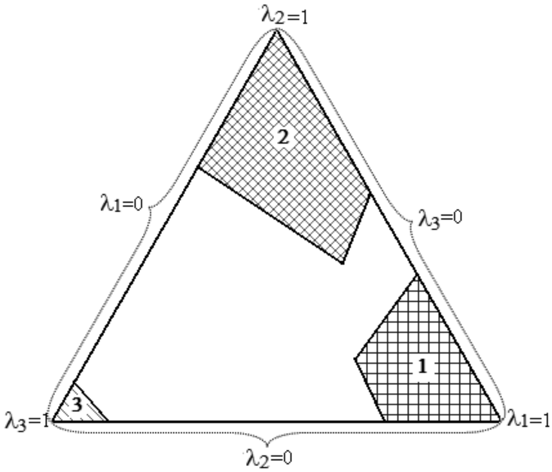


Figura 16.14: O diagrama paramétrico (para um problema com 3 funções objectivo) mostrando as regiões de indiferença correspondentes à solução que optimiza cada função objectivo

(soluções básicas) da fronteira não dominada.

A informação de preferências associada a este tipo de escalarização consiste em:

- Informação inter-critérios: escolhendo coeficientes de importância relativa (pesos) para as funções objectivo.

Este processo de cálculo de soluções eficientes pode também ser aplicado a problemas de programação inteira, inteira mista ou não linear multiobjectivo, mas não permite obter soluções eficientes não suportadas. A fig. 16.15(a) ilustra esta situação com um caso de programação inteira. Os pontos B e D são soluções óptimas alternativas da função objectivo ponderada cujo gradiente está representado na figura (z_λ). Um ligeiro aumento no peso atribuído a f_1 conduz apenas ao ponto D e um ligeiro aumento no peso de f_2 conduz apenas ao ponto B, não existindo nenhum vector de pesos que permita alcançar o ponto não dominado C (não suportado).

Num problema (não linear) em que o domínio das soluções admissíveis é convexo e as funções objectivo são côncavas é possível calcular-se todas as soluções eficientes próprias utilizando pesos estritamente positivos. Se o problema possuir soluções eficientes não próprias, estas podem também ser obtidas a partir da optimização de somas pesadas desde que se admita que os pesos podem ser nulos. A fig. 16.15(b) ilustra este caso, em que os pontos A e B só podem ser obtidos se se considerar um dos pesos igual a zero.

Minimização da distância a um ponto de referência

Uma solução de compromisso potencialmente interessante para o decisor é a que minimiza um dado tipo de distância à solução ideal, ou a outro ponto de referência do decisor que represente os níveis de cada função que gostaria de atingir (mesmo que não admissíveis). O que se pretende é então calcular uma solução eficiente admissível que esteja tão perto quanto possível (segundo uma dada métrica) das aspirações do decisor.

A fig. 16.16(a) mostra os contornos de isodistância em relação a $\mathbf{z}^*$ para as métricas L_1 (*city block*), L_2 (Euclideana) e L_∞ (de Chebyshev). Na fig. 16.16(b), os pontos $\mathbf{z}^1$, $\mathbf{z}^2$ e $\mathbf{z}^\infty$ minimizam as distâncias à solução ideal $\mathbf{z}^*$, usando as métricas L_1, L_2 e L_∞, respectivamente.

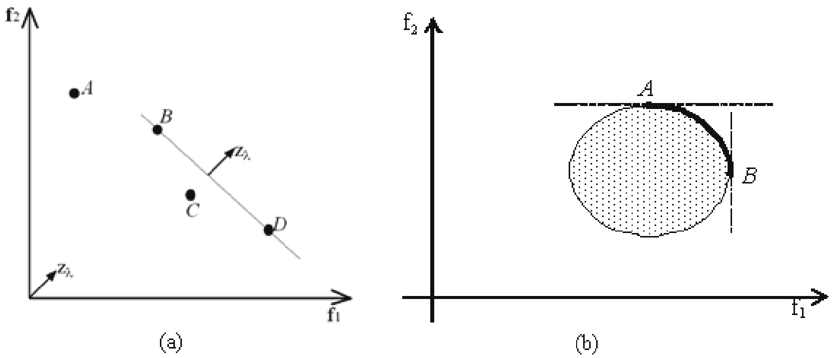


Figura 16.15: Optimização de somas ponderadas das funções objectivo (a) num problema de programação inteira e (b) num problema não linear convexo

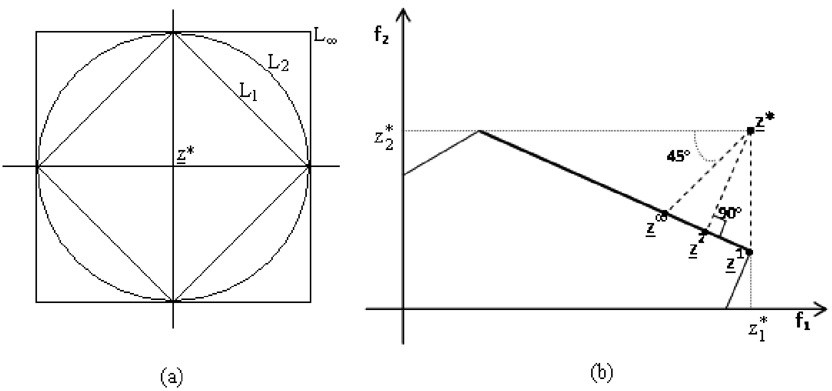


Figura 16.16: Métricas L_1, L_2 e L_∞: (a) contornos de isodistância, e (b) uso em funções distância.

Usando uma métrica L_β e tomando a solução ideal como ponto de referência o problema consiste em

$$\begin{aligned} &\min \|\mathbf{z}^* - \mathbf{f}(\mathbf{x})\|_\beta \\ &\text{s. a: } \mathbf{x} \in \mathcal{X} \end{aligned} \tag{16.5}$$

ou, considerando uma métrica L_β ponderada, com $\boldsymbol{\lambda} \geq \mathbf{0}$

$$\begin{aligned} &\min \boldsymbol{\lambda}\|\mathbf{z}^* - \mathbf{f}(\mathbf{x})\|_\beta \\ &\text{s. a: } \mathbf{x} \in \mathcal{X} \end{aligned} \tag{16.6}$$

Para $\beta = 1$ todos os desvios do ponto de referência são tomados em consideração na proporção directa da sua grandeza, enquanto para $2 < \beta < \infty$, maiores desvios vão tendo cada vez mais importância, até que para $\beta = \infty$ apenas o maior desvio é tido em conta.

Seja $\mathbf{x}^0$ solução do problema (16.5). Considerando o pior caso, isto é, a maior diferença entre as componentes dos vectores $\mathbf{z}^*$ e $\mathbf{f}(\mathbf{x}^0)$ através da métrica L_∞ (de Chebyshev) o problema

$$\min_{\mathbf{x}\in\mathcal{X}} \left\{ \max_{k=1,\ldots,p} (z_k^* - f_k(\mathbf{x})) \right\} \tag{16.7}$$

tem como solução $\mathbf{x}^0$ sse $\mathbf{x}^0$ e v^0 forem solução do problema (16.8)

$$\begin{aligned} &\min \quad v \\ &\text{s. a: } \begin{cases} v \geq z_k^* - f_k(\mathbf{x}) \quad ; \quad k = 1, \ldots, p \\ \mathbf{x} \in \mathcal{X} \\ v \geq 0 \end{cases} \end{aligned} \tag{16.8}$$

De forma mais genérica pode considerar-se uma métrica L_∞ ponderada (L_∞^λ):

$$\min_{\mathbf{x}\in\mathcal{X}} \left\{ \max_{k=1,\ldots,p} \lambda_k (z_k^* - f_k(\mathbf{x})) \right\} \tag{16.9}$$

O problema escalar (16.7) é um caso particular deste, em que todos os pesos λ_k são iguais a 1.

Note-se que, dado um problema de programação linear multiobjectivo, apenas resultam problemas escalares lineares usando as métricas L_1 ou L_∞.

