

INVESTIGAÇÃO OPERACIONAL

Junho 1990

Número 1

Volume 10

Publicação Científica da



Associação Portuguesa para o Desenvolvimento
da Investigação Operacional

INVESTIGAÇÃO OPERACIONAL

Propriedade:

APDIO — Associação Portuguesa para o Desenvolvimento
da Investigação Operacional

ESTATUTO EDITORIAL

«Investigação Operacional», órgão oficial da APDIO cobre uma larga gama de assuntos reflectindo assim a grande diversidade de profissões e interesses dos sócios da Associação, bem como as muitas áreas de aplicação da I. O. O seu objectivo primordial é promover a aplicação do método e técnicas da I.O. aos problemas da Sociedade Portuguesa. A publicação acolhe contribuições nos campos da metodologia, técnicas, e áreas de aplicação e software de I. O. sendo no entanto dada prioridade a bons casos de estudo de carácter eminentemente prático.

Distribuição gratuita aos sócios da APDIO

INVESTIGAÇÃO OPERACIONAL

volume 10 - nº 1 - Junho 1990

Publicação semestral

Direcção : J.M. Pinto Paixão
(Fac. Ciências - Universidade de Lisboa)
J.J. Júdice
(Fac. Ciências e Tecnologia - Univ. de Coimbra)

Comissão Editorial

Mordecai Avriel	(Israel)	A. Simões Monteiro	(NORMA)
João A. Branco	(IST - Univ. Técn. Lisboa)	Mohamed Najim	(ENSIAS - Argélia)
Josep Casanovas	(UPC - Espanha)	J. Manuel Oliveira	(EFACEC)
J. Dias Coelho	(FE - Univ. Nova Lisboa)	Fernando Pacheco	(Univ. Católica)
Nuno Crato	(NORMA - Açores)	A. Gouvêa Portela	(IST-Univ.Técn.Lisboa)
J.A. Romão Eusébio	(CIMPOR)	M. Baptista Rodrigues	(Partex - CPS)
A. Sousa Ferraria	(Petrogal)	A. Guimarães Rodrigues	(Univ. Minho)
D. V. Gokhale	(Estados Unidos)	Bernard Roy	(LAMSADE- França)
J. Borges Gouveia	(FE - Univ. Porto)	C. Moreira da Silva	(FE - Univ. Porto)
R. Campos Guimarães	(FE - Univ. Porto)	L. Valadares Tavares	(IST-Univ.Técn.Lisboa)
Masao Iri	(TU - Japão)	Isabel H. Themido	(IST-Univ.Técn.Lisboa)
Joaquim J. Júdice	(FC - Univ. Coimbra)	B. Calafate Vasconcelos	(FE - Univ. Porto)
A. Rinnoy Kan	(EU - Holanda)	José M. Viegas	(IST-Univ.Técn.Lisboa)
Nelson Maculan	(UFRJ - Brasil)	Andres Weintraub	(UC - Chile)

A Revista "INVESTIGAÇÃO OPERACIONAL" está registada na Secretaria de Estado da Comunicação Social sob o nº108335.

Esta Revista é distribuída gratuitamente aos sócios da APDIO. As informações sobre inscrições na Associação, assim como a correspondência para a Revista devem ser enviadas para a sede da APDIO - Associação Portuguesa para o Desenvolvimento da Investigação Operacional - CESUR, Instituto Superior Técnico, Av. Rovisco Pais, 1000 Lisboa.

Este Volume foi subsidiado por :

Instituto Nacional de Investigação Científica (INIC)

Junta Nacional de Investigação Científica e Tecnológica (JNICT)

Fundação Calouste Gulbenkian

Para efeitos de dactilografia e composição, foram utilizados equipamentos gentilmente postos à disposição pelo CEAUL (DEIOC- Faculdade de Ciências de Lisboa).

Assinatura : 3000\$00

APOIO À TOMADA DE DECISÃO NUM SISTEMA DE TRANSPORTES

Lélio Amado

RNIP – RODOVIÁRIA NACIONAL, INVESTIMENTOS E PARTICIPAÇÕES, SA
Departamento de Análise de Sistemas Operacionais
Av. Columbano Bordalo Pinheiro, 86 – 1093 LISBOA CODEX

Resumo

A presente comunicação pretende contribuir para o diagnóstico das necessidades de apoio à tomada de decisão num sistema de transportes.

Começando por descrever e caracterizar a composição de um modelo de sistema de gestão de transportes, procurar-se-á identificar a possibilidade e utilidade de aplicações da análise de sistemas e da investigação operacional, com vista à tomada de decisão na gestão dos transportes.

1. Introdução

Os problemas associados à tomada de decisão em sistemas de transportes têm vindo a tornar-se cada vez mais abundantes e simultaneamente mais complexos. Este facto, porém, decorre naturalmente da crescente escassez de recursos encontrada pelas entidades gestoras de sistemas de transporte e da necessidade imperiosa de racionalizar a afectação e a utilização dos recursos disponíveis.

Efectivamente, as empresas de transportes públicos têm sido forçados a procurar novas formas de reduzir a contradição entre a exigência de estabilização ou mesmo expansão dos serviços e a diminuição dos subsídios à exploração. Esta situação resulta dos diferentes interesses dos grupos envolvidos (entidades gestoras dos transportes públicos, políticos e passageiros) e permanece insolúvel.

No caso particular do contexto português, há ainda a ter em conta as transformações que se avizinham para o sector dos transportes, particularmente a alteração da Lei de Bases, com a adopção de um perfil liberalizante, e a integração plena de Portugal na comunidade económica

européia; estas situações exigirão às empresas de transportes portuguesas actuações de carácter mais concorrencial, conduzindo inevitavelmente à pesquisa e adopção de instrumentos mais rápidos e eficientes para apoio à tomada de decisão.

Contudo, a definição e selecção de adequados sistemas de apoio à tomada de decisão deverá passar, previamente, por uma reflexão sobre o próprio processo da tomada de decisão, articulada com as diversas fases da gestão. Por outro lado, há que identificar muito claramente os problemas, determinar as possíveis interrelações entre eles e formalizá-los com vista à sua resolução.

A presente comunicação pretende contribuir para o diagnóstico das necessidades de apoio à tomada de decisão num sistema de transportes.

Assim, no ponto que se segue apresentar-se-á uma concepção da problemática da tomada de decisão, enquadrando-se essa questão no ciclo da gestão.

No ponto 3 far-se-á a caracterização de um sistema de gestão de transportes apreciando-se a inter-relação entre as várias áreas e as necessidades de integração das decisões a nível da gestão geral do sistema.

A apreciação mais detalhada da área da exploração de transporte colectivo de passageiros constituirá o objecto do ponto 4, onde será realçada a importância da utilização de modelos matemáticos e sistemas de informação para apoio à tomada de decisão nos transportes.

2. A Problemática da Tomada de Decisão

A necessidade de tomar decisões decorre naturalmente da existência de problemas, para cuja identificação se torna indispensável a obtenção de informação. Assim, é a informação que proporciona ao decisor a percepção das relações entre os elementos do sistema em que o problema se insere, e permite o conhecimento das características deste com vista à sua resolução.

Por outro lado, tal como se mostra na figura 1, em determinadas circunstâncias poderá obter-se também informação relativa ao provável comportamento do sistema face a determinado tipo de acções do decisor, o

que permitirá a este uma estimativa das reacções a obter (decisão com experiência).

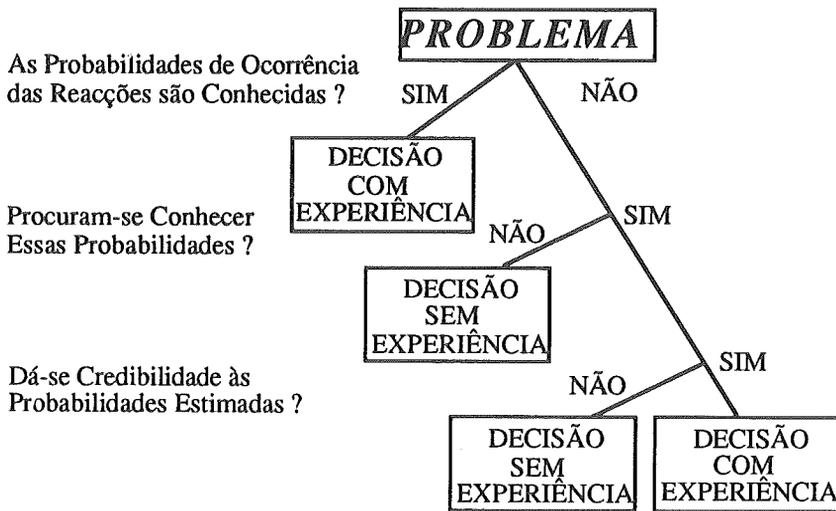


Fig. 1

É mediante esta forma de conhecimento que o agente de decisão estabelece as suas preferências e selecciona as suas acções, naturalmente na expectativa de obter os melhores resultados.

Portanto, em termos funcionais há que hierarquizar os procedimentos e os métodos para uma tomada de decisão racional, começando o centro de decisão por recolher informação conducente à identificação do problema, à caracterização do sistema e à análise do seu comportamento histórico, passando depois à respectiva formalização e obtenção de solução através da utilização de instrumentos adequados, prosseguindo com a execução prática das acções decorrentes da solução encontrada e estando atento aos resultados reais, que analisados comparativamente com os previstos, proporcionarão a actualização da base de conhecimento inicial e eventualmente conduzirão à reformulação da decisão no período seguinte (Fig. 2).

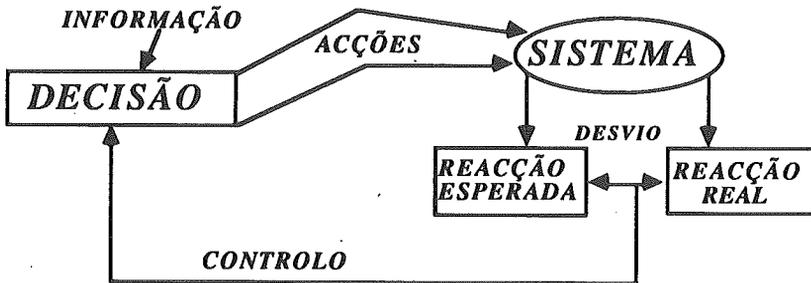


Fig. 2

O modelo que se acabou de descrever e se ilustra na figura anterior é facilmente identificável com o processo de decisão que deve estar associado a qualquer sistema de gestão, no qual se interligam quatro funções fundamentais, constituindo igual número de fases do processo de gestão: informação, planeamento/programação, organização/implantação e controlo (Fig. 3).

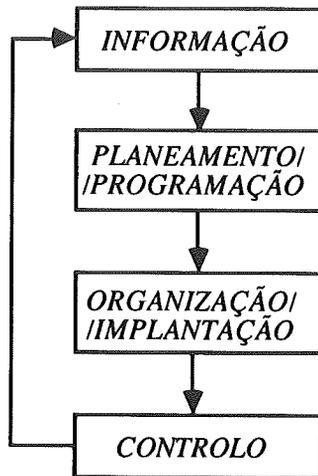


Fig. 3

Os problemas, no entanto, poderão ter âmbitos diferentes, quer em termos do horizonte temporal que abrangem, quer em termos da estrutura do sistema que envolvem. Assim, no que respeita à correspondente tomada de decisão, é frequente adoptar-se uma classificação que, em geral, permite

também a identificação de equivalentes níveis de gestão: **Estratégico, Tático e Operacional.**

De acordo com aquela classificação, é possível caracterizar claramente as decisões sob diversos aspectos, tal como é feito no Quadro I.

QUADRO I

Características	Decisões		
	Estratégicas	Táticas	Operacionais
Objectivo	Aquisição de Recursos	Utilização de Recursos	Execução
Horizonte Temporal	Longo Prazo	Médio Prazo	Curto Prazo
Nível de Management	Alto	Médio	Baixo
Alcance	Grande	Médio	Pequeno
Fonte de Informação	Externa e Interna		Interna
Nível de Detalhe da Informação	Muito Agregado	Moderado	Pouco
Grau de Incerteza	Elevado	Moderado	Fraco
Grau de Risco	Elevado	Moderado	Fraco

(Robert N. Anthony – "Planning and Control Systems")

A diversidade de características de decisão que se observam no quadro anterior pode ainda ser conjugada com dois tipos de decisão, derivados de situações repetitivas ou de situações únicas, o que é bem indiciador da importância que se deve atribuir à escolha de adequados instrumentos de apoio à tomada de decisão. Assim, historicamente os métodos de apoio à tomada de decisão têm vindo a tornar-se mais elaborados (Quadro II), proporcionando simultaneamente maior rapidez e rigor na resposta e possibilitando decisões melhores e mais oportunas.

QUADRO II

Tipo de Decisão	Métodos de Apoio	
	Convencionais	Modernos
* Programada (Estruturada)	Estrutura Organizacional	Processamento Automático da Informação
- Rotina	Procedimentos	
- Repetitiva	Regras Habituação	Modelos Matemáticos
* Não Programada (Não Estruturada)	Políticas	Hierarquização do Sistema
- Única	Intuição	Teoria da Decisão
- Complexa	Treino Promoção	Heurísticas

(Herbert Simon – *"The Shape of Automation for Men and Management"*)

3. Caracterização de um Sistema de Gestão de Transportes

Encarando o transporte numa perspectiva empresarial em que a finalidade é a produção e venda desse serviço, importa definir a composição de um modelo de sistema de gestão que satisfaça as exigências que lhe estão subjacentes, e caracterizar os problemas típicos que se levantam no âmbito dessa actividade, com vista à identificação das necessidades de apoio à tomada de decisão na gestão de transportes.

Como em qualquer outro ramo, também na indústria de transportes é fundamental a detecção das necessidades de mercado por satisfazer, a qual suporta a determinação da produção a realizar; no exercício da actividade produtiva, é então necessário conjugar a utilização de diferentes factores de produção (equipamentos, recursos humanos e recursos financeiros), bem como adoptar adequadas políticas de aquisição de materiais de consumo e/ou subcontratar a prestação de serviços de apoio.

De uma forma sintética, ficamos assim com a relação dos principais problemas que se encontram num sistema de gestão de transportes, o que nos permite definir o modelo representado na figura 4 e antever algumas situações de tomada de decisão.

No modelo representado na referente figura podem-se assim identificar seis áreas funcionais (Produção, Comercial, Manutenção, Aprovisionamento, Recursos Humanos e Finanças), as quais constituem subsistemas do sistema de gestão de transportes. No entanto, dadas as particularidades do sector de transportes, em que a produção e a comercialização são simultâneas, por não ser possível produzir para stock, e em que na maior parte a sua oferta é considerada serviço público, ficando o exercício da actividade sujeito a tarifas e normas rígidas impostas pelo governo, é frequente encontrarmos as funções comercial e produção diluídas numa função mais geral que se tem designado por "exploração".

Dada a finalidade que começamos por enunciar para um sistema empresarial desta natureza, é natural que a área da exploração seja considerada a mais nobre, já que é nela que se consubstancia a própria razão de ser do sistema. Contudo, isto não significa que o funcionamento do sistema resulte exclusivamente das decisões tomadas neste âmbito, pois há que adquirir equipamentos através de decisões de investimento, há que manter os equipamentos em boas condições de funcionamento e utilização, para o que são importantes decisões nas áreas de manutenção e aprovisionamento, e há que contratar, formar e aperfeiçoar os recursos humanos indispensáveis ao suporte da estrutura do sistema. A todas estas decisões aparece ainda associada uma decisão de financiamento, já que é indispensável garantir os meios financeiros necessários à aquisição dos recursos e, em última análise, uma decisão de gestão que valide, ou não, as decisões parcelares a nível de cada subsistema em função dos resultados globais pretendidos, os quais podem ser avaliados exclusivamente em termos financeiros ou com base em critérios de ordem mais qualitativa.

Afectação de Serviços a Estações
Rotas Óptimas para Transportes Escolares e para Sistemas
Dial-a-Ride
Afectação de Meios

*** Área da Manutenção**

Planeamento de Intervenções (Manutenção Preventiva)
Planeamento de Substituição de Frota

*** Área de Aprovisionamento**

Gestão de Stocks
Gestão de Materiais (M.R.P.)

*** Área de Recursos Humanos**

Planeamento de Efectivos de Pessoal
Gestão de Pessoal/Contratação Colectiva

*** Área Financeira**

Gestão de Tesouraria e Financeira
Análise Económica-Financeira de Projectos de Investimento

Cada um destes problemas poderá pois ser objecto de análise, com vista à respectiva formalização e construção de um modelo representativo que permita, através de métodos e algoritmos adequados, pesquisar as melhores soluções.

Importa sublinhar, no entanto, que embora possamos desenvolver e implantar modelos específicos de apoio à decisão para cada um dos problemas que se arrolaram por área funcional, há que manter a visão global do sistema de forma a compreender o objectivo final da tomada de decisão: seleccionar a melhor solução para todo o sistema. Efectivamente, os vários subsistemas estão interligados entre si, pelo que diferentes decisões em cada um terão diferentes implicações nos restantes; por outro lado, é comum que os objectivos dos subsistemas tenham alguns pontos de conflito, e que, por consequência, o óptimo do sistema global não seja o somatório dos óptimos dos vários subsistemas, mas antes uma solução de compromisso entre todos eles.

Dir-se-á, nestas circunstâncias, que estamos perante um problema com objectivos ou critérios múltiplos, o que, desde logo, nos faz sentir a necessidade de integrar as decisões, atribuindo adequados coeficientes aos subobjectivos, que constituirão igualmente factores de conversão de escala para que os objectivos parcelares sejam somáveis. Apesar de tudo, esta forma de abordagem do problema encerra algumas dificuldades, nomeadamente porque os factores de conversão podem diferir de pessoa para pessoa.

Pelo que se acabou de referir, têm vindo a adquirir grande importância e desenvolvimento as metodologias de análise multicritério para apoio à tomada de decisão.

Embora sem negar a importância e as potencialidades dos instrumentos de análise multicritério, continua-se a reconhecer grande utilidade aos modelos específicos de cada um dos problemas atrás enumerados, tanto mais que se partilha a opinião de que os modelos, como meros instrumentos de apoio à decisão, não devem ter sempre o mesmo grau de complexidade da vida real, por forma a que a sua resolução seja possível dentro de relações aceitáveis de custo-benefício.

Dado que, de todas as áreas identificadas no modelo de sistema de gestão atrás apresentado, apenas a "exploração" é específica dos transportes, a seguir apreciar-se-ão mais em detalhe os problemas de gestão encontrados a esse nível.

4. Apoio à Decisão na Exploração de Transportes Colectivos de Passageiros

Com vista ao enquadramento dos vários tipos de instrumentos cuja aplicação se considera adequada para apoio à tomada de decisão na exploração de transportes colectivos de passageiros, concebeu-se o modelo exposto na figura 5, o qual segue a metodologia de gestão definida na figura 3.

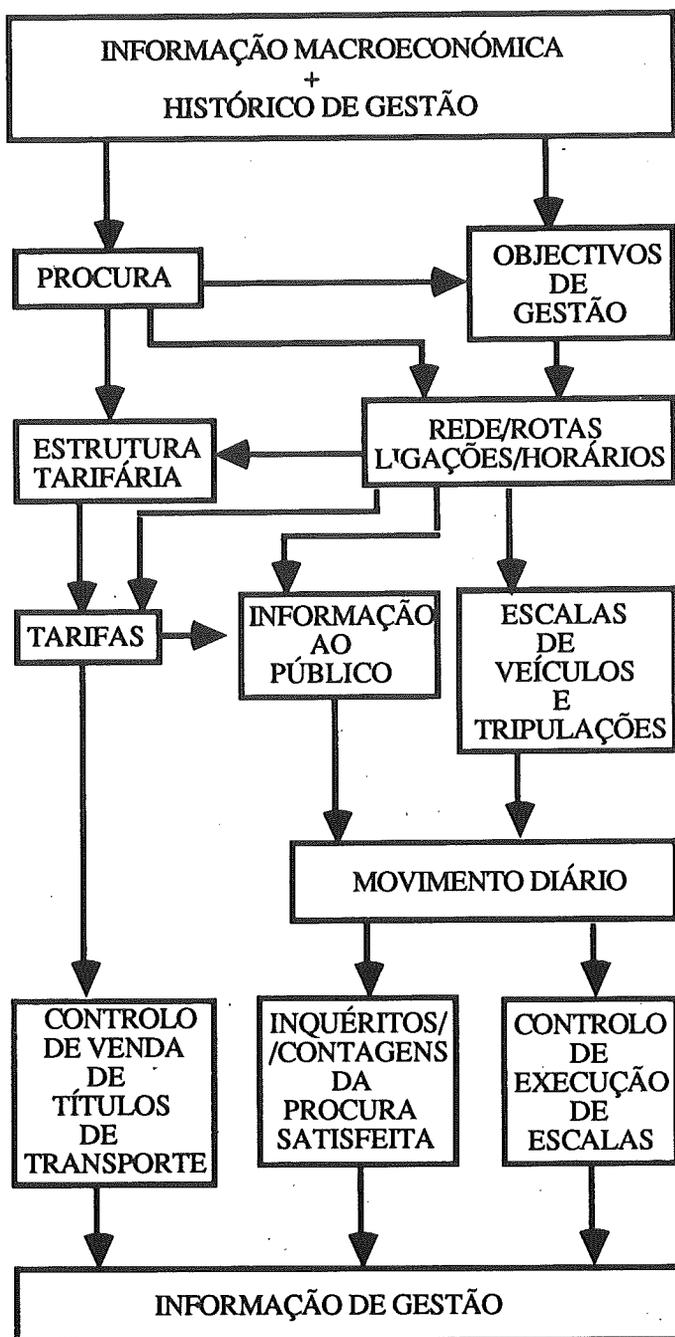


Fig. 5

Assim, partindo-se de uma base de informação composta por indicadores macroeconómicos e informação de gestão histórica, pode-se proceder a uma análise e modelização da procura de transporte, a qual, associada à situação da conjuntura macroeconómica, permitirá aos responsáveis da gestão traçar os objectivos a atingir em termos de produção.

Em face dos objectivos de gestão, da quantificação e da qualificação da procura, pode-se, uma vez reconhecida a rede de infra-estruturas, proceder à elaboração do plano de produção, consubstanciado no traçado de uma rede de transporte, com identificação das rotas, ligações e horários a oferecer.

A estrutura tarifária, bem como as tarifas a aplicar deverão resultar de uma avaliação da cobertura dos custos e da remuneração dos factores produtivos em função da previsível utilização que o sistema irá ter, atendendo ainda a critérios de adequação das formas de cobrança às tipologias de serviço oferecidas. Refira-se, porém, que no actual contexto português e para a parcela do transporte classificada de serviço público, esta decisão não cabe aos operadores de transporte, já que as tarifas são definidas administrativamente.

Uma vez traçado o plano de produção, põe-se o problema de programar a utilização racional dos veículos e tripulações (elaborar as escalas), o que, sublinhe-se, deverá ser feito de uma forma algo interactiva com o plano de produção.

No modelo apresentado integra-se também uma missão de informação ao público, a qual, em função da maior ou menor complexidade do sistema de transportes deverá apoiar a tomada de decisão do passageiro para a melhor opção de utilização do transporte.

Refira-se, finalmente, a inclusão de um subsistema de controlo que, incidindo sobre diferentes aspectos da produção e da venda do serviço de transporte, proporcionará a adequada informação com vista à avaliação dos resultados da gestão.

Do apresentado na figura 5 ressaltam vários aspectos em que a tomada de decisão de gestão deve ser apoiada em adequados modelos matemáticos e sistemas de informação, cuja utilização fica actualmente bastante facilitada pelo recurso aos microcomputadores.

No que respeita ao estudo da procura de transporte, a figura 6 ilustra a metodologia clássica de abordagem do problema, em que se aplicam diversos tipos de modelos matemáticos: **modelos de regressão, modelos gravitacionais, modelos de redes** (caminho mais curto, fluxo máximo, fluxo de custo mínimo, etc.).

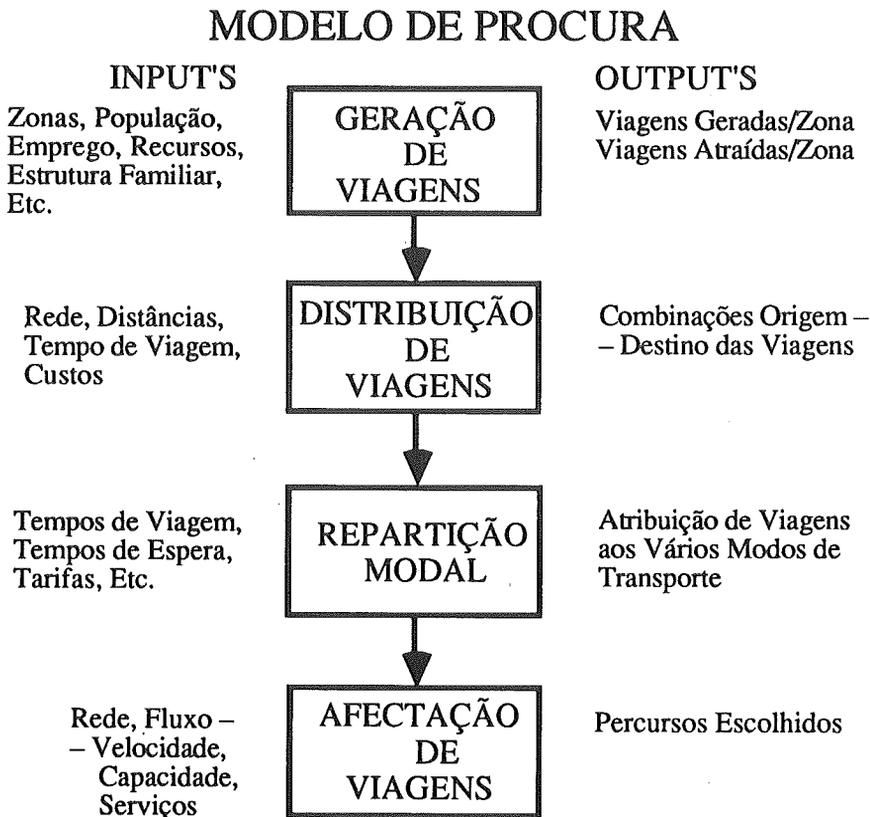


Fig. 6

O planeamento da oferta de transporte público em termos de redes, linhas e horários pode também beneficiar da ajuda de instrumentos de tratamento gráfico interactivo, modelos de simulação de tráfego de veículos e passageiros e mesmo de modelos de programação matemática, como no caso da definição de cadências óptimas de autocarros em linhas urbanas, em que se pode efectuar a seguinte formalização:

Maximizar Intervalo entre Partidas

sujeito a:

$$\text{Prob}(\text{Carga na Paragem}_j > \text{Carga Máx. Adm.}) \leq \beta_1, \text{ (todo e qq. } j)$$

$$\text{Prob}(\text{Espera na Paragem}_j > \text{Espera Máx. Adm.}) \leq \beta_2, \text{ (todo e qq. } j)$$

sendo β_1 e β_2 as probabilidades máximas admitidas para ultrapassagem de determinados "plafonds" caracterizadores da qualidade de serviço.

O planeamento da utilização racional de veículos e tripulações em sistemas de transporte colectivo também é um problema que carece de resolução cuidada e criteriosa, devendo as decisões neste domínio ser suportadas por modelos matemáticos orientados por um objectivo de custo mínimo.