Para garantir que as soluções obtidas por meio de (16.9) são estritamente eficientes (e não apenas fracamente eficientes) pode usar-se a métrica de Chebyshev (ponderada) aumentada ($L_\infty^{\lambda\epsilon}$):

$$\min_{\mathbf{x}\in\mathcal{X}} \left(\left\{ \max_{k=1,\ldots,p} \lambda_k (z_k^* - f_k(\mathbf{x})) \right\} + \rho \sum_{k=1}^{p} (z_k^* - f_k(\mathbf{x})) \right) \tag{16.10}$$

O segundo termo em (16.10) é uma perturbação acrescentada ao problema min-max normal, com ρ um escalar positivo suficientemente pequeno, que se destina a assegurar a eficiência estrita da solução.

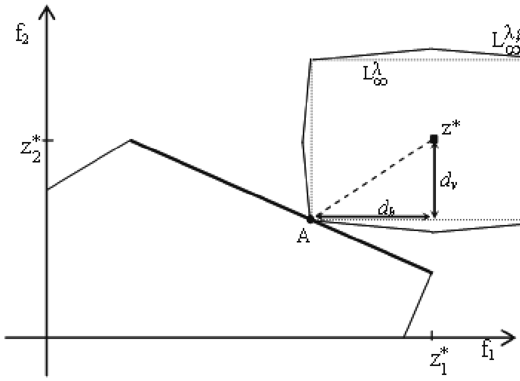


Figura 16.17: Minimização da distância de Chebyshev ponderada e aumentada à solução ideal

O problema escalar (16.10) pode escrever-se de forma equivalente

$$\min \quad v + \rho \sum_{k=1}^{p} (z_k^* - f_k(\mathbf{x}))$$

$$\text{s. a:} \begin{cases} v \geq \lambda_k (z_k^* - f_k(\mathbf{x})) \quad k = 1, \ldots, p \\ \mathbf{x} \in \mathcal{X} \\ v \geq 0 \end{cases} \tag{16.11}$$

Note-se que a função objectivo do problema (16.11) pode ser substituída por $v - \rho \sum_{k=1}^{p} f_k(\mathbf{x})$ visto que o termo restante é constante.

A minimização da distância à solução ideal segundo a métrica de Chebyshev ponderada (L_∞^λ) ou ponderada e aumentada ($L_\infty^{\lambda\epsilon}$) é ilustrada na fig. 16.17. O ponto A é a solução que minimiza a distância a $\mathbf{z}^*$ segundo (L_∞^λ) ou ($L_\infty^{\lambda\epsilon}$) considerando um vector particular de pesos $\boldsymbol{\lambda}$ em que $\lambda_1 < \lambda_2$. O vector $\boldsymbol{\lambda}$ determina a direcção de projecção do ponto de referência $\mathbf{z}^*$ na região não dominada, tal que $d_v/d_h = \lambda_1/\lambda_2$.

Demonstra-se que se $\mathbf{x}' \in \mathcal{X}$ é uma solução para o problema de Chebyshev ponderado aumentado (16.11) para algum $\boldsymbol{\lambda} \geq \mathbf{0}$, então $\mathbf{x}'$ é uma solução eficiente para o problema multiobjectivo original. Habitualmente usam-se pesos normalizados, i.e. $\boldsymbol{\lambda} \in \Lambda^0 = \left\{\boldsymbol{\lambda} \in \mathbb{R}^p : \ \lambda_k \geq 0, \ k = 1, \ldots, p, \ \sum_{k=1}^{p} \lambda_k = 1\right\}$.

Considerando $\boldsymbol{\lambda} \in \Lambda^0$ e a solução ideal deslocada $\mathbf{z}^{**} = \mathbf{z}^* + \boldsymbol{\epsilon}$ como ponto de referência (em que os ϵ_k são constantes arbitrárias estritamente positivas), (Steuer, 1986) prova que existe sempre um ρ positivo que garante que qualquer solução eficiente de um problema com região admissível poliédrica ou um conjunto discreto finito possa ser obtida a partir do problema escalar (16.11). Assim, para ρ positivo suficientemente pequeno, além da optimização de (16.11) ser condição suficiente de eficiência – o que é válido para o caso geral – é também condição necessária para os casos referidos, que incluem a programação linear multiobjectivo e a programação inteira multiobjectivo. A métrica pesada e aumentada de Chebyshev não permite, contudo, caracterizar por completo o conjunto das soluções eficientes de problemas com soluções eficientes não próprias. Para estes casos, (Steuer, 1986) propõe uma abordagem lexicográfica, que também pode ser aplicada aos outros casos, mas que é mais difícil de implementar pois são necessárias duas fases de optimização. Na primeira fase considera-se a métrica L_∞ (ponderada) não aumentada, ou seja, apenas se minimiza a variável v. Na segunda fase, minimiza-se $\sum_{k=1}^{p}(z_k^* - f_k(\mathbf{x}))$ no conjunto das soluções que minimizam v de modo a eliminar as soluções fracamente eficientes obtidas na primeira fase.

Nos problemas escalares anteriores é adoptada a solução ideal como ponto de referência, embora seja possível considerar outros pontos de referência (em geral, níveis de aspiração para cada função objectivo especificados pelo decisor). Se o ponto de referência usado não verificar $\mathbf{z}^r \geq \mathbf{f}(\mathbf{x})$ para todo o $\mathbf{x} \in \mathcal{X}$, então a variável v no problema escalar deve ser considerada sem restrição de sinal ($v \in \mathbb{R}$). Nestes casos poderá não se tratar da minimização de uma distância, já que $\mathbf{z}^r$ pode atravessar a fronteira não dominada e tornar-se atingível, deixando de se poder falar em métrica de Chebyshev. No entanto, garante-se igualmente que se obtém uma solução pelo menos fracamente eficiente no programa escalar não aumentado e estritamente eficiente no programa escalar aumentado. Estes programas minimizam funções que podem ser designadas genericamente por funções escalarizantes de *realização* (*achievement scalarizing functions* segundo (Lewandowski e Wierzbicki, 1988)).

A minimização da distância à solução ideal (ou a outro ponto de referência que represente níveis de aspiração para o decisor) pode ser interpretada como a minimização do desconforto de obter uma solução não dominada de compromisso $\mathbf{z}^0$ em vez do ponto ideal $\mathbf{z}^*$.

Através do problema (16.11) é possível obter todos os pontos da fronteira não dominada de um problema de programação linear multiobjectivo, quer sejam ou não vértices da região admissível.

A informação de preferências associada a este tipo de escalarização consiste em:

- Informação intra-critérios: estabelecimento de pontos de referência;
- Informação inter-critérios: coeficientes de ponderação da métrica L_∞.

4. Métodos multiobjectivo de apoio à decisão

Embora não seja fácil categorizar os diversos tipos de métodos de apoio à decisão multiobjectivo propostos na literatura, uma das classificações mais geralmente adoptadas baseia-se na fase do processo de decisão em que é requerida a intervenção do decisor. Assim, podem distinguir-se métodos em que:

(a) Não há articulação de preferências do decisor (métodos geradores). No caso de problemas de PLMO são geradas todas as soluções extremas não dominadas (e eventualmente todas as faces não dominadas de dimensão máxima) que são apresentadas ao decisor para escolha da solução preferida, face à informação gerada pelas sucessivas etapas de cálculo (que não requerem qualquer tipo de intervenção por parte do decisor)[1].

(b) A articulação de preferências do decisor é feita:

(b1) *a priori* (onde se destacam os métodos de função valor, ou utilidade) – inicialmente é obtida do decisor informação de preferências que serve para construir uma função valor que agrega numa única dimensão todos os critérios explicitamente considerados no modelo. A construção desta função valor (ou utilidade, quando há probabilidades associadas aos resultados de cada acção potencial, embora esta distinção seja muitas vezes ignorada na literatura) tem *tanto de arte como de ciência* (Keeney e Raiffa, 1976; Yu, 1985, cap. 6). Ou seja, apesar de existir modelação multicritério, o problema é depois reduzido ao caso monocritério e a optimização desta função valor conduz à solução de compromisso óptima.

(b2) progressivamente (métodos interactivos) – etapas de cálculo são alternadas com etapas de diálogo, em que o decisor é chamado a expressar as suas preferências perante soluções

[1] Neste caso, como a intervenção do decisor ocorre somente após a geração de todo o conjunto de soluções não-dominadas, também se diz que as preferências são informadas *a posteriori*.

que lhe são propostas, até se atingir uma condição de paragem (variável de método para método) que pode pressupor ou não a existência de uma função valor implícita do decisor. A assunção da existência implícita de uma função valor do decisor significa que este é incapaz de especificar a forma exacta da função, mas está em condições de fornecer informação local sobre as suas preferências, coerente com essa função, face a questões colocadas pelo método.