Para o eficiente acompanhamento da execução das acções decididas na fase de planeamento, é ainda imprescindível um conjunto de regras e procedimentos, bem como uma estrutura organizativa e circuitos de informação adequados a uma verdadeira função de controlo, sem a qual o decisor não poderá ter o "feed-back" que lhe permita avaliar o efectivo alcance das suas decisões e tomar novas decisões, reformulando ou não as anteriores.

No entanto, o apoio à decisão num sistema de transporte colectivo de passageiros não se deve restringir ao transportador, já que, para o bom funcionamento do sistema, também é importante uma boa tomada de decisão pelos passageiros.

Assim, um sistema de informação ao público deve prever a necessidade de fornecer respostas actualizadas e "inteligentes" ao passageiro, nomeadamente, qual a melhor rota para uma deslocação: que linhas utilizar, quais os eventuais pontos de transbordo, qual o tempo previsível de viagem, etc.

Para este efeito torna-se necessário dispor de sistemas de informação, equipamentos, modelos e algoritmos apropriados para uma resposta rápida e acessível a qualquer cidadão que possa ler e escrever, como existe em Paris o sistema "SITU", o qual em cada um dos terminais de rua existentes ao longo daquela cidade pode determinar a melhor forma de ir para qualquer destino da

cidade indicado pelo passageiro, de acordo com as opções por ele referidas para o mínimo tempo de percurso ou modos de transporte preferidos.

5. Conclusões

Dada a necessidade de se tomarem boas decisões de gestão, reveste-se de grande importância a utilização de metodologias, técnicas e instrumentos de apoio à tomada de decisão nessa função.

Pela sua concepção sistémica da realidade, pela metodologia científica que a caracteriza e pelas técnicas a que recorre, a Investigação Operacional pode dar um importante contributo no apoio à tomada de decisão num sistema de transportes, devendo a utilização das suas técnicas ser conjugada com formas de tratamento automático da informação.

Considerando-se a grande utilidade dos sistemas de informação, das metodologias de análise, dos modelos matemáticos, dos algoritmos e dos instrumentos de cálculo e dos instrumentos de controlo para a gestão dos transportes, pode-se concluir que existem vastas possibilidades de aplicação da análise de sistemas e da I.O. para apoio à tomada de decisão num sistema de transportes.

Referências

- [1] ANTHONY, R.N. – "Planning and Control Systems"
- [2] BANA E COSTA, C. – "Introdução Geral às Abordagens Multicritério de Apoio à Tomada de Decisão" - Investigação Operacional, Vol. 8, N^o 1, (Junho/1988), 117-139.
- [3] FONSECA, M.C. – "Implementação de uma Heurística para Determinação das Rotações de Serviço do Pessoal Tripulante" - Investigação Operacional, Vol. 8, N^o 1, (Junho/1988), 55-64.
- [4] GREGA, W. – "Decomposition Method for Generation of Optimal Timetables for Urban Bus Network" - Control, Computers, Communications in Transportation, (Paris, Sept./1989), 433-437.

- [5] GUPTA, A.K.; JAIN, S.S.; KHANNA, S.K. – "Simulation Model for Operational Analysis and Management of Urban Bus Routes" - Control, Computers, Communications in Transportation, (Paris, Sept./1989), 379/385.
- [6] PAIAS, A.M. – "Sistema Interactivo para Geração dos Horários de Condutores de Viaturas" - Investigação Operacional, Vol. 8, Nº 1, (Junho/1988), 47-53.
- [7] PAIXÃO, J.; BRANCO, I.M.; CAPTIVO, M.E.; PATO, M.V.; ROMÃO EUSÉBIO, J.A.; AMADO, L. - "Bus and Crew Scheduling on a Microcomputer" - OR Models on Microcomputers, North-Holland (1986), 79-95.
- [8] ROMÃO EUSÉBIO, J.A.; AMADO, L. – "Modelos de Planeamento/Gestão para Optimização de Viaturas e Tripulações numa Rede de Transportes" - II Congresso da APDIO, (Porto/1984).
- [9] ROMÃO EUSÉBIO, J.A.; AMADO, L.; FRAGOSO, L.; PAIXÃO, J. – "Desenvolvimento e Implementação de um Sistema Automático de Escalas de Autocarros e Tripulações na Rodoviária Nacional" - Investigação Operacional, Vol. 8, Nº 1, (Junho/1988), 35-46.
- [10] SIMON, Herbert – "The Shape of Automation for Men and Management".
- [11] SOUSA, J.P.; GUIMARÃES, R.C. – "A Decision Support System for Vehicle and Crew Scheduling in Urban Transport Services" - OR Models on Microcomputers, North-Holland (1986), 65-78.
- [12] STATHOPOULOS, N. – "RETIS: Un Outil Interactif Graphique d'Aide à la Conception de Reseau" - Control, Computers, Communications in Transportation, (Paris, Sept./1989), 457-464.

MEMSOR — METODO MULTICRITÉRIO PARA SEGMENTAÇÃO ORDENADA

Carlos A. Bana e Costa
IST/CESUR
Av. Rovisco Pais
1000 Lisboa

Manuel Campos de Almeida
Gabinete de Estudos e Planeamento do
Ministério das Obras Públicas, Transportes e Comunicações
Av. Columbano Bordalo Pinheiro, 5
1000 Lisboa

Resumo

Muitos contextos de tomada de decisão são caracterizados por apenas se ter acesso a informação de natureza ordinal sobre as preferências intra-critério dos decisores, a par do desejo destes em disporem não só de uma ordenação final das alternativas, mas também de uma indicação de mérito global absoluto para cada uma delas. A conjugação destes requisitos torna difícil o recurso na prática a métodos multicritério, dos vários tipos de abordagens operacionais conhecidas, uma vez que algum ou alguns dos seus pressupostos de aplicação serão sempre violados em tais situações. Para ultrapassar esta situação propõe-se neste artigo um novo processo de agregação que designamos por MEMSOR, Metodo Multicritério para Segmentação Ordenada.

1. Contexto Metodológico

Dois tipos fundamentais de abordagens operacionais têm sido utilizadas para apoiar a tomada de decisão perante problemas discretos de comparação de várias alternativas tendo em conta múltiplos critérios de avaliação: a "abordagem do critério único de síntese" - englobando a "Escola Americana" da Teoria do Valor ou Utilidade (Keeney e Raiffa, 1976) e, também, o Método de Saaty (Saaty, 1988) - e a "abordagem da subordinação de síntese" - base dos métodos multicritérios inseridos na vulgarmente chamada "Escola Francesa" (Roy, 1990).

A primeira apresenta como vantagem o facto de ser possível definir uma medida de mérito (valor) global para cada alternativa, indicadora da sua posição relativa numa ordenação final; no entanto, é necessário dispôr de informação completa (cardinal) sobre as preferências intra-critério dos decisores, o que na maioria dos casos práticos é muito difícil ou mesmo impossível de obter.

Por sua vez, uma das grandes vantagens da abordagem da subordinação de síntese é precisamente permitir ordenar (pelo menos parcialmente) as alternativas em termos relativos, mesmo quando a informação de que se dispõe sobre as preferências critério a critério é pobre; no entanto, não é possível a indicação do mérito global de cada alternativa.

2. Objectivos do Método MEMSOR

A aplicação de qualquer método de análise multicritério pressupõe a especificação anterior de qual o objectivo que o decisor pretende alcançar quando se propõe comparar entre si várias alternativas de decisão recorrendo a múltiplos critérios de avaliação.

Por influência do paradigma da optimização, característico da investigação operacinal (clássica), durante bastante tempo aceitou-se sem discussão que um decisor procura essencialmente escolher a "melhor" alternativa de entre um conjunto pré-definido. No entanto, muitas vezes o decisor pretende não só escolher a melhor alternativa em termos relativos, mas também saber se essa "melhor" alternativa tem um valor absoluto global suficientemente bom para ser tomada como decisão a implementar. Em terminologia da Teoria da Medição ("Measurement Theory", ver (Roberts, 1979)) dir-se-á que, como vários psicólogos têm feito notar (por exemplo, Blumenthal (1977)) existe uma distinção clara entre medição relativa e medição absoluta de valor.

No campo da análise multicritério este problema tem merecido pouca atenção. Refiram-se as preocupações de Saaty (1986 e 1987) e o conceito de problemática da segmentação introduzido por Bernard Roy (1985) – ver definição precisa em (Bana e Costa, 1986). No contexto da "Abordagem de Subordinação", Moscarola e Roy (1977) desenvolveram um método

adequado à afectação das alternativas de um dado conjunto a classes (segmentos) pré-definidos, com referência a limiares mínimos de aceitabilidade e máximos de rejeição – a chamada segmentação tricotómica (ver, também, (Moscarola, 1977) e (Roy, 1981)).

No entanto, este processo não responde ao duplo problema da comparação por mérito relativo e absoluto, isto é, não permite simultaneamente fornecer indicação sobre o posicionamento relativo das alternativas e, por outro lado, definir o mérito absoluto de cada uma delas. Foi com este objectivo que os autores desenvolveram o método MEMSOR – Método Multicritério para Segmentação Ordenada – que se apresenta neste artigo.

3. Desenvolvimento do Método: afectação de alternativas a classes ordenadas

Seja A um conjunto de alternativas $a_1, \dots, a_1, \dots, a_n$ que são apreciadas segundo um conjunto C de critérios $C_1, \dots, C_j, \dots, C_m$. Para cada critério C_j é definida uma escala de avaliação (pontuação) com diferentes níveis, variando de c_j^* a c_j^* (respectivamente o pior e o melhor níveis de avaliação atribuíveis a qualquer alternativa), expressando as preferências do decisor nesse critério. Designe-se por c_{ij} a pontuação atribuída pelo decisor à alternativa a_i ($i = 1, \dots, n$) segundo o critério C_j ($j = 1, \dots, m$).

O método que se propõe consiste em afectar cada alternativa a_i a um segmento do espaço definido pelos critérios e pontuações, através da comparação de a_i com alternativas virtuais de referência. Este processo tem algumas analogias metodológicas com a análise tricotómica de Moscarola-Roy referida no ponto 2, uma vez que o conjunto A das alternativas em confronto é segmentado segundo vários níveis (multidimensionais) de valoração absoluta.

No entanto, o método que se apresenta vai mais longe, pois que, para além de cada alternativa ser afectada a um determinado segmento, todos os segmentos estão hierarquizados, existindo uma preferência estrita entre os níveis consecutivos S_k e S_{k-1} definidores de cada segmento.

Vejamos como se define o perfil de pontuações que segundo cada critério definem os limiares S_k e S_{k-1} de cada segmento.

Tratemos o caso mais simples em que a pontuação segundo cada critério varia numa escala ordinal de 1 a 5, idêntica para todos os critérios e em que cada ordinal tem uma designação deste tipo:

- 5 – Muito Bom
- 4 – Bom
- 3 – Média
- 2 – Aceitável
- 1 – Sem Interesse

Para a construção das alternativas virtuais de referências que constituem os limiares da segmentação adoptou-se a regra:

- 5 = Muito bom em todos os critérios
- 4 = Bom em todos os critérios
- 3 = Média em todos os critérios
- 2 = Aceitável em todos os critérios
- 1 = Sem interesse em nenhum critério.

É em relação a estas alternativas virtuais de referência que é comparada cada alternativa a_j , podendo para o efeito utilizar-se qualquer dos métodos de ordenação conhecidos, desde que sejam adequados ao tratamento de informação ordinal nos critérios. No presente artigo optou-se pelo ELECTRE III (Roy, 1978) para proceder à afectação por segmentos.

O processo consiste em n aplicações do método, tantas quantas o número de alternativas de A , sendo cada corrida feita sobre o conjunto formado por uma alternativa a_j e todas as alternativas virtuais S . Anote-se que se posteriormente se pretender ordenar as alternativas posicionadas num mesmo segmento, poderá recorrer-se à aplicação do ELECTRE III apenas sobre o sub-conjunto de alternativas contidas no mesmo conjunto.

O interesse do método proposto reside em que além de comparar as alternativas de A entre si, elas são localizadas numa escala absoluta de valoração global. Assim, poder-se-á não só afirmar que entre duas alternativas uma é melhor que outra, mas também efectuar o seu posicionamento numa escala absoluta emitindo um juízo de valor sobre o mérito global.

4. Aplicação do Método

A aplicação do método descrito vai ser efectuada recorrendo a um caso real que já foi objecto de análise no âmbito da Comunidade Económica Europeia na acção COST 307, "Política de Conservação de Energia nos Transportes".

O problema consistia em seleccionar as medidas de política a testar em modelo, a partir das preferências manifestadas pelos seis países intervenientes no projecto, além da própria Comissão. As entidades representantes dos vários países manifestaram as suas preferências atribuindo pontuações a cada medida segundo um conjunto de doze critérios.