A utilização de métodos geradores em problemas complexos do mundo real revela-se proibitiva. Por um lado, o esforço computacional para calcular todos os pontos extremos não dominados é demasiado elevado mesmo para problemas lineares de dimensão média. Por outro, a quantidade de resultados gerada (sem qualquer intervenção do decisor no processo) é por vezes enorme, sem que isso se traduza num aumento da qualidade da informação colocada à disposição do decisor.

Os métodos que assumem a existência de uma função valor (a ser optimizada) reflectindo as preferências de um decisor racional têm sido sujeitos a muitas críticas, sobretudo com base em estudos empíricos que demonstraram que os decisores não são sempre coerentes com as funções valor destinadas a agregar os critérios e representar as suas preferências, agindo geralmente mais de acordo com a sua própria maneira de ver o problema do que procurando conciliar esta com os axiomas que são a justificação teórica da existência da função valor (Bell e Farquhar, 1986).

Os métodos interactivos, em que existe uma articulação progressiva das preferências do decisor ao longo do processo interactivo, são os mais adequados para constituírem o núcleo de sistemas de apoio à decisão para problemas multiobjectivo, pela capacidade em incorporar aprendizagem e adaptação na expressão de preferências ao longo do processo de decisão, onde o decisor é chamado a desempenhar um papel activo. Os métodos interactivos alternam etapas de cálculo de soluções não dominadas com etapas de diálogo, em que a intervenção do decisor é usada para guiar o processo interactivo de decisão, reduzindo o âmbito da pesquisa e o esforço computacional associado.

Podem distinguir-se duas implementações de metodologias interactivas resultantes de duas concepções básicas para o papel da interactividade:

(a) Orientada para a procura. O decisor age de acordo com uma estrutura de preferências pré-existente e estável (representada por uma função valor implícita), e permanece coerente com ela ao longo da utilização da metodologia, respondendo de forma determinística às questões colocadas. Ou seja, o objectivo da interacção é a procura de uma proposta óptima (que não depende da evolução do processo interactivo) face à estrutura de preferências, que é assim descoberta durante o processo. Nestas circunstâncias parece razoável impor a convergência matemática do método interactivo. A convergência é garantida e controlada pelo método.

(b) Orientada para a aprendizagem. É recusada a assunção de que existe uma estrutura de preferências pré-existente e estável com a qual o decisor é coerente. As preferências do decisor podem ser parcialmente instáveis e evolutivas. O objectivo da interacção é a aprendizagem das preferências, ou seja, a clarificação do que, na perspectiva do decisor, possa ser uma solução satisfatória. A convergência não é garantida e é controlada pelo decisor, fazendo então sentido falar em convergência psicológica (significando a identificação de uma solução eficiente como satisfatória graças à informação entretanto recolhida). O processo termina com uma solução de compromisso satisfatória de acordo com a informação disponível e a vontade do decisor quando não parecer possível a emergência de novas intuições sobre o problema.

Face aos métodos geradores, os métodos interactivos revelam vantagens inequívocas quer ao reduzir o esforço computacional, quer ao limitar o esforço cognitivo imposto ao decisor. Face aos métodos baseados em funções valor (ou utilidade) os métodos interactivos evitam a obtenção prévia de informação de preferências para formular a função valor, que se supõe ser a sua representação analítica, permitindo a aprendizagem e a evolução de preferências à medida que vai sendo reunida informação ao longo do processo interactivo.

A interactividade significa oferecer ao decisor um ambiente operacional que suscite a exploração, a reflexão, a emergência de novas intuições, permitindo-lhe compreender com mais profundidade o problema de decisão em causa, contribuindo para moldar e fazer evoluir as suas preferências no sentido de guiar o processo de pesquisa, centrando-o nas regiões onde se localizam as soluções que julga mais interessantes, até encontrar uma solução de compromisso satisfatória ou chegar à conclusão que é necessário rever os dados iniciais e/ou o reformular o modelo matemático. A aprendizagem deve ser entendida não apenas no sentido do aumento do conhecimento disponível, mas sobretudo no do aperfeiçoamento das capacidades do decisor de modo a fazer um uso adequado desse conhecimento.

Os métodos interactivos tendem a ser vistos como a abordagem mais adequada para o estudo destes problemas: *...the future of multiple objective programming is in its interactive application* (Steuer, 1986, p. 386). Por outro lado, face à acessibilidade de elevadas capacidades de cálculo e potencialidades gráficas com custos cada vez mais acessíveis, encontra-se disponível a base tecnológica para a construção de sistemas de apoio à decisão flexíveis conjugando as características interactivas dos métodos multiobjectivo com adequados ambientes de interacção e meios gráficos de apresentação de informação.

Métodos interactivos em PLMO

Os métodos interactivos podem ser classificados de acordo com uma grande diversidade de parâmetros, não existindo uma categorização aceite de forma universal. Nesta secção são apresentadas algumas formas possíveis de categorização dos métodos interactivos de PLMO. A primeira baseia-se na estratégia de redução do âmbito da pesquisa (explícita ou implicitamente), segundo a qual se podem distinguir métodos de (Steuer, 1986; Clímaco et al., 2003):

- Redução progressiva da região admissível;
- Redução progressiva do diagrama paramétrico (espaço dos pesos);
- Contracção progressiva do cone dos gradientes das funções objectivo;
- Pesquisa direccional.

As três primeiras estratégias reduzem progressivamente a região eficiente que pode ser pesquisada em fases subsequentes.

Perante a diversidade de técnicas de cálculo de soluções eficientes, podem classificar-se os métodos de acordo com o tipo de função escalar substituta (função escalarizante) usada para agregar temporariamente os objectivos:

- Optimização de uma função objectivo restringindo as outras;
- Soma ponderada das funções objectivo;
- Minimização de uma distância a um ponto de referência.

A possibilidade de o decisor intervir para solicitar a realização de uma dada operação em qualquer momento do processo de decisão interactivo ou ser obrigado a uma sequência pré-determinada de fases de cálculo e de diálogo, permite classificar os métodos em:

- Não estruturados;
- Estruturados.

Estas classificações não são mutuamente exclusivas, dado que mesmo dentro de cada parâmetro de categorização há métodos que combinam diferentes estratégias de redução do âmbito da pesquisa e técnicas de cálculo de soluções eficientes. Por outro lado, estas classificações não são exaustivas, pois existem métodos que é difícil englobar em qualquer delas.

Seguidamente são brevemente descritos os métodos STEM, de Zionts e Wallenius, *Interval Criterion Weights*, *Pareto Race* e TRIMAP, considerados representativos das várias categorias anteriormente apresentadas. Estes métodos são detalhadamente descritos, com exemplos ilustrativos do respectivo funcionamento, em (Clímaco et al., 2003).

Método STEM

O *Step Method* (STEM), desenvolvido por Benayoun et al. (1971) é um método interactivo de redução da região admissível. Em cada fase de diálogo, após uma fase de cálculo, o decisor é chamado a especificar as quantidades que está disposto a sacrificar nas funções objectivo cujos valores considera satisfatórios de modo a melhorar as restantes. Em cada fase de cálculo é minimizada uma distância ponderada de Chebyshev à solução ideal.

Em cada iteração, o problema a optimizar reflecte as escolhas do decisor feitas em iterações precedentes, através da redução da região admissível. É apresentada ao decisor a solução de compromisso calculada em cada iteração minimizando uma distância ponderada de Chebyshev à solução ideal. Se os valores das funções objectivo são considerados satisfatórios, o processo termina; caso contrário, o decisor deve especificar quais pretende relaxar, e de que quantidade, por forma a melhorar os outros objectivos. A região admissível é então reduzida através das limitações nos valores das funções objectivo (calculadas com base nas quantidades de relaxamento introduzidas pelo decisor).

O método de Zionts e Wallenius

O método de Zionts e Wallenius (1976; 1983) reduz progressivamente o diagrama paramétrico (espaço dos pesos), de acordo com as preferências do decisor expressas através das respostas em cada interacção face a comparações entre pares de soluções e julgamentos sobre as tendências de variação unitária ao longo de arestas que têm origem na solução corrente e conduzem a outras soluções não dominadas. Em cada fase de cálculo é optimizada uma soma ponderada das funções objectivo.

Partindo das respostas dadas pelo decisor, o método introduz então restrições no diagrama paramétrico, reduzindo progressivamente o domínio admissível para a selecção de um novo vector de pesos. O processo termina quando o diagrama paramétrico for reduzido a uma região suficientemente pequena de forma a que seja identificada uma solução final, convergindo para o óptimo de uma função valor implícita não necessariamente linear do decisor (na realidade, para o vértice eficiente que garante o maior valor para a função), ou a informação de preferências expressa pelo decisor torne a solução actual mais interessante do que as que lhe são propostas para comparação. Parte-se do princípio que as respostas do decisor nas fases de diálogo são coerentes com essa função valor implícita, embora no caso de serem detectadas incoerências seja dada a possibilidade de eliminar as restrições mais antigas no diagrama paramétrico.

O método *Interval Criterion Weights*

O método *Interval Criterion Weights* (ICW), desenvolvido por Steuer (1977; 1986), é um procedimento interactivo que reduz progressivamente o cone dos critérios (cone convexo gerado pelos gradientes das funções objectivo). Esta redução é feita de acordo com as preferências manifestadas pelo decisor ao escolher a solução preferida numa amostra de soluções não dominadas que lhe é apresentada em cada fase de diálogo. Esta amostra é obtida por técnicas de filtragem a partir de um conjunto de soluções não dominadas, de modo a colocar à apreciação do decisor um número não muito elevado de soluções. Em cada fase de cálculo são optimizadas várias somas ponderadas das funções objectivo.

Por um lado, o método ICW usa várias combinações regularmente dispersas no diagrama paramétrico (espaço dos pesos) para definir um conjunto de somas ponderadas, evitando assim obter informação sobre as preferências do decisor através da indicação explícita de pesos. Por outro lado, este método também não faz apelo à especificação de valores satisfatórios para as funções objectivo usados para reduzir a região admissível (como, por exemplo, no STEM). A informação dada pelo decisor ao escolher a solução da amostra que corresponde mais de perto às suas preferências é usada para contrair o cone dos critérios em torno do gradiente da função objectivo soma ponderada que deu origem a essa solução, resultando o cone de critérios reduzido da iteração seguinte. O cone dos critérios é assim gradualmente contraído e deslocado até se focar numa pequena porção da superfície da região admissível, contendo o vértice eficiente com maior valor para a função valor implícita do decisor.

O método *Pareto Race*

O método *Pareto Race*, proposto por Korhonen e Wallenius (1988) baseado em trabalhos de Korhonen (1987) e Korhonen e Laakso (1986b; 1986a), é um método de pesquisa direccional livre, que permite ao decisor mover-se por qualquer ponto sobre a região eficiente. A informação requerida ao decisor consiste fundamentalmente na especificação das funções objectivo a melhorar alterando a direcção do movimento. As soluções eficientes são obtidas por meio da optimização de uma função escalarizante baseada em ponto de referência e programação paramétrica em relação aos termos independentes das restrições (metas não flexíveis).

A partir de níveis de aspiração para os valores das funções objectivo, especificados inicialmente pelo decisor, é construída uma direcção de referência (direcção que, partindo de um ponto admissível no espaço dos objectivos, oferece uma variação nos valores das funções objectivo que está de acordo com as preferências do decisor). Esta direcção de referência é então projectada sobre o conjunto das soluções eficientes, gerando uma trajectória (subconjunto das soluções eficientes) que é apresentada ao decisor. O decisor pode assim percorrer a fronteira eficiente, controlando a direcção do movimento (privilegiando diferentes funções objectivo) e a velocidade (obtendo soluções mais ou menos próximas umas das outras) como se estivesse a conduzir um automóvel (daí a denominação *Pareto Race*) sobre essa superfície.

O método TRIMAP

O método TRIMAP, desenvolvido por Clímaco e Antunes (1987; 1989), é um método de pesquisa livre baseado na aprendizagem progressiva e selectiva do conjunto das soluções não dominadas, que combina a redução da região admissível com a redução do diagrama paramétrico (espaço dos pesos). O decisor pode especificar limitações inferiores para os valores da funções objectivo e impor restrições nos pesos. Em cada fase de cálculo é optimizada uma soma ponderada das funções objectivo.

O TRIMAP combina três procedimentos fundamentais: decomposição do diagrama paramétrico, introdução de restrições no espaço dos objectivos e introdução de restrições nos pesos. As limitações introduzidas nos valores das funções objectivo podem ser traduzidas automaticamente para o diagrama paramétrico, que é usado como um valioso meio para recolher e apresentar a informação ao decisor. Esta possibilidade é muito útil dado que os decisores estão geralmente mais à vontade quando o diálogo é conduzido em termos dos valores das funções objectivo, que constitui o espaço que lhes é mais familiar. Inicialmente são calculadas as soluções eficientes que optimizam cada uma das funções objectivo, fornecendo ao decisor uma primeira informação sobre a gama de valores de cada objectivo, e a solução eficiente que minimiza uma distância ponderada de Chebyshev à solução ideal. Através da análise comparativa dos gráficos do diagrama paramétrico e do espaço dos objectivos, o decisor pode fazer uma cobertura progressiva e selectiva do diagrama paramétrico, decidindo em cada interacção o interesse de pesquisar soluções em regiões deste espaço ainda não exploradas. A redução do âmbito da pesquisa é feita principalmente impondo limitações nos valores das funções objectivo, que são traduzidas para o diagrama paramétrico.

O método TRIMAP será analisado mais detalhadamente na secção 6, descrevendo-se através de um exemplo ilustrativo como o método pode constituir um utensílio para apoio à decisão em problemas de programação linear com três funções objectivo.

5. O Método TRIMAP (programação linear TRIobjectivo - Método de Aprendizagem Progressiva das soluções não dominadas)

O método TRIMAP, desenvolvido por Clímaco e Antunes (1987; 1989), é um método de pesquisa livre onde se pretende proporcionar ao decisor uma aprendizagem progressiva e selectiva do conjunto das soluções não dominadas. O TRIMAP combina a redução da região admissível com a redução do diagrama paramétrico. O decisor pode especificar as suas preferências por meio do estabelecimento de limitações inferiores para os valores da funções objectivo e da imposição de restrições no diagrama paramétrico.

A finalidade essencial do método é a de ajudar o decisor a eliminar os subconjuntos de soluções não dominadas que não lhe interessam, e não a de assegurar a convergência para a solução de compromisso óptima de uma função utilidade implícita, que se supõe pré-existente ao processo de decisão. O processo interactivo termina quando o decisor considera conhecer o bastante sobre o conjunto das soluções não dominadas, sendo sempre possível rever decisões e julgamentos anteriores à medida que vai sendo reunido maior conhecimento sobre o problema. Ou seja, a informação de preferências do decisor não necessita ser coerente em sucessivas interacções. Usando a terminologia de (Roy, 1987), no TRIMAP a convergência (de uma função utilidade) deve dar lugar à criação, de modo que o processo interactivo seja um processo construtivo e não a determinação de algo pré-existente.

O TRIMAP combina três procedimentos fundamentais: decomposição do diagrama paramétrico, introdução de restrições no espaço dos objectivos e introdução de restrições no diagrama paramétrico. A característica mais inovadora do método é a possibilidade de traduzir as limitações introduzidas nos valores das funções objectivo para o diagrama paramétrico, que é usado como um meio coerente para recolher e apresentar a informação ao decisor. O TRIMAP é vocacionado para problemas com três funções objectivo, o que, embora constituindo uma limitação, permite o uso de meios gráficos adequados ao diálogo com o decisor. O principal objectivo é possibilitar ao decisor um preenchimento progressivo e selectivo do diagrama paramétrico, que lhe dê informação sobre a forma da superfície não dominada, evitando deste modo um estudo exaustivo de regiões onde os valores das funções objectivo sejam muito semelhantes (o que frequentemente acontece em casos reais).

A redução do âmbito da pesquisa é feita principalmente impondo limitações nos valores das funções objectivo (que é o tipo de informação que menos esforço exige da parte do decisor), que são traduzidas para o diagrama paramétrico. A introdução destas limitações adicionais pode ainda ser usada para obter soluções não dominadas que não sejam pontos extremos do poliedro admissível. É também possível impor restrições directamente no diagrama paramétrico. Através da análise comparativa do diagrama paramétrico e do espaço dos objectivos, o decisor pode fazer uma cobertura progressiva e selectiva do diagrama paramétrico, decidindo em cada interacção o interesse de pesquisar soluções em áreas do triângulo ainda não exploradas.