As medidas de política consideradas foram as indicadas na tabela da página seguinte e os critérios considerados para a sua avaliação foram:

- C₁ – Quantidade de energia economizada
- C₂ – Aplicabilidade a todos os países
- C₃ – Realização de economias a curto prazo
- C₄ – Efeitos positivos contra a poluição
- C₅ – Outros efeitos positivos (tráfego urbano)
- C₆ – Não ter efeitos negativos colaterais
- C₇ – Custo para a sua introdução
- C₈ – Custo de controle de efeitos
- C₉ – Custo da sua manutenção
- C₁₀ – Problemas para a sua implementação
- C₁₁ – Tempo para a sua habituação
- C₁₂ – Aceitação pela colectividade

Medidas de Política Alternativas

<u>Legislativas</u>	<u>Infraestruturas/Gestão do Tráfego</u>	
L1 rendimento do veículo	controle de circulação rodoviário	I1
L2 limite ao consumo	racionalização do transporte aéreo	I2
L3 carga útil	comboios de alta velocidade	I3
L4 preferência ao transporte público	transporte combinado terra-mar	I4
L5 limite de velocidade	transporte combinado rodo-ferroviário	I5

L6 protecção do meio ambiente (poluição)	criação de novas infraestruturas	I6
<u>Financeiras</u>	<u>Promoção de Investigação</u>	
F1 custo dos veículos-energético	desenvolvimento tecnológico	P1
F2 subvenções ao investimento	sugestões de circulação	P2
F3 preços especiais para os transportes públicos	substituição de produtos energéticos	P3
F4 tarifação rodoviária	planificação socio-económica do território	P4
F5 diferenciação do preço dos carburantes	telemática	P5
F6 medidas fiscais sobre os veículos	<u>Comportamento do condutor</u>	
	educação	C1
	informação	C2

Como era objectivo dos decisores seleccionar uma única alternativa de cada grupo de medidas de política, a metodologia proposta neste artigo será apenas aplicada a um único grupo, neste caso o das medidas de carácter legislativo.

Por uma questão de simplificação da apresentação do método vamos apenas efectuar a classificação considerando as preferências manifestadas por uma única entidade (decisor), e considerando os pesos a atribuir a cada critério que esta entidade adoptou.

		Matriz Multicritério											
Medidas		C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈	C ₉	C ₁₀	C ₁₁	C ₁₂
legislativas	Pesos	10	6	6	8	8	8	5	4	5	5	5	8
L ₁		3	3	2	3	1	2	3	3	2	2	2	2
L ₂		3	2	2	3	1	2	3	3	5	3	4	4
L ₃		2	2	2	2	2	1	3	3	4	3	3	2
L ₄		4	2	5	4	5	3	3	3	3	2	2	2
L ₅		3	3	4	4	1	2	2	2	5	2	2	2
L ₆		1	3	1	5	1	1	1	2	2	5	4	4

Para a comparação de cada alternativa/medida legislativa com as alternativas virtuais utilizaram-se os conceitos de índice de concordância e índice de discordância no mesmo sentido com que são utilizados no método ELECTRE III (ver Roy, 1990). Adoptou-se para todos os critérios um nível de discordância $q > 2$ que é tomado como nível de veto.

Para a aplicação do método que se designa por MEMSOR I comparou-se cada alternativa/medida legislativa com os diferentes limiares $S_5 \dots S_1$.

Limiares	1ª Fase	2ª Fase
5 - Muito bom	[5 ... 5] \Rightarrow	
4 - Bom	[4 ... 4] \Rightarrow	
		L_4
3 - Médio	[3 ... 3] \Rightarrow	L_6
		$L_1 L_2 L_3 L_5$
		$L_2 \rightarrow L_5 \rightarrow L_6 \rightarrow L_3 \rightarrow L_1$
2 - Aceitável	[2 ... 2] \Rightarrow	
1 - sem interesse	[1 ... 1] \Rightarrow	

Da primeira fase de aplicação resultou que:

- i) a medida legislativa L_4 estaria acima do nível Médio e abaixo do nível Bom;
- ii) a medida legislativa L_6 é pior que o nível Bom, melhor que o nível Aceitável, mas é incomparável com o nível Médio.
- iii) as medidas legislativas L_1, L_2, L_3 e L_5 estão entre os níveis Aceitável e Médio.

A segunda fase de aplicação consiste na ordenação das alternativas L_1, L_2, L_3, L_5 pela utilização do ELECTRE III, incluindo-se também a alternativa L_6 neste grupo. Obteve-se a seguinte ordenação: $L_2 \rightarrow L_5 \rightarrow L_6 \rightarrow L_3 \rightarrow L_1$.

Em complemento ao processo seguido no método MEMSOR I, pode-se recorrer a outro processo operacional para a segmentação e ordenação simultânea das alternativas. Este novo método, designado por MEMSOR II, consiste na aplicação do método ELECTRE III directamente sobre o conjunto formado por todos os limiares ou alternativas virtuais e todas as alternativas/medidas legislativas propostas. A segmentação/ordenação obtida foi a seguinte:

Limiares

5 - Muito bom	[5 ... 5] ⇒	
4 - Bom	[4 ... 4] ⇒	
3 - Médio	[3 ... 3] ⇒	L_4
		$L_2, L_6 \rightarrow L_5 \rightarrow L_3 L_1$
2 - Aceitável	[2 ... 2] ⇒	
1 - sem interesse	[1 ... 1] ⇒	

Verifica-se que com a aplicação do MEMSOR I L_4 se localizava acima do nível Médio e que aplicando o método MEMSOR II L_4 é incomparável com aquele nível, sendo ainda de notar que a ordenação das alternativas situadas acima do limiar Aceitável resulta menos discriminante.

Estas alterações no ordenamento das alternativas são devidas ao facto da posição relativa final de duas alternativas quaisquer não ser independente das restantes alternativas. Ao aplicar-se o método MEMSOR I cada alternativa a_i é comparada isoladamente com as alternativas fictícias definidoras dos limiares de segmentação, enquanto que ao aplicar-se o método MEMSOR II todas as alternativas a_i são comparadas simultaneamente com as alternativas fictícias S_k .

5. Conclusões

A vantagem do método MEMSOR é a de se poderem testar as alternativas em presença face a alternativas virtuais definidas de um modo preciso e que têm uma valoração e significado absoluto. Cada alternativa a_i é situada no conjunto das alternativas virtuais, as quais estão associadas a uma escala hierarquizada (que no exemplo tratado varia de "muito bom" a "sem interesse"). Assim, no caso tratado, ao seleccionar a medida de política L_4 porque é a melhor em termos relativos, sabemos também que estamos a seleccionar uma medida de mérito global absoluto entre médio e bom.

Bibliografia

- [1] BANA E COSTA, C.A. (1986) – *A problemática da rejeição em decisão multicritério*, *Investigação Operacional*, 6(1/2) (95-104).
- [2] BLUMENTHAL, A.L. (1977) – *The Process of Cognition*, Prentice-Hall, Englewood Cliffs.
- [3] KEENEY, R., RAIFFA, H. (1976) – *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, Wiley, New York.
- [4] MOSCAROLA, J. (1987) – *Multicriteria decision aid: two applications in education management*, *Cahier du Lamsade* N° 11, Université de Paris-Dauphine.
- [5] MOSCAROLA, J., ROY, B. (1977) – *Procédure automatique d'examen de dossier, fondés sur une procédure trichotomique en présence de critères multiples*, R.A.I.R.O., Vol. 11, N° 2.
- [6] ROBERTS, F.S. (1979) – *Measurement Theory with applications to Decision Making, Utility and the Social Sciences*, Addison-Wesley, Reading, Massachusetts.
- [7] ROY, B. (1978) – *ELECTRE III: un algorithme de classement fondé sur une représentation floue des préférences en présence de critères multiples*, *Cahier du CERO*, 20 (3-24).
- [8] ROY, B. (1981) – *A multicriteria analysis for trichotomic segmentation problems*, em P.Nijkamp and J.Spronk (eds.), *Multiple Criteria Analysis: Operational methods*, Gower, Aldershot (245-257).
- [9] ROY, B. (1985) – *Methodologie Multicritère d'Aide à la Decision Economica*, Paris.
- [10] ROY, B. (1990) – *The outranking approach and the foundations of ELECTRE methods*, em C.Bana e Costa (eds.), *Readings in Multiple Criteria Decision Aid*, Springer-Verlag, Berlim (155-183).
- [11] SAATY, T.L. (1986) – *Absolute and relative measurement with the AHP: the most livable cities in the United States*, *Socio-Econ.Plann.Sci.*, Vol. 20, N° 6 (327-331).

- [12] SAATY T.L. (1987) – *Rank generation, preservation, and reversal in the analytic hierarchy decision process*, Decision Science, Vol. 18, Nº 2 (157-177).
- [13] SAATY, T.L. (1989) – *The Analytic Hierarchy Process*, University of Pittsburg.

ESTIMAÇÃO DA DISTRIBUIÇÃO DA TAXA DE PRODUÇÃO DE SISTEMAS INDUSTRIAIS COM MÁQUINAS EM PARALELO POR UM MÉTODO DE AGREGAÇÃO E DECOMPOSIÇÃO

Rui Alves
Faculdade de Economia
Universidade do Porto

Resumo

A avaliação do desempenho de sistemas industriais do tipo fluxo contínuo (como as linhas de montagem) cujo equipamento não é fiável por estar sujeito a avarias de natureza aleatória é um tópico importante e que foi extensamente estudado para configurações em linha. A tendência recente para automatizar este tipo de sistemas industriais (com a utilização de robots, por exemplo) tem sido acompanhada pela adopção de configurações em que existem máquinas em paralelo, embora pouca atenção tenha sido dada à avaliação do desempenho para este género de configurações. Nesta comunicação os sistemas deste tipo são representados por Cadeias de Markov em tempo contínuo, e a distribuição da respectiva taxa de produção é estimada através de um método de agregação e decomposição que permite obter resultados aproximados que, de forma geralmente conservadora, se aproximam bastante dos valores exactos. É de realçar, no entanto, que o método se torna exacto para as configurações em série pura e em paralelo puro.

1. Introdução

Os sistemas industriais do tipo fluxo contínuo caracterizam-se por terem o equipamento e os postos de trabalho dispostos na sequência pela qual são executadas as operações e por fabricarem grandes quantidades de apenas um ou poucos produtos semelhantes (ou seja, sujeitos às mesmas operações). Destes, são considerados nesta comunicação os sistemas cuja produção é constituída por unidades discretas (e não por um fluído) e cujos tempos de operação podem ser considerados determinísticos, sendo as linhas de

montagem e as linhas automatizadas (*Automatic transfer lines*, na terminologia inglesa) os exemplos mais típicos. A configuração em linha tem sido escolhida devido à sua simplicidade e à facilidade com que permite movimentar elevados volumes de produção; contudo, tal configuração tem a desvantagem de o fluxo de produção ser facilmente perturbado por qualquer avaria ao longo do sistema.

As linhas de montagem tradicionais (manuais) não apresentam grandes problemas de fiabilidade e por isso costumam funcionar de forma síncrona, ou seja, os produtos são movimentados de cada estação para a seguinte a intervalos regulares e constantes (a que se chama *ciclo*). O problema mais importante das linhas de montagem é o da afectação das operações às estações, procurando-se fazê-lo de modo a conseguir tanto quanto possível balancear a linha, o que é equivalente a minimizar os tempos mortos. Tal problema tem sido objecto de grande atenção desde há mais de trinta anos, tendo sido Salveson [1955] quem primeiro apresentou a sua formulação como um programa linear. Seguiram-se-lhe Bowman [1960] com a formulação em programação inteira, e Jackson [1956] que propôs uma formulação e método de resolução por programação dinâmica. Desde então aparecerem diversas variantes das formulações apresentadas, assim como dos métodos de resolução. O problema do balanceamento de linhas de montagem é NP-hard (Baybars [1986]), não sendo por isso de estranhar que tenham sido propostos vários métodos inexactos (isto é, métodos que não dão qualquer garantia de obter soluções óptimas). Baybars [1986] apresenta um estudo dos métodos exactos, ao passo que em Baybars [1984] se passam em revista os métodos inexactos. Primeiro Mastor [1970] e mais recentemente Talbot et al. [1986] efectuaram estudos comparativos de diferentes métodos de resolução, embora o último trabalho tenha incluído um maior número de métodos (alguns dos quais exactos). Estes estudos demonstraram claramente que existem métodos para obter soluções óptimas ou quase-óptimas com um esforço computacional muito pequeno.

As linhas automatizadas caracterizam-se por terem uma série de estações ligadas por um mecanismo transportador (como um tapete rolante ou *conveyor*), sendo destinadas à produção de grandes quantidades de determinado artigo. Tratando-se de sistemas mecanizados, os tempos das

operações podem ser considerados determinísticos, mas, dada a ocorrência de avarias (com as necessárias reparações), a sua fiabilidade é um problema importante. A forma tradicional de prevenir contra a perda de produção durante os tempos de reparação é a constituição de "stocks intermédios" (Buffer stocks, na terminologia inglesa) ao longo da linha. Um dos primeiros estudos sobre este tema foi realizado por Buzacott [1967], o qual, pressupondo que os tempos inter-avarias e de reparação seguiam distribuições geométricas, modelou o sistema como uma cadeia de Markov em tempo discreto com o objectivo de estimar o aumento da taxa média de produção do sistema na presença de tais stocks. Entre muitos outros estudos realizados sobre este tema desde então podemos citar Ohmi [1981], que desenvolveu um método de aproximação baseado em modelos de Markov para linhas com um qualquer número de estações, Kubat e Sumita [1985], que resolveram o problema da colocação dos stocks ao longo da linha através de programação dinâmica, e de Koster [1987], que apresenta um método de agregação baseado num tratamento analítico que usa pontos de regeneração.