Inicialmente são calculadas as soluções eficientes que optimizam cada uma das funções objectivo, fornecendo ao decisor uma primeira informação sobre a gama de valores de cada objectivo. Como informação auxiliar é também calculada a solução eficiente (que não é, em geral, um vértice) que minimiza uma distância ponderada de Chebyshev à solução ideal, como no método STEM.

A selecção dos pesos para o cálculo de soluções eficientes, pode ser feita de duas formas:

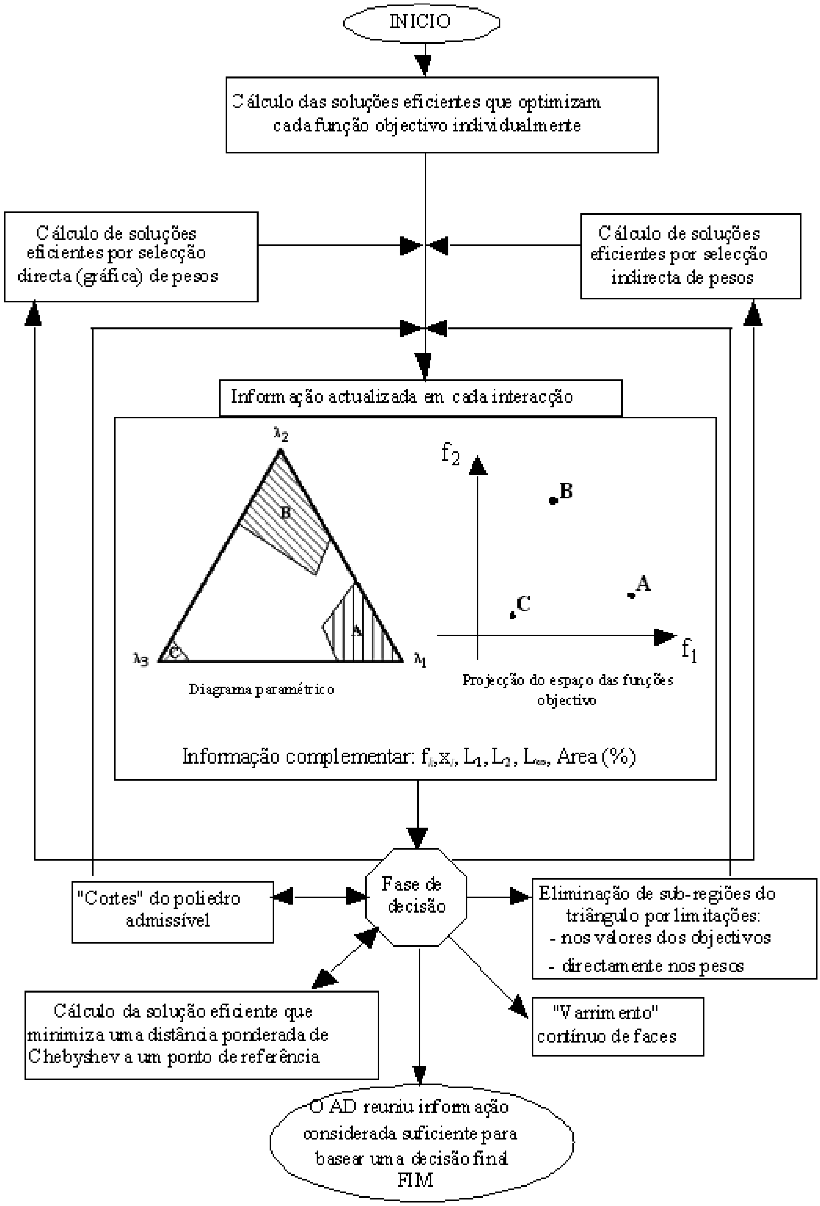


Figura 16.18: Fluxograma do método TRIMAP

- Directamente, em que o decisor escolhe um vector de pesos de uma zona do triângulo não preenchida, que lhe parece importante para continuar a pesquisa;
- Indirectamente, construindo uma função ponderada cujo gradiente é normal ao plano que passa por três soluções não dominadas já calculadas escolhidas pelo decisor. Se os pesos assim obtidos não forem todos não negativos, é feita uma pequena perturbação no gradiente da função objectivo ponderada por forma a assegurar essa condição.

A introdução de limitações adicionais nos valores das funções objectivo, e a respectiva tradução para o diagrama paramétrico, permite que o diálogo com o decisor seja feito em termos dos valores das funções objectivo acumulando a informação resultante no diagrama paramétrico. O estabelecimento da limitação adicional $f_k(\mathbf{x}) \geq e_k$, $(e_k \in \mathbb{R},\ k \in \{1,2,3\})$ conduz à construção do problema auxiliar

$$\begin{aligned} &\max f_k(\mathbf{x}) \\ &\text{s. a: } \mathbf{x} \in \mathcal{X}_a \equiv \{\mathbf{x} \in \mathcal{X} : f_k(\mathbf{x}) \leq e_k\} \end{aligned} \tag{16.12}$$

Maximizando $f_k(\mathbf{x})$ em $\mathcal{X}_a$ são obtidas soluções (básicas) óptimas alternativas (note-se que o gradiente da função a maximizar tem a mesma direcção do gradiente da restrição adicional). Os vértices eficientes do poliedro admissível $\mathcal{X}_a$ que optimizam (16.12) são seleccionados e as sub-regiões do diagrama paramétrico que correspondem a cada um destes pontos são calculadas e representadas graficamente. Estas sub-regiões do diagrama paramétrico são as regiões de indiferença definidas por $\left\{\boldsymbol{\lambda}^T W \geq 0,\ \boldsymbol{\lambda} \in \Lambda\right\}$, relativas a cada base eficiente alternativa. A união de todas estas regiões de indiferença determina a sub-região do diagrama paramétrico onde a limitação adicional no valor da função objectivo é satisfeita. Se o decisor estiver apenas interessado em soluções não dominadas que satisfaçam $f_k(\mathbf{x}) \geq e_k$, então é suficiente restringir a pesquisa a esta sub-região. Se o decisor pretender impor mais do que uma limitação então (16.12) é resolvido para cada uma delas e as sub-regiões correspondentes no diagrama paramétrico são preenchidas com diferentes padrões (ou cores) permitindo assim visualizar claramente as zonas onde existem intersecções. A introdução destas limitações adicionais pode também ser usada para obter soluções não dominadas que, em geral, não são vértices do poliedro admissível original.

É também possível eliminar regiões do diagrama paramétrico impondo limitações directamente na variação dos pesos (dos tipos $\lambda_i/\lambda_j \geq u_{ij},\ i,j \in \{1,2,3\},\ i \neq j,\ u_{ij} \in \mathbb{R}^+$, ou $0 < u_L \leq \lambda_k \leq u_H < 1$, com $k \in \{1,2,3\}$).

O TRIMAP permite ainda percorrer faces não dominadas, com uma interface semelhante ao do método *Pareto Race*, entre dois dos seus vértices previamente calculados. É mostrada uma linha que avança ou recua sobre a projecção da face a uma velocidade controlada pelo utilizador. Os valores das funções objectivo correspondentes aos pontos percorridos são dinamicamente apresentados num gráfico de barras.

O decisor pode identificar um ponto de referência no espaço dos objectivos e o programa calcula a solução não dominada que minimiza uma distância ponderada de Chebyshev a esse ponto.

Em cada interacção de TRIMAP são apresentados ao decisor dois gráficos principais. O primeiro é o diagrama paramétrico mostrando as regiões de indiferença correspondentes às soluções extremas não dominadas já conhecidas. O segundo é uma projecção do espaço dos objectivos mostrando as soluções não dominadas já calculadas. Estão também disponíveis indicadores complementares sobre cada solução: distâncias L_1, L_2 e L_∞ à solução ideal e a área da região de indiferença (percentagem ocupada da área total do triângulo). Outro tipo de informação gráfica disponível sobre cada solução é o gráfico em teia de aranha (*spider web*), também designado por mapas de radar (*radar charts*) ou estrutura em estrela (*star structures*), que constitui uma forma útil de condensar as várias dimensões de avaliação num único gráfico (Kasanen et al., 1991). O gráfico em teia de aranha usado em TRIMAP considera o ponto ideal (ou outro ponto de referência) no centro do gráfico, fornecendo uma impressão visual da diminuição de distâncias necessária para atingir esse ponto.