Por outro lado, poucos estudos têm sido feitos sobre sistemas não fiáveis com tempos de operação determinísticos e máquinas em paralelo. Buzacott [1986] foi provavelmente o primeiro autor a debruçar-se sobre tais sistemas ao considerar o impacto dos diferentes esquemas básicos de redundância (máquinas em paralelo, em *stand-by* e em *splitting*) na eficiência do sistema. Estes três esquemas caracterizam-se da seguinte forma (para uma estação com duas máquinas): no caso de máquinas em paralelo ambas trabalham ao mesmo tempo e, se uma avariar, a taxa de produção da estação fica reduzida à da outra máquina; no caso de máquinas em *stand-by* apenas funciona uma máquina de cada vez, e após a ocorrência de uma avaria a produção é encaminhada para a outra máquina; no caso de máquinas em *splitting* funcionam ambas as máquinas a metade da sua taxa potencial, e quando uma delas avaria a outra duplica a sua taxa de produção durante o tempo de reparação. Ignall e Silver [1977] consideram um sistema com duas estações e um stock intermédio de capacidade finita, em que cada estação tem uma ou mais máquinas não fiáveis, e apresentam uma heurística para estimar a respectiva taxa de produção. O mesmo tipo de sistema é estudado por Elsayed e Hwang [1986], os quais utilizam análise de regressão para estimar a

eficiência do sistema em função da dimensão do stock intermédio e das taxas de avarias e de reparações. Quase todos os sistemas considerados são de reduzida dimensão, tendo no máximo duas estações e/ou duas máquinas por estação. Mais recentemente Alves [1988], [1989] considerou de forma explícita a avaliação do desempenho de sistemas com várias estações e um qualquer número de máquinas em paralelo em cada uma. Esta comunicação, baseada largamente no primeiro daqueles trabalhos, considera o problema da obtenção da distribuição da taxa de produção para tais sistemas. Acontece que o paralelismo em sistemas industriais tem vindo a receber cada vez mais atenção (Bookbinder e Kotwa [1987]), sendo possível demonstrar (Alves [1988]) que tais configurações têm a vantagem de reduzir a variância da distribuição da taxa de produção.

A organização desta comunicação é a seguinte. Na secção 2 o sistema industrial considerado é descrito em pormenor, e é apresentada a cadeia de Markov que o modela. Dada a dimensão do respectivo espaço de estados é praticamente impossível obter a solução exacta, e por isso na Secção 3 é apresentado um método de agregação e decomposição que permite obter soluções aproximadas. A Secção 4 contém a descrição do algoritmo de resolução e a discussão dos resultados dos trabalhos computacionais, enquanto na Secção 5 se apresentam as principais conclusões deste trabalho. A última secção lista a bibliografia referida no texto da comunicação.

2. Caracterização do Sistema e dos Modelos

Nesta secção é descrito o sistema industrial considerado e são definidos os processos estocásticos usados para o modelar. O sistema é constituído por uma série de estações, cada uma das quais executa um conjunto de tarefas e pode ter apenas uma ou diversas máquinas em paralelo. Os principais pressupostos acerca do sistema e do seu modo de operar são os seguintes:

- (1) É fabricado apenas um artigo (embora não seja difícil estender a análise a certos casos de produção de mais do que um artigo).

- (2) As máquinas pertencentes à mesma estação são idênticas; este pressuposto é feito para facilidade de exposição, podendo ser abandonado com relativa facilidade.
- (3) Não existem stocks intermédios; todas as estações têm de operar à velocidade da(s) mais lenta(s) – o(s) estrangulamento(s) – e o sistema opera de forma síncrona. É suposto que a primeira estação tem sempre material para processar e que a última nunca se encontra bloqueada por não ter para onde enviar os artigos produzidos.
- (4) As máquinas não são fiáveis, isto é, estão sujeitas a avarias que ocasionam a sua paragem até estarem reparadas.
- (5) A taxa de avarias (aqui considerada como o número de avarias por artigo processado) de cada máquina é constante no tempo e proporcional à velocidade de operação; quer isto dizer que apenas são consideradas avarias causadas pelo funcionamento da máquina (segundo Buzacott e Hanifin [1978] a maioria das avarias são deste tipo), e que o tempo inter-avarias segue uma distribuição exponencial dependente da velocidade de funcionamento. Este pressuposto é geralmente encontrado dada a inerente facilidade de tratamento do modelo.
- (6) A taxa de reparações (definida como o número de reparações por unidade de tempo) de cada máquina é constante no tempo e independente do número de máquinas avariadas; isto quer dizer que os tempos de reparação também seguem distribuições exponenciais (dados empíricos recolhidos por Bookbinder e Kotwa [1987] parecem justificar tal pressuposto), e que o número de máquinas avariadas nunca excede a capacidade das instalações de reparação, o que se justifica pelo facto dos sistemas existentes não terem um grande número de estações nem muitas máquinas em paralelo.
- (7) Os tempos de operação são determinísticos e após uma avaria não há necessidade de considerar qualquer desperdício nem de repetir operações já executadas; os tempos determinísticos justificam-se pela pouca variabilidade encontrada nos equipamentos automatizados, e o outro pressuposto é feito para facilidade de exposição.

Dado que não existem stocks intermédios, a taxa de produção do sistema é determinada pela estação mais lenta, ou seja, a estação que apresenta a taxa de produção potencial mínima. Isto implica que a taxa de produção real de qualquer estação é sempre igual à taxa do sistema e que as estações com taxas potenciais superiores têm de operar aquém da sua capacidade. Está-se assim a pressupor que as máquinas podem operar a qualquer velocidade entre zero e a velocidade máxima; embora este pressuposto seja irrealista, o mesmo resultado é alcançado através da programação dos tempos de espera apropriados entre as operações, o que é relativamente fácil de fazer em instalações produtivas automatizadas.

A notação utilizada é a seguinte:

S = número de estações;

n_i = número de máquinas idênticas na estação i ;

$y_i(t)$ = variável aleatória que representa o número de máquinas não avariadas na estação i no momento t ;

r_i = taxa máxima de produção de cada máquina não avariada na estação i (número de artigos processados por unidade de tempo);

f_i = taxa de avarias de cada máquina na estação i , medida em número de avarias por artigo processado;

μ_i = taxa de reparações de cada máquina na estação i , medida em número de reparações por unidade de tempo.

A taxa de produção potencial da estação i no momento t depende da velocidade das suas máquinas e do número de máquinas não avariadas, sendo dada por $y_i(t)r_i$. No momento t uma ou mais estações constituirão o estrangulamento do sistema e determinarão a respectiva taxa de produção $s(t)$, ou seja,

$$s(t) = \min_{i=1, \dots, S} \{y_i(t)r_i\} \quad (2.1)$$

A taxa de produção máxima do sistema, atingida quando não existem máquinas avariadas, é dada por

$$r_{\max} = \min_{i=1, \dots, S} \{n_i r_i\} \quad (2.2)$$

Assim que uma estação tenha todas as máquinas avariadas e, portanto, taxa de produção potencial igual a zero, a taxa do sistema cai também a zero.

Da definição das taxas de reparações deduz-se que as reparações na estação i no momento t ocorrem a uma taxa dada por $[n_i - y_i(t)]\mu_i$. Por outro lado, dada a definição das taxas de avarias e a sua natureza, o tempo médio inter-avarias depende da taxa de produção do sistema; no momento t a taxa de produção real da estação i é $s(t)$, e assim a taxa de avarias na estação i é dada por $s(t)f_i$.

Seja $\{\underline{Y}(t), t \geq 0\}$ o processo estocástico em tempo contínuo, em que os elementos do vector $\underline{Y}(t)$ representam o número de máquinas não avariadas em cada estação no momento t , isto é,

$$\underline{Y}(t) = [y_1(t), \dots, y_i(t), \dots, y_S(t)] \quad (2.3)$$

As taxas de transição deste processo são dadas por:

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \frac{\Pr\{\underline{Y}(t+\epsilon) = \underline{Y}^1 \mid \underline{Y}(t) = \underline{Y}^0\}}{\epsilon} &= v(\underline{Y}^0, \underline{Y}^1) \\ &= \begin{cases} (n_j - y_j^0)\mu_j, & \text{se } y_j^1 = y_j^0 + 1, y_k^1 = y_k^0 (k \neq j) \\ s_0 f_j, & \text{se } y_j^1 = y_j^0 - 1, y_k^1 = y_k^0 (k \neq j) \\ 0, & \text{em caso contrário} \end{cases} \quad (2.4) \end{aligned}$$

em que $s^0 = \min_i \{y_i^0 r_i\}$. Vemos assim que o processo estocástico $\{\underline{Y}(t)\}$ é uma cadeia de Markov em tempo contínuo, sendo o seu espaço de estados representado por $\mathcal{N} = \prod_{i=1}^S \{0, 1, \dots, n_i\}$. Esta cadeia não se decompõe por estação porque a taxa à qual ocorrem avarias numa qualquer estação depende do número de máquinas não avariadas nas outras, de acordo com (2.4) e (2.1).

Estando interessados em calcular o valor esperado da taxa de produção do sistema ($E[s]$) e em conhecer a respectiva distribuição de probabilidade, é necessário resolver o sistema em ordem às probabilidades estacionárias da

cadeia $\{\underline{Y}(t)\}$. O número de estados de \mathcal{N} é dado por $\prod_{i=1}^S (n_i + 1)$, e assim um processo $\{\underline{Y}(t)\}$ representando um sistema com dez estações e três máquinas por estação teria mais de um milhão de estados; apesar do carácter esparsa da matriz de transições, a determinação da solução exacta é praticamente impossível mesmo para sistemas relativamente pequenos. Vemos assim a necessidade de dispor de um algoritmo eficiente que permita obter uma distribuição de probabilidade aproximada para a taxa de produção do sistema.

Para reduzir o número de estados procedemos à sua agregação. Concretamente, todos os estados $[y_i]$ tais que a taxa de produção do sistema seja determinada pela estação j e assumam o valor s são agregados no estado $[s, j]$. O processo agregado é representado por $\{Z(t), t \geq 0\}$, sendo \mathcal{M} o seu espaço de estados. Em qualquer momento, a estação que determina a taxa de produção do sistema é aquela que apresenta a taxa de produção potencial mínima, e se existirem duas ou mais estações em tal situação supõe-se que a responsável pela taxa do sistema é a que tem menor índice, ou seja, se $Z(t) = [s, j]$ então $j = \min\{i : y_i r_i = s\}$. Esta regra simples, embora arbitrária, permitiu obter resultados altamente satisfatórios. O estado correspondente à taxa de produção máxima ($s(t) = r_{\max}$) é representado por $[r_{\max}, *]$. Este caso é tratado de forma diferente porque não existe apenas uma estação responsável pela taxa máxima de produção, já que o sistema só produz à taxa r_{\max} se todas as estações tiverem uma taxa de produção potencial igual ou superior a r_{\max} . Uma forma alternativa de tratar este caso especial seria considerar a existência fictícia de uma estação 0, a qual teria uma máquina perfeitamente fiável cuja taxa de produção fosse r_{\max} .

A dimensão de \mathcal{M} é bastante mais pequena do que a do conjunto \mathcal{N} ; na verdade, $(1 + \sum_{i=1}^S n_i)$ constitui um limite superior ao número de estados do processo $\{Z(t)\}$. Para o sistema que já serviu de ilustração, com dez estações e três máquinas por estação, o processo teria no máximo trinta e um estados.

Existe uma relação entre \mathcal{N} , o espaço de estados da cadeia $\{\underline{Y}(t)\}$, e \mathcal{M} , o espaço de estados do processo $\{Z(t)\}$. Esta relação, representada por $\mathcal{F}(\cdot)$, é definida da seguinte maneira:

$$\mathcal{F}([y_i]) = \begin{cases} [r_{\max}, *] & \text{se } y_i r_i \geq r_{\max}, \forall i \\ [s, j], & \text{se } y_i r_i \geq s, \\ & \forall i \text{ e } j = \min \{ i : y_i r_i = s < r_{\max} \} \end{cases} \quad (2.5)$$

Sejam $\phi([s, j], [u, k])$ as taxas de transição do processo $\{Z(t)\}$. Segundo a teoria da agregação para cadeias de Markov devida a Takahashi, estas taxas de transição são obtidas da seguinte forma (Schweitzer [1984], Takahashi [1975]):

$$\phi([s, j], [u, k]) = \sum_{[y'_i]: \mathcal{F}([y'_i])=[s,j]} p([y'_i]) \sum_{[y_i]: \mathcal{F}([y_i])=[u,k]} v([y'_i], [y_i]) \quad (2.6)$$

em que $[y_i]$ e $[y'_i]$ são estados de $\{Y(t)\}$ tais que $\mathcal{F}([y_i]) = [u, k]$ e $\mathcal{F}([y'_i]) = [s, j]$, respectivamente, e $p([y'_i])$ é a probabilidade condicional de estar no estado $[y'_i]$ dado que $[y'_i]$ pertence ao conjunto de estados que correspondem ao bloco $[s, j]$, ou seja,

$$p([y'_i]) = \frac{\pi([y'_i])}{\sum_{[y_k]: \mathcal{F}([y_k])=[s,j]} \pi([y_k])}, \text{ para } [y'_i] : \mathcal{F}([y'_i]) = [s, j] \quad (2.7)$$

Pode-se ver em (2.6) que as taxas de transição de $\{Z(t)\}$ são médias ponderadas de algumas das taxas de transição $\{Y(t)\}$. Se os coeficientes de ponderação $p([y'_i])$ forem calculados com base nos valores exactos das probabilidades estacionárias $\pi([y'_i])$ de $\{Y(t)\}$, segundo (2.7), então as taxas de transição (2.6) são exactas, no sentido de que o processo $\{Z(t)\}$ pode ser considerado Markov para o efeito da determinação das respectivas probabilidades estacionárias; dado que os valores de $\pi([y'_i])$ não estão disponíveis têm de ser utilizadas estimativas, sendo assim obtidos valores aproximados quer para as taxas de transição quer para as probabilidades estacionárias de $\{Z(t)\}$. A próxima secção descreve em pormenor o método de agregação e decomposição utilizado para obter estas taxas aproximadas.

3. Método de Agregação e Decomposição

Nesta secção o método de agregação e decomposição é descrito em pormenor. Desde logo é de notar que, dada a estrutura da cadeia de Markov $\{Y(t)\}$, existe no máximo uma transição entre um dado estado $[y'_i]$ no bloco $[s, j]$ e todos os estados $[y_i]$ tais que $\mathcal{F}([y_i]) = [u, k]$, a qual corresponde ou a uma avaria ou a uma reparação na estação j ou k . Assim a expressão (2.6) pode-se escrever

$$\hat{\phi}([s, j], [u, k]) = \sum_{[y'_i]: \mathcal{F}([y'_i])=[s, j]} p([y'_i]) v([y'_i], [y_i]) \quad (3.1)$$

em que $\hat{\phi}$ representam as taxas de transição aproximadas de $\{Z(t)\}$ e, por simplicidade de notação, $[y_i]$ representa o estado no bloco $[u, k]$ que comunica com o estado $[y'_i]$.

As transições no processo $\{Z(t)\}$ também são causadas por avarias e reparações ocorridas nas estações. Uma avaria na estação i pode provocar uma transição de um estado $[s, i]$ ou $[s, j]$ para outro estado $[u, i]$, isto é, se uma avaria ocorrida na estação i provocar uma transição em $\{Z(t)\}$ então esta estação torna-se o estrangulamento do sistema. Uma reparação na estação i pode provocar uma transição de um estado $[s, i]$ para outro estado $[u, i]$ ou $[u, k]$, isto é, após uma reparação na estação i que provoque uma transição em $\{Z(t)\}$, a estação i pode (ou não) continuar a ser o estrangulamento do sistema.

Estamos interessados em estimar as ponderações $p([y'_i])$ para as transições causadas pelas diferentes estações. Para tal decomposmos o sistema por estações, ou seja, consideramos as estações uma de cada vez tomando todas as outras ("o resto do sistema") de forma agregada. O resto do sistema em relação à estação i compreende todas as estações $j, j \neq i$, e apresenta um comportamento (quanto a transições) semelhante ao do próprio sistema. No que se segue, o expoente (i) representa algo relativo ao resto do sistema em relação à estação i .

Definamos o processo estocástico em tempo contínuo $\{Z^{(i)}(t), t \geq 0\}$, para todo o i , em que $Z^{(i)}(t)$ representa a taxa de produção do resto do sistema em relação à estação i no momento t e a estação responsável por tal taxa, ou

seja $Z^{(i)}(t) = [s^{(i)}(t), j]$, em que $s^{(i)}(t) = \min_{\forall k \neq i} \{y_k(t)r_k\}$ e $j = \min\{l : l \neq i \text{ e } y_l(t)r_l = s^{(i)}(t)\}$. Sejam $\phi^{(i)}([s, j], [u, k])$, para $j, k \neq i$, as taxas de transição do processo $\{Z^{(i)}(t)\}$. Consideremos agora um processo estocástico em tempo contínuo associado a cada estação i , $\{Z_i(t), t \geq 0\}$, tal que $Z_i(t) = (Y_i^{(t)}, Z^{(i)}(t))$. O espaço de estados de $\{Z_i(t)\}$ é dado pelo produto cartesiano dos espaços de $\{Y_i(t)\}$ e de $\{Z^{(i)}(t)\}$, podendo este processo apresentar transições que tanto sejam devidas à estação i como ao resto do sistema.

Vamos agora argumentar que $\{Z_i(t)\}$ não goza da propriedade de Markov. Suponhamos que no momento t_1 $Y_i(t_1) = y_i$ e $Z^{(i)}(t_1) = [s^{(i)}, j]$, de modo que $Z_i(t_1) = (y_i, [s^{(i)}, j])$, e consideremos o processo $\{Z_i(t)\}$ no momento $t_1 + \Delta$, $\Delta > 0$. Se, durante o intervalo Δ , todas as transições forem provocadas pela estação i (isto é, o resto do sistema manteve-se no estado $[s^{(i)}, j]$) então $\Pr\{Z_i(t_1 + \Delta) = (y_i', [s^{(i)}, j])\}$ depende do estado em que o processo se encontrava no momento t_1 mas é independente dos estados visitados antes de t_1 . Por outro lado, se alguma(s) transição(ões) ocorreram no resto do sistema durante Δ então $\Pr\{Z_i(t_1 + \Delta) = (y_i', [u^{(i)}, k])\}$ depende não só do estado $Z_i(t_1)$ mas também dos estados visitados antes de t_1 , uma vez que estes determinam o novo estado $[u^{(i)}, k]$ do resto do sistema. Vemos assim que $\{Z_i(t)\}$ não apresenta a propriedade da "perda de memória". Vamos, porém, pressupor que os processos estocásticos $\{Z_i(t)\}$, para todo o i , são Markov.

Sejam as taxas de transição (variáveis no tempo) do processo $\{Z_i(t)\}$ representadas por $p_{i,t}[(y_i^1, [s^{(i)}, j]), (y_i^2, [u^{(i)}, k])]$, em que $y_i^1 = y_i^2$ ou $[s^{(i)}, j] = [u^{(i)}, k]$. De acordo com o pressuposto anteriormente enunciado, estas taxas serão consideradas invariantes no tempo e as suas estimativas serão representadas por \hat{p}_i . Sejam as probabilidades estacionárias do processo $\{Z_i(t)\}$ representadas por π_i , ou seja, $\pi_i(y_i, [s^{(i)}, j]) = \lim_{t \rightarrow \infty} \Pr\{Z_i(t) = (y_i, [s^{(i)}, j])\}$. Os valores aproximados destas probabilidades serão representados por $\hat{\pi}_i(y_i, [s^{(i)}, j])$.

Quando os processos $\{Y_i(t)\}$ e $\{Z^{(i)}(t)\}$ são combinados no processo $\{Z_i(t)\}$ é necessário proceder a algumas correcções nas taxas de transição

correspondentes a avarias dado que, no sistema como um todo, as taxas de avarias são proporcionais à taxa de produção do sistema $\min\{y_i r_i, s^{(i)}\}$, e não $y_i r_i$ ou $s^{(i)}$ como é suposto nos processos isolados. Assim, as taxas de transição \hat{p}_i são estimados da seguinte maneira:

$$\hat{p}_i[(y_i, [s^{(i)}, j]), (y_i - 1, [s^{(i)}, j])] = \theta(y_i, [s^{(i)}, j]) f_i, \forall y_i > 0 \quad (3.2a)$$

$$\hat{p}_i[(y_i, [s^{(i)}, j]), (y_i + 1, [s^{(i)}, j])] = (n_i - y_i) \mu_i, \forall y_i < n_i \quad (3.2b)$$

$$\hat{p}_i[(y_i, [s^{(i)}, j]), (y_i, [u^{(i)}, k])] = \hat{\phi}^{(i)}([s, j], [u, k]) \gamma_{y_i}([s, j], [u, k]) \quad (3.2c)$$

As transições em (3.2a) são causadas por avarias ocorridas na estação i , e os factores de correcção $\theta(y_i, [s^{(i)}, j])$ (que representam a taxa de produção do sistema) são dados por

$$\theta(y_i, [s^{(i)}, j]) = \min\{y_i r_i, s^{(i)}\} \quad (3.3)$$

As transições em (3.2b) são causadas por reparações que tiveram lugar na estação i , e não é necessária qualquer correcção uma vez que as taxas de reparações não dependem da taxa de produção do sistema. As transições em (3.2c) são causadas pelo resto do sistema (em relação à estação i), e os factores de correcção são dados por

$$\gamma_{y_i}([s, j], [u, k]) = \begin{cases} \frac{y_i r_i}{s}, & \text{se } s > u \text{ e } s \geq y_i r_i \\ 1, & \text{se } s \leq u \text{ ou } s < y_i r_i \end{cases} \quad (3.4)$$

Isto implica que, quando a estação i é o estrangulamento do sistema, as taxas de avarias são diminuídas pela razão entre a taxa de produção do sistema (igual à da estação i) e a taxa de produção do resto do sistema, ao passo que quando a taxa de produção do sistema é determinada pelo resto do sistema as taxas de avarias se mantêm inalteradas. O uso destes factores de correcção, embora baseado em argumentos puramente heurísticos, traduziu-se numa melhoria significativa do algoritmo e da precisão das aproximações. As taxas de transição (não ajustadas) do processo $\{Z^{(i)}(t)\}$ representativo do resto do sistema em relação à estação i são estimadas de um modo semelhante ao que é

usado para as taxas de transição do processo $\{Z(t)\}$, embora ignorando as transições causadas pela estação i .

As taxas de transição $\hat{\phi}$ do processo $\{Z(t)\}$ que são causadas pela estação i são estimadas através das seguintes expressões, em que as frações representam os coeficientes de ponderação apresentados em (2.6) e (2.7):

$$\hat{\phi}([y_i r_i, i], [(y_i - 1)r_i, i]) = \frac{\sum_{v \geq y_i r_i} \hat{\pi}_i(y_i, [v^{(i)}, k])}{\Pr([y_i r_i, i])} (y_i r_i) f_i \quad (3.5a)$$

$$\hat{\phi}([y_i r_i, i], [(y_i + 1)r_i, i]) = \frac{\sum_{v > (y_i + 1)r_i} \hat{\pi}_i(y_i, [v^{(i)}, k])}{\Pr([y_i r_i, i])} (n_i - y_i) \mu_i \quad (3.5b)$$

$$\hat{\phi}([y_i r_i, i], [u, k]) = \frac{\hat{\pi}_i(y_i, [u^{(i)}, k])}{\Pr([y_i r_i, i])} (n_i - y_i) \mu_i \quad (3.5c)$$

$$\hat{\phi}([s, j], [y_i r_i, i]) = \frac{\hat{\pi}_i(y_i + 1, [s^{(i)}, j])}{\Pr([s, j])} s f_i \quad (3.5d)$$

As expressões (3.5a, b) correspondem a avarias e reparações ocorridas na estação i que não provocam qualquer alteração na estação responsável pela taxa de produção do sistema, as (3.5c) a reparações na estação i que tornam outra estação a responsável pela taxa de produção do sistema, e as (3.5d) a avarias na estação i que a transformam no estrangulamento do sistema. Considerando transições causadas por todas as estações i , para $i = 1, \dots, S$, são obtidas todas as taxas de transição $\hat{\phi}$ do processo $\{Z(t)\}$ (assim como as taxas $\hat{\phi}^{(i)}$ dos processos $\{Z^{(i)}(t)\}$).

Os denominadores em (3.5) representam os valores relativos das probabilidades do processo $\{Z(t)\}$ estar num determinado estado, ou de a cadeia $\{Y(t)\}$ estar em determinado bloco. São estimados da seguinte maneira:

$$\Pr([r_{\max}, *]) = \frac{1}{S} \sum_{k=1}^S \left[\sum_{y_k: y_k r_k \geq r_{\max}} \hat{\pi}_k(y_k, [r_{\max}, *]) \right] \quad (3.6a)$$

$$\Pr([y_i r_i, i]) = \sum_{v^{(i)} > y_i r_i} \hat{\pi}_i(y_i, [v^{(i)}, k]) + \sum_{\substack{v^{(i)} = y_i r_i \\ i < k}} \hat{\pi}_i(y_i, [v^{(i)}, k]) \quad (3.6b)$$

$\Pr([s, j])$ reduz-se a (3.6a) quando $s = r_{\max}$ e a (3.6b) quando $s < r_{\max}$. Dado que cada processo $\{Z_i(t)\}$ fornece uma estimativa da probabilidade da taxa de produo do sistema ser r_{\max} , (3.6a) utiliza tais estimativas ao calcular a respectiva mdia. Na expresso (3.6b) a probabilidade do sistema estar a produzir a uma dada taxa $[y_i r_i, i]$  calculada considerando apenas as probabilidades do processo $\{Z_i(t)\}$. A primeira parcela corresponde s probabilidades de a taxa de produo do resto do sistema ser superior a $y_i r_i$ e de y_i mquinas no estarem avariadas (fazendo com que a taxa de produo do sistema seja $y_i r_i$ e determinada pela estaco i). A segunda parcela corresponde ao caso de duas estaces (i e k) constituirem ao mesmo tempo o estrangulamento do sistema e determinarem uma taxa de produo igual a $y_i r_i$; de acordo com a regra anteriormente definida, a estaco i  considerada o estrangulamento se e so se $i < k$. Estas probabilidades, depois de normalizadas, constituem um bom conjunto de valores iniciais para as probabilidades estacionrias $\hat{\pi}([s, j])$ de $\{Z(t)\}$.

Uma vez que se pressupe que os processos $\{Z(t)\}$ e $\{Z_i(t)\}$ para todo o i so Markov quando na verdade no o so, as probabilidades $\hat{\pi}([s, j])$ no so exactas mas meras aproximaes. , no entanto, curioso notar que no caso das configuraes em srie pura e em paralelo puro o mtodo se torna exacto. Na verdade, nestes casos o processo $\{Z(t)\}$  Markov porque, excluindo os estados transientes de $\{Y(t)\}$, existe uma correspondncia biunvoca entre os estados de $\{Y(t)\}$ e os estados de $\{Z(t)\}$. Daqui decorre que os processos $\{Z_i(t)\}$ tm so Markov, e ento as probabilidades $\hat{\pi}(y_i, [u^{(i)}, j])$ estimam de forma correcta $\pi([y_i])$. Ento, as taxas de transio dadas por (3.5) so exactas, e tm so as probabilidades $\pi([s, j])$.

4. Algoritmo de Resoluo e Resultados Computacionais

Nesta seco o mtodo desenvolvido na Seco 3  apresentado sob a forma de algoritmo, e so discutidos os resultados computacionais levados a cabo para testar a qualidade das aproximaes e o desempenho do mtodo. O algoritmo  o seguinte:

Inputs:

1. Nmero de estaes (S);
2. Para cada estao $i = 1, \dots, S$:
 - 2.1. Nmero de mquinas, n_i ;
 - 2.2. Taxa de produo de cada mquina, r_i ;
 - 2.3. Taxa de avarias de cada mquina, f_i ;
 - 2.4. Taxa de reparaes de cada mquina, μ_i .