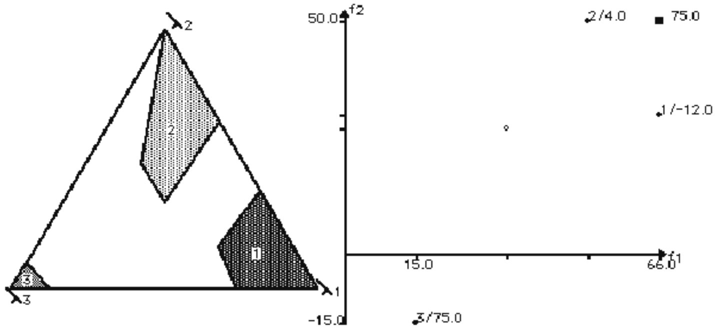


Figura 16.19: O diagrama paramétrico e uma projecção do espaço dos objectivos após a optimização individual de f_1, f_2 e f_3

Exemplo ilustrativo da aplicação do TRIMAP

Para ilustrar a aplicação de TRIMAP a problemas lineares multiobjectivo, consideremos o seguinte problema, com quatro variáveis de decisão, três funções objectivo a maximizar e três restrições:

$$\text{Max} \begin{bmatrix} 3 & 1 & 2 & 1 \\ 1 & -1 & 2 & 4 \\ -1 & 5 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

$$\text{s. a:} \begin{bmatrix} 2 & 1 & 4 & 3 \\ 3 & 4 & 1 & 2 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \leq \begin{bmatrix} 60 \\ 60 \\ 50 \end{bmatrix} \qquad \mathbf{x} \geq \mathbf{0}\,, \quad \mathbf{x} \in \mathbb{R}^4$$

x_5, x_6, x_7 são variáveis folga (*slacks*) correspondentes às restrições 1, 2 e 3, respectivamente.

O TRIMAP começa por calcular as soluções não dominadas que optimizam separadamente cada função objectivo. A fig. 16.19 mostra o diagrama paramétrico com as regiões de indiferença correspondentes às soluções que optimizam f_1, f_2 e f_3 (soluções 1, 2 e 3, respectivamente), e a projecção ortogonal em $f_1 \times f_2$ do espaço dos objectivos. O valor a seguir à identificação da solução representa a função que é projectada (f_3 neste caso). O quadrado preto representa a solução ideal. O círculo representa a solução que minimiza uma distância ponderada de Chebyshev a um ponto de referência (inicialmente a solução ideal).

A análise do diagrama paramétrico revela-nos que existe pelo menos outra solução alternativa à solução 2 que também optimiza f_2. Colocando o cursor do rato na zona do diagrama paramétrico à esquerda da região de indiferença da solução 2, e escolhendo um vector de pesos para construir uma soma ponderada das funções objectivo encontramos a solução 4 (fig. 16.20) que é um óptimo alternativo da solução 2 em relação a f_2. A aresta não dominada que liga estas duas soluções extremas não dominadas é também já conhecida.

A tabela seguinte mostra a informação disponível sobre as soluções não dominadas já conhecidas:

Solução	f_1	f_2	f_3	Área (%)	$\mathbf{x}_B$	L_∞
1	66.0	30.0	-12.0	13.43	$x_1 = 18.0$; $x_3 = 6.0$; $x_7 = 14.0$	87.0
2	51.0	50.0	4.0	17.01	$x_1 = 14.0$; $x_4 = 9.0$; $x_5 = 5.0$	71.0
3	15.0	-15.0	75.0	1.25	$x_2 = 15.0$; $x_5 = 45.0$; $x_7 = 20.0$	65.0
4	12.5	50.0	25.0	7.87	$x_4 = 12.5$; $x_5 = 22.5$; $x_6 = 35.0$	53.5

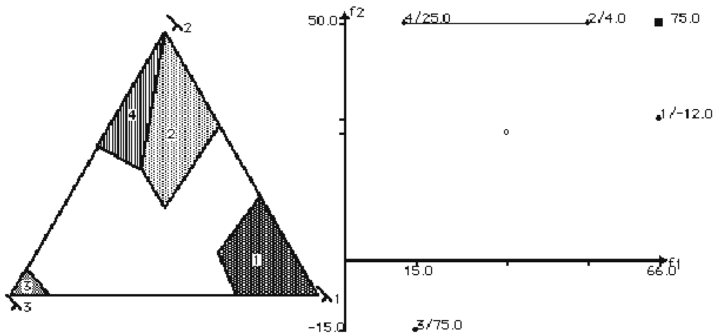


Figura 16.20: Cálculo da solução 4 por posicionamento do rato sobre uma zona não coberta do diagrama paramétrico

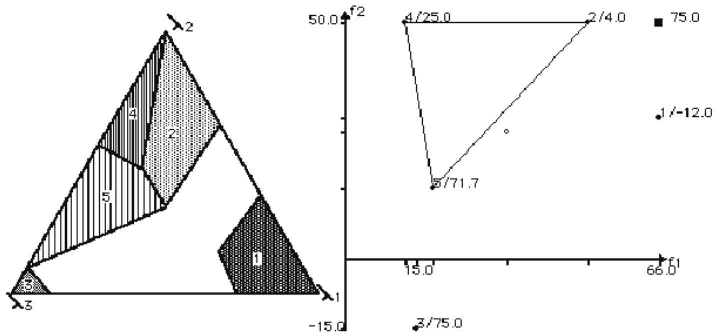


Figura 16.21: Cálculo da solução 5 optimizando uma função ponderada cujo gradiente é normal ao plano que passa pelas soluções 2, 3 e 4 no espaço dos objectivos.

Sobre a solução que minimiza uma distância ponderada (com pesos que se destinam a ter em conta as ordens de grandeza em que são expressas as funções objectivo) de Chebyshev à solução ideal está disponível a informação: $f_1 = 33.9$; $f_2 = 26.9$; $f_3 = 44.2$; $x_1 = 5.7$, $x_2 = 7.2$, $x_3 = 1.6$, $x_4 = 6.3$; $L_\infty = 32.1$.

Nesta fase do processo de decisão interactivo o método TRIMAP é útil ao possibilitar quer explorar regiões não cobertas no diagrama paramétrico, quer analisar mais detalhadamente a região eventualmente já identificada como interessante pelo decisor.

A solução 5 (fig. 16.21) foi calculada optimizando uma função ponderada cujo gradiente é normal ao plano que passa pelas soluções 2, 3 e 4 seleccionadas pelo decisor. A análise do diagrama paramétrico na fig. 16.21 revela-nos que é já conhecida uma face não dominada definida pelas soluções extremas 2, 4 e 5.

Solução	f_1	f_2	f_3	Área (%)	$\mathbf{x}_B$	L_∞
5	18.3	15.0	71.7	17.0	$x_2 = 11.7$; $x_4 = 6.7$; $x_5 = 28.3$	47.7

Para além de possibilitar uma pesquisa livre, TRIMAP é especialmente útil para o estabelecimento de limitações nos valores das funções objectivo. Esta informação é traduzida para o diagrama paramétrico, sendo apresentadas as regiões correspondentes às soluções não dominadas que satisfazem as restrições impostas pelo decisor. Suponhamos que em relação a f_1 o decisor não está interessado em soluções distantes da solução ideal mais do que 40. Esta informação de preferências corresponde