Inicializao:

1. Identificar todos os estados $[s, j]$ de $\{Z(t)\}$.
2. Inicializar as probabilidades $\hat{\pi}(y_i, [s^{(i)}, j])$, para todo o i , como segue:

$$\hat{\pi}(y_i, [s^{(i)}, j]) = \begin{cases} \frac{1}{n_i+1}, & \text{se } [s^{(i)}, j] = [r_{\max}, *] \\ 0, & \text{caso contrrio} \end{cases}$$

3. Calcular as taxas de transio $\hat{\phi}([s, j], [u, k])$, determinadas por todas as estaes i , de acordo com (3.5).

Iterao Bsica:

Repetir o seguinte ciclo at que as probabilidades $\hat{\pi}(y_i, [s^{(i)}, j])$ converjam para todos os estado $(y_i, [s^{(i)}, j])$ e para todas as estaes i .

Para cada estao $i = 1, \dots, S$:

1. Calcular $\hat{\phi}^{(i)}([s, j], [u, k])$ de acordo com (3.5).
2. Calcular $\gamma_{y_i}([s, j], [u, k])$ de acordo com (3.4).
3. Calcular as taxas de transio \hat{p}_i de acordo com (3.2).
4. Resolver em ordem s probabilidades estacionrias $\hat{\pi}_i(y_i, [s^{(i)}, j])$.
5. Actualizar as taxas de transio $\hat{\phi}([s, j], [u, k])$ de acordo com (3.5).

Passo Final:

1. Inicializar as probabilidades $\hat{\pi}[s, j]$ normalizando as probabilidades dadas por (3.6).
2. Resolver em ordem às probabilidades estacionárias $\hat{\pi}[s, j]$ de $\{Z(t)\}$.

A técnica da uniformização para cadeias de Markov em tempo contínuo (Keilson [1979]) foi utilizada para obter os valores das probabilidades estacionárias dos processos $\{Z(t)\}$ e $\{Z_i(t)\}$, para todo o i . Para o caso de sistemas relativamente fiáveis, foram necessárias de cinco a sete iterações básicas para que os valores de $\hat{\pi}_i(y_i, [s^{(i)}, j])$ convergissem com uma precisão de três casas decimais; para o caso de sistemas menos fiáveis foram necessárias de oito a dez iterações básicas. Em cada iteração apenas foi permitido um máximo de quarenta substituições para cada $i = 1, \dots, S$. No passo final foram geralmente necessárias menos de cem iterações para conseguir que as probabilidades convergissem com uma precisão de cinco casas decimais. O número de iterações neste passo final tende a crescer com o número de estações do sistema.

Para validar o algoritmo é necessário comparar os resultados aproximados com os valores exactos. Estes apenas podem ser obtidos para sistemas de dimensão pequena e média, e para tal há que calcular as probabilidades estacionárias da cadeia $\{Y(t)\}$. Foi utilizado o esquema de substituições sucessivas e limites iterativos apresentado em Schweitzer e van der Wal [1987]. O desempenho computacional deste esquema é muito bom, embora dependa em larga medida do grau de fiabilidade do sistema. Na verdade, o estado de sistema perfeito (isto é, inexistência de máquinas avariadas) foi escolhido para ser o "estado 0" no referido esquema e o método é bastante sensível à escolha de tal estado, no sentido de que quanto mais provável ele for menos iterações são necessárias para conseguir a convergência desejada. Para os sistemas mais fiáveis foram geralmente necessárias menos de trinta iterações para atingir convergência com uma precisão de quatro casas decimais, ao passo que para os sistemas menos fiáveis foram necessárias entre sessenta e cem iterações.

Para sistemas de grande dimensão é impossível obter os valores exactos das probabilidades, e então os resultados aproximados são comparados com os resultados dados pela simulação do sistema. Nesta simulação foi utilizado um gerador de números aleatórios congruencial com módulo $2^{31} - 1$ e multiplicador 397204094, o qual, segundo Fishman e Moore [1982], dá excelentes resultados.

O método de agregação e decomposição foi extensivamente testado com o objectivo de avaliar a qualidade dos resultados aproximados. Os sistemas considerados tinham entre uma e quatro máquinas por estação, variando o número de estações entre duas e trinta e seis. Em todos os casos a taxa máxima de produção do sistema era de dez artigos por unidade de tempo. Foram também considerados dois cenários: o dos sistemas mais fiáveis, em que as taxas de avarias e de reparações eram tais que a taxa média de produção fosse cerca de 65% a 75% do máximo, e o dos sistemas menos fiáveis, em que as taxas de reparações foram empiricamente ajustadas de modo a obter uma taxa média de produção de cerca de 35% a 45% do máximo.

Começamos por estudar um pequeno sistema com três estações com duas, três, e duas máquinas, respectivamente. O valor esperado da taxa de produção era de 6.4489 artigos por unidade de tempo, sendo o valor aproximado de 6.3712. O erro era 0.0777, ou 1.20%. Com o objectivo de analisar o desempenho do método para sistemas com características diversas, foram comparados os valores exactos e aproximados para diferentes valores dos parâmetros de input, isto é, as taxas de avarias e de reparações.

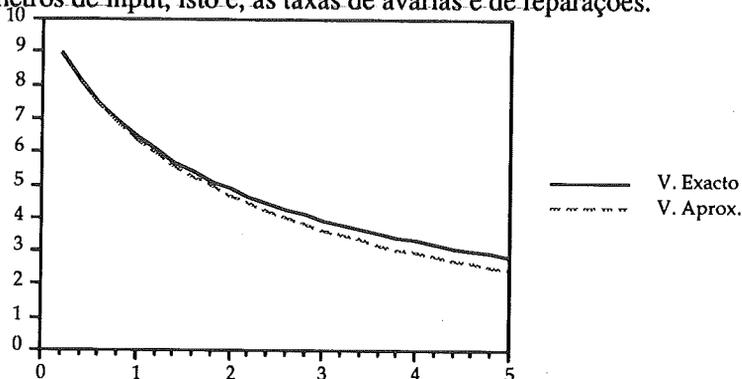


Figura 1

Valor Esperado da Taxa de Produção vs. $k \times$ Taxas de Avarias para um Sistema Pequeno

Na Figura 1 estão representados graficamente ambas as taxas médias de produção (exacta e aproximada) para diferentes valores da taxa de avarias, tendo todas elas sido modificadas na mesma proporção (ou sejam, foram multiplicadas pelo mesmo factor k). É evidente que quanto mais fiável for melhor é a solução aproximada. O erro máximo é 0.4436 artigos por unidade de tempo (ou 15.61%) e o erro mínimo é 0.0021 (ou 0.02%), sendo o gráfico convexo. A Figura 2 apresenta um gráfico similar, mas desta feita para diferentes valores das taxas de reparação, sendo o gráfico concavo. Mais uma vez é bem patente que a solução aproximada está mais próxima do valor exacto quanto mais fiável o sistema for. O erro máximo é 0.4436 artigos por unidade de tempo (ou 15.61%) e o erro mínimo é 0.0021 (ou 0.02%). Estes valores coincidem com os obtidos para a figura anterior dado que em ambos os casos estamos a considerar sistemas cinco vezes mais fiáveis ($k = 5$) e cinco vezes menos fiáveis ($k = 0.2$) do que a situação dos valores base, sendo apenas diferentes porque enquanto na Figura 1 são manipuladas as taxas de avarias, na Figura 2, são ajustadas as taxas de reparações.

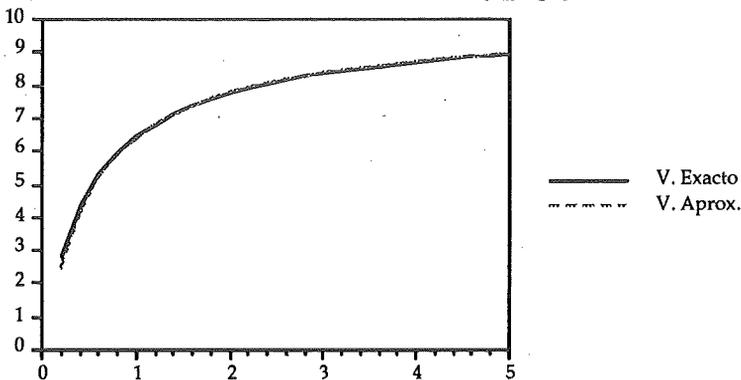


Figura 2

Valor Esperado da Taxa de Produção vs. $k \times$ Taxas de Reparções para um Sistema Pequeno

Passamos então a considerar sistemas de dimensão média, podendo ter de duas até dez estações. Para estes sistemas tanto os valores exactos como os aproximados dos primeiro e segundo momentos da distribuição da taxa de produção foram obtidos, tendo isto sido feito para dois cenários: o dos sistemas mais fiáveis e o dos sistemas menos fiáveis. O maior sistema para o qual foi obtida a solução exacta tinha perto de 83000 estados; no cenário mais fiável o tempo de computação foi de 1165.5 segundos de CPU (num IBM

4361), ao passo que no cenário menos fiável foram precisos 2146.9 segundos.

Na Figura 3 o valor esperado (exacto e aproximado) da taxa de produção está representado graficamente à medida que o número de estações vai aumentando. É óbvio que o erro dos resultados aproximados é menor quanto mais estações o sistema tem. O erro também é menor para o cenário mais fiável; na verdade, para este caso o erro médio é 0.0554 artigos por unidade de tempo (ou 0.84%), sendo 0.2941 (ou 7.55%) no cenário menos fiável.

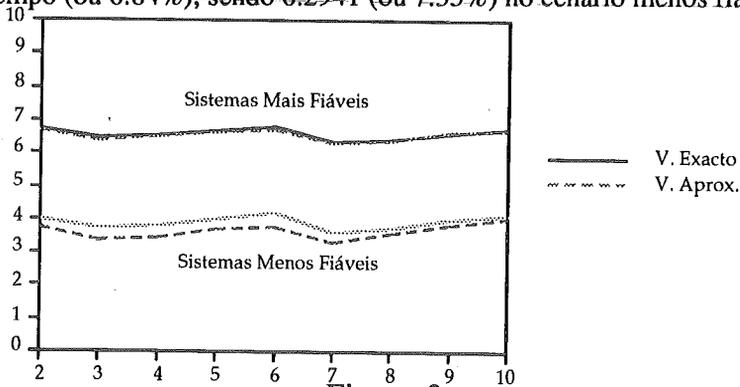


Figura 3

Valor Esperado da Taxa de Produção vs. Número de Estações para Sistemas de Dimensão Média

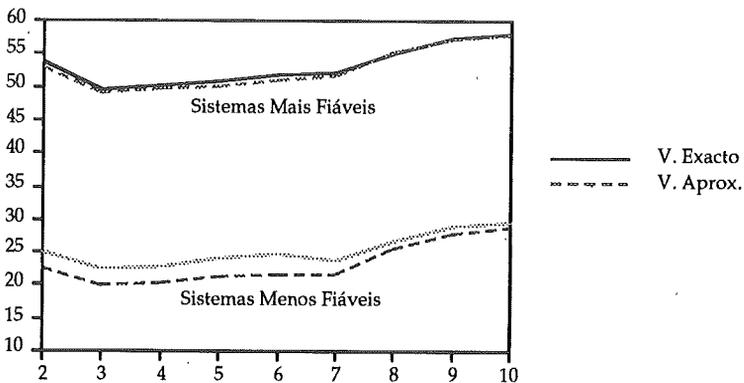


Figura 4

Segundo Momento da Distribuição da Taxa de Produção vs. Número de Estações para Sistemas de Dimensão Média

A Figura 4 apresenta um gráfico similar para o segundo momento da distribuição, sendo fácil verificar as mesmas conclusões. No cenário mais fiável o erro médio é 0.5946 (artigos)² por unidade de tempo (ou 1.12%) e no menos fiável esse valor é 2.2717 (ou 8.96%).

A complexidade do algoritmo depende grandemente do número de estações, sendo também importante o número de máquinas em cada uma. Estes mesmo parâmetros determinam igualmente a dimensão do espaço de estados do processo $\{Y(t)\}$, e desse modo a complexidade do algoritmo que obtém a solução exacta. A Figura 5 exhibe o gráfico dos tempos de computação (em segundos de CPU num IBM 4361) necessários à obtenção das soluções exacta e aproximada para o cenário mais fiável, sendo clara a vantagem do método desenvolvido para sistemas com mais de seis/oito estações. No cenário menos fiável a obtenção da solução exacta requer bastante mais tempo devido à escolha do "estado 0", tal como foi previamente explicado. O tempo de computação para o método de agregação e decomposição à medida que a fiabilidade do sistema diminui também aumenta, mas de forma mais moderada.

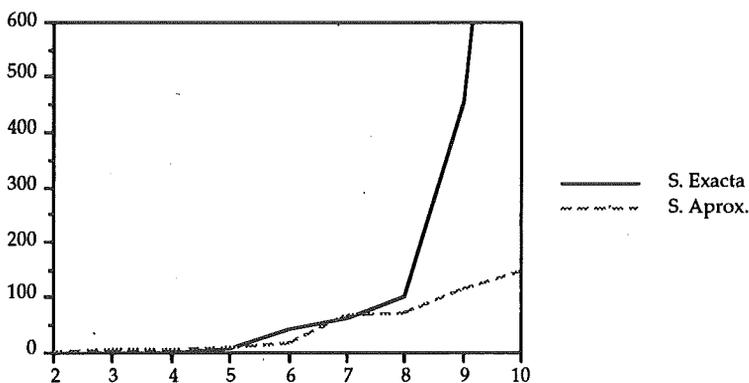


Figura 5
Tempo de Computação (segundos de CPU) vs. Número de Estações para Sistemas Mais Fiáveis de Dimensão Média

Por fim, o método de agregação e decomposição foi testado para sistemas de grande dimensão, atingindo trinta e seis estações. Para tais sistemas os resultados aproximados foram comparados com valores dados por uma simulação do sistema. Todas as simulações gastaram mais de 2147 segundos de CPU (o máximo tempo que pode ser medido). Para sistemas com doze e dezasseis estações o método desenvolvido requereu 300.4 e 972.4 segundos de CPU, respectivamente; não foi possível obter o tempo de computação para sistemas com vinte ou mais estações, mas verificou-se que requeria bastante menos tempo que a simulação. Os resultados dados pela simulação levaram à

construção de intervalos de confiança a 95%, o que permite estabelecer limites superior e inferior ao erro cometido.