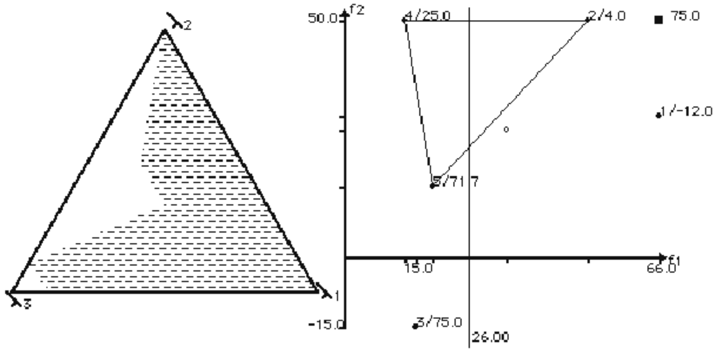


Figura 16.22: Introdução da restrição adicional f1(x) ? 26

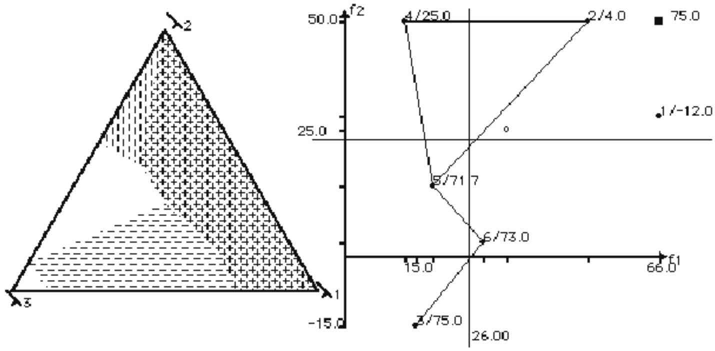


Figura 16.23: Introdução das restrições adicionais $f_1(\mathbf{x}) \geq 26$ e $f_2(\mathbf{x}) \geq 25$.

a introduzir a restrição adicional $f_1(\mathbf{x}) \geq 26$, que é satisfeita na região do diagrama paramétrico assinalada na fig. 16.22. As soluções extremas 1 e 2 (bem como a solução que minimiza a distância de Chebyshev à solução ideal) já conhecidas satisfazem esta restrição adicional imposta pelo decisor.

Vamos supor que o decisor escolhe, por selecção directa (utilizando o posicionamento do rato), um conjunto de pesos de uma zona do triângulo não preenchida satisfazendo esta limitação. Deste modo é obtida a solução 6 (fig. 16.24).

À medida que vai reunindo informação o decisor pode ir diminuindo o âmbito da pesquisa, sobretudo à custa da imposição de limitações adicionais nas funções objectivo, que são traduzidas para o diagrama paramétrico. Suponhamos que para além de manter a informação de preferências correspondente a $f_1(\mathbf{x}) \geq 26$, o decisor está agora também em condições de expressar uma limitação inferior em f_2, por exemplo $f_2(\mathbf{x}) \geq 25$. Na fig. 16.23 as regiões do diagrama paramétrico correspondentes a cada limitação são preenchidas com tramas diferentes colocando em evidência a respectiva intersecção, onde o decisor deve prosseguir a pesquisa se pretender manter-se coerente com estas expressões das suas preferências actuais.

Pesquisando novas soluções na região onde estas duas limitações nos valores das funções objectivo são simultaneamente satisfeitas, encontramos a solução 7 (fig. 16.24). Nesta situação conhecemos a face não dominada definida pelas soluções extremas 2-4-5; existe uma outra face definida pelas soluções extremas 3-5-6 que é dominada pelos pontos situados sobre as arestas 3-6 e 6-5 (comparar com a projecção do espaço dos objectivos), e existem pelo menos duas outras faces não dominadas ainda não completamente conhecidas, uma que inclui as soluções 2-5-6-7 e outra que inclui as soluções

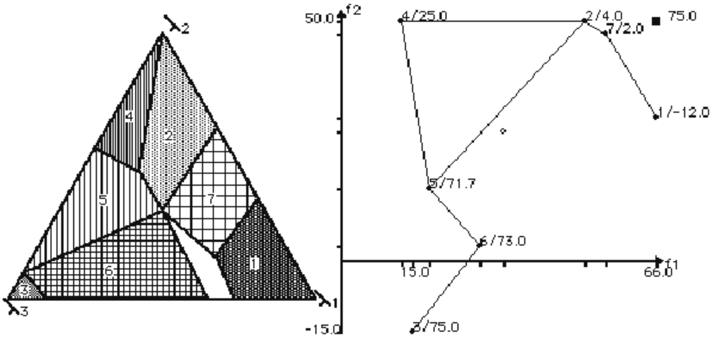


Figura 16.24: Cálculo da solução 7 na região onde as limitações $f_1(\mathbf{x}) \geq 26$ e $f_2(\mathbf{x}) \geq 25$ são simultaneamente satisfeitas

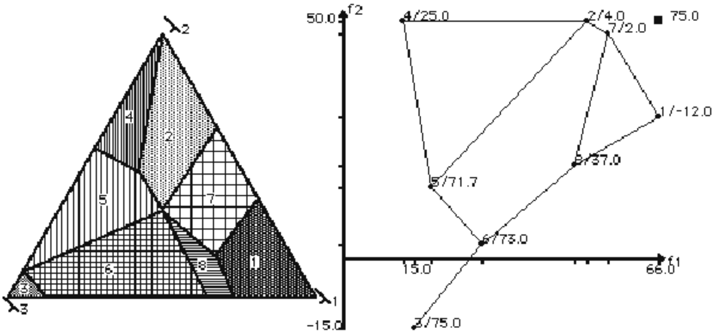


Figura 16.25: Nesta situação são conhecidas todas as soluções (vértices) não dominadas

1-7.

De modo a tentar obter mais informação sobre estas faces não dominadas vamos supor que o decisor, por indicação de um ponto – conjunto de pesos – nesta zona não preenchida do triângulo, calculou a solução 8 (fig. 16.25). Na fig. 16.25 são conhecidas todas as soluções extremas não dominadas, dado que o diagrama paramétrico está totalmente preenchido. Note-se que não é este, em geral, o objectivo da aplicação do TRIMAP.

O problema tem 8 soluções extremas não dominadas e existem 3 faces não dominadas (uma definida pelas soluções 2-4-5, outra por 2-5-6-8-7 e outra pelas soluções 1-7-8).

Solução	f_1	f_2	f_3	Área (%)	$\mathbf{x}_B$	L_∞
6	29.0	3.0	73.0	23.71	$x_2 = 13.0$; $x_3 = 8.0$; $x_5 = 15.0$	47.0
7	55.5	47.5	2.0	15.09	$x_1 = 14.5$; $x_3 = 2.5$; $x_4 = 7.0$	73.0
8	48.5	19.5	37.0	4.62	$x_1 = 7.5$; $x_2 = 7.0$; $x_3 = 9.5$	38.0

6. Integração entre Métodos em Sistemas de Apoio à Decisão

A experiência adquirida com o desenvolvimento das implementações computacionais e com as aplicações de métodos interactivos de programação linear multiobjectivo, bem como a análise comparativa das suas características conceptuais, fizeram surgir a necessidade de dispor de ferramentas computacionais mais flexíveis, capazes de constituir uma base de experimentação quer da adequação dos métodos a tipos de problemas e a fases do processo interactivo de decisão, quer das possibilidades de comutação entre métodos assegurando a máxima transferência de informação, quer mesmo da desagregação dos métodos em procedimentos elementares. Esta linha de evolução teve continuidade com os ambientes interactivos TOMMIX e SOMMIX.

O TOMMIX é um ambiente interactivo vocacionado para problemas com três funções objectivo, incluindo os métodos STEM, Zionts-Wallenius, TRIMAP, Pareto Race e Interval Criterion Weights (representativos de diferentes estratégias de redução do âmbito da pesquisa, técnicas de cálculo de soluções eficientes e formas de informação de preferências), onde a estrutura interna dos métodos é respeitada (Antunes et al., 1992). A partir da avaliação das características conceptuais de cada método e das suas implementações isoladas foi desenvolvida a base integrada de métodos TOMMIX que permite tirar partido da combinação de diferentes tipos de métodos interactivos de programação linear multiobjectivo, tornando possível a transferência de informação entre cada um deles utilizável nas fases subsequentes do processo de decisão, e usando meios de interacção com o utilizador e formas de apresentação de resultados comuns. A base de métodos está especialmente adaptada (embora não limitada) a problemas de programação linear com três funções objectivo, o que permite o uso de meios gráficos particularmente úteis no diálogo com o decisor. Por esta razão foi-lhe dado o nome de TOMMIX (*three-objective methods mixed*).

A base de métodos inclui cinco métodos interactivos de programação linear multiobjectivo: STEM, Zionts-Wallenius (ZW), *Interval Criterion Weights* (ICW), *Pareto Race* (PR) e TRIMAP. Estes métodos:

- São representativos de diferentes estratégias para reduzir o âmbito da pesquisa:
  - redução da região admissível (STEM, TRIMAP),
  - redução do espaço dos pesos (ZW, TRIMAP),
  - contracção do cone dos critérios (ICW),
  - pesquisa direccional (PR);
- Admitem ou não a intervenção do utilizador para solicitar a realização de uma dada função em qualquer momento do processo interactivo:
  - estruturado (STEM, ZW, ICW),
  - não estruturado (PR, TRIMAP);
- Usam diferentes técnicas de cálculo de soluções não dominadas:
  - soma ponderada das funções objectivo (ZW, ICW, TRIMAP),
  - minimização de uma distância a um ponto de referência (STEM, PR, TRIMAP),
  - programação paramétrica em relação aos termos independentes das metas flexíveis (PR);
- Requerem (directa ou indirectamente) distinta informação de preferências do decisor:
  - níveis de aspiração para funções objectivo (PR, TRIMAP),
  - limitações inferiores para funções objectivo (STEM, PR, TRIMAP),
  - limitações nos pesos (TRIMAP),
  - comparações par a par entre soluções (ZW),
  - escolha de uma solução preferida de entre uma amostra de soluções não dominadas (ICW),
  - avaliação de tendências de variação unitária ao longo de arestas eficientes (ZW);

- Apresentam diferentes formas de comunicação de informação com utilizador.

O TOMMIX permite o uso isolado de cada método oferecendo também a possibilidade de trocar de método em qualquer interacção com o decisor. É, assim, possível tirar partido das potencialidades de cada método em diferentes fases do processo de decisão, captar distintas atitudes do decisor, bem como avaliar a qualidade da aplicação de cada um dos métodos em problemas de decisão particulares. Esta base de métodos de programação linear multiobjectivo permite total flexibilidade nas transições entre métodos, dado que as limitações adicionais (que reduzem o âmbito da pesquisa) introduzidas usando um dado método não interferem com as características fundamentais dos outros métodos.

A investigação tendente a estender de forma coerente este ambiente para problemas com mais de três funções objectivo deu origem ao ambiente interactivo SOMMIX (Clímaco et al., 1997), que é baseado num painel de controlo que oferece ao utilizador uma gama de comandos, constituindo o resultado da extensão de TOMMIX ao caso genérico de programação linear multiobjectivo. As ferramentas para calcular novas soluções eficientes, bem como os processos de diálogo com o utilizador são do mesmo tipo dos usados no TOMMIX, onde a estrutura dos métodos era preservada. Os comandos correspondem, no essencial, a todos os procedimentos elementares que é possível identificar nos métodos integrados no TOMMIX, e que podem ser usados de forma independente dos próprios métodos. A informação gráfica sobre as soluções eficientes já calculadas é fornecida sobretudo através de gráficos de barras, *spider web* e projecções do espaço dos pesos. Detalhes sobre o funcionamento e potencialidades de utilização do SOMMIX podem ser consultados em (Clímaco et al., 1997).

7. Breve referência a estudos de aplicação

Estas ferramentas de apoio à decisão em problemas de programação linear multiobjectivo têm sido usadas na realização de vários estudos nas áreas dos planeamento da expansão de sistemas produtores de energia eléctrica, planeamento da modernização de redes de telecomunicações face à introdução de novas tecnologias e novos serviços, e planeamento regional baseado em análise input-output e estudo das interacções energia-economia-ambiente.

Um modelo de planeamento energético considera três funções objectivo quantificando o custo total dos planos de expansão (a minimizar), a fiabilidade do sistema produtor (a maximizar) e o impacte ambiental (a minimizar). As variáveis de decisão referem-se ao tipo de tecnologias de geração (fuel, nuclear e carvão) consideradas para a expansão do sistema produtor. As restrições podem ser classificadas em três categorias: de carga, operacionais das unidades geradoras e orçamentais. Detalhes sobre a modelação matemática deste problema bem como o respectivo estudo usando o método TRIMAP podem ser consultados em (Clímaco et al., 1995). Em (Clímaco e Antunes, 1990) um modelo semelhante é usado para efectuar uma análise comparativa entre os métodos STEM, Zionts-Wallenius e TRIMAP.

Esta abordagem foi depois aprofundada no sentido do desenvolvimento de um novo modelo prestando particular atenção quer à gestão do lado da procura, quer aos impactes ambientais. Neste caso foram também consideradas três funções objectivo a minimizar: custo total de expansão, impacte ambiental associado à capacidade instalada e impacte ambiental associado à produção de energia. A gestão do lado da procura (demand-side management, DSM) é modelada como um novo grupo gerador, competindo com os grupos geradores do lado da oferta. O modelo considera cinco categorias de restrições relacionadas com a fiabilidade do sistema produtor, a disponibilidade das unidades geradoras, a capacidade do grupo gerador equivalente DSM, a capacidade total a instalar ao longo do período de planeamento e as emissões de poluentes. As tecnologias consideradas para expansão do sistema produtor são, para além do grupo equivalente DSM, fuel, carvão, nuclear e gás natural. O estudo foi realizado o usando o método TRIMAP e os detalhes da modelação, bem como alguns resultados ilustrativos, podem ser vistos em (Martins et al., 1996).

Outros dos estudos realizados dizem respeito ao planeamento estratégico da modernização de redes de telecomunicações, face à introdução de novas tecnologias e novos serviços. Em (Antunes et al., 1993) é desenvolvido um modelo de programação linear multiobjectivo baseado num diagrama de transição de estados (cujos nodos caracterizam uma linha de assinante num dado instante do período de planeamento). O modelo considera três funções objectivo que quantificam o valor associado à evolução das linhas de assinante, o grau de modernização associado à desejabilidade de novos serviços (ambos a maximizar), e o custo da dependência externa associada às estratégias de evolução (a minimizar). As restrições expressam limitações inferiores para a satisfação da procura estimada e para a penetração das tecnologias de suporte, limitações técnicas associadas à instalação de linhas e limitações de carácter orçamental. Os resultados foram obtidos usando o método TRIMAP.

Este modelo foi depois estendido para o estudo dos cenários de evolução de redes de acesso de banda larga, num contexto residencial e de pequenas e médias empresas (Antunes et al., 1998). Este modelo é baseado numa tabela de transições de estado que expressa as combinações admissíveis de categorias de serviço (serviço telefónico tradicional, serviço melhorado tipo acesso básico em banda estreita ISDN, serviço comutado de banda larga assimétrico, serviço comutado de banda larga simétrico, serviço distributivo de banda larga tipo CATV tipicamente não comutado e serviço comutado de banda larga avançado) e de arquitecturas tecnológicas para a rede de acesso (par de cobre, par de cobre melhorado tipo ADSL, híbrido fibra-coaxial HFC, FTTC).

Em (Alves et al., 1997) é apresentado um modelo de planeamento regional com quatro funções objectivo, baseado no quadro input-output de uma região de Portugal, que quantificam o consumo privado, o nível de emprego (ambas a maximizar), o deficit da balança comercial da região e o consumo de energia (ambas a minimizar). As restrições derivam de relações de equilíbrio entre os diferentes sectores da economia, limitações impostas na variação de algumas variáveis do modelo, bem como restrições definidoras (por exemplo, o consumo privado definido como uma função do rendimento das famílias). A tabela input-output está organizada em 12 sectores (uma agregação dos 49 sectores das contas nacionais). O output de cada sector é representado por uma variável de decisão. Todas as variáveis são consideradas exógenas à excepção da formação bruta de capital fixo e do consumo colectivo. Este estudo foi realizado com o SOMMIX, com o intuito de demonstrar as potencialidades no apoio à decisão de uma ferramenta baseada num painel de controlo em relação a métodos como o STEM e o Zionts-Wallenius.

O TRIMAP foi usado para o estudo das interacções entre o sistema energético, a economia a nível nacional e os impactes ambientais resultantes de diferentes políticas (Antunes et al., 2002). O modelo de programação linear multobjectivo, baseado numa tabela input-output na qual são considerados 44 sectores económicos, com dados representativos da realidade Portuguesa, inclui como funções objectivo o produto interno bruto, as importações de energia e a auto-produção de electricidade. As restrições dizem respeito à capacidade de armazenamento e stocks de segurança de hidrocarbonetos, às emissões de CO2, ao défice público, às limitações para as exportações e para as importações, à capacidade de produção da economia, ao valor acrescentado bruto e à balança de pagamentos.