Os Quadros 1 e 2 apresentam os resultados para os primeiro e segundo momentos da distribuição da taxa de produção, respectivamente. Os resultados aproximados são bastante bons, encontrando-se sempre entre os limites dos intervalos de confiança a 95% dados pela simulação.

Número de Estações	Valor Esperado da Taxa de Produção (Unidades: artigos por unidade de tempo)		
	Aproximação	Simulação	Erro
12	6.9524	[6.9224, 6.9909]	[-0.0300, +0.0385]
16	7.3888	[7.3574, 7.4325]	[-0.0314, +0.0437]
20	7.5783	[7.5241, 7.5930]	[-0.0542, +0.0147]
24	7.4489	[7.4274, 7.4845]	[-0.0215, +0.0356]
28	7.2888	[7.2499, 7.3088]	[-0.0389, +0.0200]
32	7.1692	[7.1601, 7.1953]	[-0.0091, +0.0261]
36	7.1752	[7.1547, 7.2034]	[-0.0205, +0.0282]

Quadro 1.

Valor Esperado da Taxa de Produção para Sistemas de Grande Dimensão

Número de Estações	Segundo Momento da Distribuição da Taxa de Produção (Unidades: (artigos) ² por unidade de tempo)		
	Aproximação	Simulação	Erro
12	60.1736	[59.9405, 60.5710]	[-0.2331, +0.3974]
16	64.6091	[64.3275, 64.9958]	[-0.2816, +0.3867]
20	66.8091	[66.3598, 66.9915]	[-0.4493, +0.1824]
24	64.3398	[64.2294, 64.7243]	[-0.1104, +0.3845]
28	61.6727	[61.3354, 61.8821]	[0.3374, +0.2094]
32	60.7368	[60.6437, 60.9673]	[-0.0931, +0.2305]
36	60.4202	[60.2882, 60.7122]	[-0.1320, +0.2920]

Quadro 2.

Segundo Momento da Distribuição da Taxa de Produção para Sistemas de Grande Dimensão

5. Conclusões

Esta comunicação abordou o problema da estimação da taxa média de produção de sistemas industriais com máquinas não fiáveis e uma configuração em série-paralelo. Estas configurações têm vindo a suscitar cada vez mais interesse, especialmente se são utilizados equipamentos automatizados. O conhecimento da distribuição da taxa de produção em tais casos pode ser muito útil para comparar o desempenho de configurações alternativas e para ajudar a configurar o sistema da melhor forma.

Pressupondo que as taxas de avarias e de reparações eram invariantes no tempo foi possível modelar o sistema como uma cadeia de Markov em tempo contínuo; a dimensão do espaço de estados, porém, torna impossível obter soluções exactas para sistemas reais. Foi, então, desenvolvido um método de agregação e decomposição que permite obter soluções aproximadas. Embora se tenha pressuposto que todas as avarias eram causadas pelo funcionamento das máquinas, é fácil estender o método de modo a incorporar outros tipos de avarias.

As experiências numéricas levadas a cabo mostraram que o método desenvolvido permite obter boas aproximações aos valores exactos. As aproximações tornam-se melhores à medida que a fiabilidade do sistema e o número de estações aumenta. O esforço computacional necessário à obtenção da solução aproximada é bastante menor do que o necessário à obtenção da solução exacta ou o da simulação do sistema, tornando assim possível o estudo de sistemas de grande dimensão. Contudo, este tempo de computação depende grandemente do número de estações e do número de máquinas por estação.

Vimos que quando as taxas de avarias diminuem a taxa média de produção aumenta, sendo os rendimentos marginais crescentes e sendo este efeito mais forte quando todas as taxas sofrem essa diminuição. A taxa média de produção também aumenta quando as taxas de reparações aumentam, mas desta vez os rendimentos marginais são decrescentes; tal como no caso anterior, este efeito é mais acentuado quando todas as taxas são aumentadas. O conhecimento deste facto é particularmente importante ao tomar decisões com o objectivo de aumentar a taxa de produção do sistema. Aumentar às

taxas de reparações parece ser a melhor escolha para sistemas menos fiáveis, e à medida que a taxa de produção cresce passa a valer a pena diminuir às taxas de avarias.

É de notar que as aproximações são de um modo geral conservadoras, pois o método subestima quase sempre o valor esperado exacto da taxa de produção do sistema.

6. Bibliografia

- [1] ALVES, R. [1988], *Performance Evaluation of Continuous Flow Manufacturing Systems with Parallel Machines*, tese de doutoramento, William E. Simon Graduate School of Business Administration, University of Rochester, Rochester, NY 14627.
- [2] ALVES, R. [1989], *Performance Evaluation of Series-Parallel Systems*, submetido e aceite para publicação no JMOM – Journal of Manufacturing and Operations Management.
- [3] BAYBARS, I. [1984], *A Survey of Inexact Algorithms for the Simple Assembly Line Balancing Problem*, Working Paper, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, Pennsylvania.
- [4] BAYBARS, I. [1986], *A Survey of Exact Algorithms for the Simple Assembly Line Balancing Problem*, Management Science, Vol. 32, Nº 8, pp.909-932.
- [5] BOOKBINDER, J.H. e KOTWA, T.R. [1987], *Modeling an AGV Automobile Body-Framing System*, Interfaces, Vol. 17, Nº 6, pp.41-50.
- [6] BOWMAN, E.H. [1960], *Assembly Line Balancing by Linear Programming*, Operations Research, Vol. 8, pp.385-389.
- [7] BUZACOTT, J.A. [1967], *Automatic Transfer Lines with Buffer Stocks*, International Journal of Production Research, Vol. 5, Nº 3, pp.183-200.

- [8] BUZACOTT, J.A. [1968], *Prediction of the Efficiency of Production Systems Without Internal Storage*, International Journal of Production Research, Vol. 6, N° 3, pp.173-188.
- [9] BUZACOTT, J.A. e HANIFIN, L.E. [1978], *Models of Automatic Transfer Lines with Inventory Banks—A Review and Comparison*, AIIE Transactions, Vol. 10, N° 2, pp197-207.
- [10] ELSAYED, E.A. e HWANG, C.C. [1986], *Analysis of Two-Stage Manufacturing Systems with Buffer Storage and Redundant Machines*, International Journal of Production Research, Vol. 24, N° 1, pp.187-201.
- [11] FISHMAN, G.G. e MOORE, L.R. [1982], *A Statistical Evaluation of Multiplicative Congruential Random Number Generators with Modulus $2^{31}-1$* , Journal of the American Statistical Association, Vol. 77, N° 377, pp.129-136.
- [12] IGNALL, E. e SILVER, A. [1977], *The Output of a Two-Stage System with Unreliable Machines and Limited Storage*, AIIE Transactions, Vol. 9, N° 2, pp.183-188.
- [13] JACKSON, J.R. [1956], *A Computing Procedure for a Line Balancing Problem*, Management Science, Vol. 2, pp.261-271.
- [14] KEILSON, J. [1979], *Markov Chain Models—Rarity and Exponentiality*, Springer-Verlag, New York.
- [15] KOSTER, M.B.M. de [1987], *Estimation of Line Efficiency by Aggregation*, International Journal of Production Research, Vol. 25, N° 4, pp.615-626.
- [16] KUBAT, P. e SUMITA, U. [1985], *Buffers and Backup Machines in Automatic Transfer Lines*, International Journal of Production Research, Vol. 23, N° 6, pp.1259-1270.
- [17] MASTOR, A.A. [1970], *An Experimental Investigation and Comparative Evaluation of Production Line Balancing Techniques*, Management Science, Vol. 16, pp.728-746.
- [18] OHMI, T. [1981], *An Approximation for the Production Efficiency of Automatic Transfer Lines with In-Process Storages*, AIIE Transactions, Vol. 13, N° 1, pp.22-28.

- [19] SALVESON, M.E. [1955], *The Assembly Line Balancing Problem*, Journal of Industrial Engineering, Vol. 6, pp.18-25.
- [20] SCHWEITZER, P.J. [1984], *Aggregation Methods for Large Markov Chains*, in Mathematical Computer Performance and Reliability, G.Iazeolla, P.J.Courtois, e A. Hordijk (editores), Elsevier North-Holland, pp.275-286.
- [21] SCHWEITZER, P.J. e WAL, J. van der [1987], *Iterative Bounds on the Equilibrium Distribution of a Finite Markov Chain*, Probability in the Engineering and Information Science, Vol. 1, Nº 1, pp.117-131.
- [22] TAKAHASHI, Y. [1975], *A Lumping Method for Numerical Calculations of Stationary Distributions of Markov Chains*, Research Report B-18, Department of Information Sciences, Tokyo Institute of Technology, Tokyo, Japan.
- [23] TALBOT, F.B., PATTERSON, J.H. e GEHRLEIN, W.V. [1986], *A Comparative Evaluation of Heuristic Line Balancing Techniques*, Management Science, Vol. 32, Nº 4, pp.430-454.

CONDIÇÕES NECESSÁRIAS DE OPTIMALIDADE PARA PROCESSOS COM TRAJECTÓRIAS DESCONTÍNUAS

Fernando Manuel Ferreira Lobo Pereira

DEEC-FEUP

Rua dos Bragas

4099 Porto Codex

INESC-N

Largo de Mompilher, 22

4000 Porto

Resumo

Apresentamos condições necessárias de optimalidade na forma de um Princípio do Máximo de Pontryagin para problemas de controlo óptimo que poderão admitir soluções cujas trajectórias de estado são descontínuas. Estas surgem, por exemplo, em aplicações aeroespaciais e económicas, onde determinadas variáveis de controlo podem idealmente ser consideradas ilimitadas.

Na nova formulação do problema de controlo óptimo que propomos, a dinâmica possui dois termos: um convencional e outro que reflecte a intervenção de controlos num espaço de medidas vectoriais.

Baseando-se na aproximação do problema dado por um outro convencional e na aplicação do Teorema de Ekeland, os argumentos por nós usados permitem hipóteses consideravelmente menos restritivas do que aquelas mediante as quais resultados similares tinham sido provados anteriormente. De facto, os elementos envolvidos na componente convencional do nosso problema quase satisfazem as hipóteses mais fracas até hoje reconhecidas como necessárias para deduzir condições de optimalidade para problemas com trajectórias absolutamente contínuas.

1. Introdução

A teoria do controlo óptimo convencional trata de problemas cuja dinâmica é representada por equações diferenciais ordinárias podendo portanto ter, como soluções, processos cujas trajectórias são apenas absolutamente contínuas.

Todavia existem determinados problemas para os quais é natural considerar soluções que apresentam descontinuidades. Algumas estratégias óptimas para a exploração de recursos naturais (ver Clark [1]) poderão incluir extracções de quantidades significativas de recursos num intervalo de tempo extremamente curto.

Também nalguns problemas de navegação aeroespacial (Lawden [5], Marec [6] e Neustadt [7]) se verifica que o consumo mínimo corresponde a trajectórias em que uma parte importante do combustível é consumida numa fracção muito pequena da duração total da missão. As trajectórias associadas ao tipo de estratégias descritas acima apresentam variações bruscas que, idealmente, são modelizadas por descontinuidades.

Uma vez que a teoria do controlo convencional trata apenas problemas com trajectórias absolutamente contínuas torna-se necessário criar uma nova formulação que permita derivar condições de optimalidade caracterizando as soluções da classe de problemas de que mencionamos alguns exemplos.

Após uma introdução às questões envolvidas através de um exemplo simples, apresentamos uma formulação do problema de controlo óptimo em que o espaço dos controlos tradicional (subconjunto de funções mensuráveis) foi alargado de modo a incluir elementos de um espaço de medidas passando, portanto, a trajectória a ser obtida por um integral de Lebesgue-Stieltjes. Esta extensão consiste em incluir, além da componente da dinâmica usual (a absolutamente contínua em relação á medida de Lebesgue), uma outra em que o controlo, eventualmente singular, entra de forma afim. Deste modo, esta formulação pode ser vista como uma extensão natural de um problema de controlo óptimo convencional (ver Neustadt [7]) em que algumas componentes do controlo que afectam a dinâmica poderão ser ilimitadas.

Seguidamente apresentamos condições necessárias de optimalidade para o problema de controlo óptimo por nós definido e comparamos o nosso resultado com outros anteriormente obtidos. A nossa metodologia (ver Vinter e Pereira [13] e Pereira [8]) é bastante diferente das usadas por outros autores (Rishel [10], Rockafellar [11] e [12] e Warga [14] e [15]) que no passado deduziram resultados semelhantes para o problema de controlo óptimo impulsional. Esta abordagem não só permite a obtenção das condições necessárias de optimalidade assumindo hipóteses bastante menos restritivas, como também permite restrições afins no estado e restrições na componente convencional do controlo variantes no tempo.

2. Caracterização do Problema

Considere-se o seguinte problema de consumo mínimo

$$\text{Minimizar } \int_0^1 u(t)dt \quad (2.1)$$

onde $u \in L^1(0, 1)$ satisfaz

$$\dot{x}(t) = Ax(t) + bu(t) \text{ q.s. } t \in [0, 1] \quad (2.2)$$

$$x(0) = 0$$

$$x(1) = x_1$$

$$u \in \mathcal{U} := \left\{ u \in L^1(0, 1) : u(t) \geq 0 \text{ e } \int_0^1 u(t)dt \leq M \right\}$$

sendo A, b, x_1 e $M \in \mathbb{R}$, $x_1 > 0$ e M suficientemente grande. $L^1(0, 1)$ denota o espaço das funções integráveis e "q.s. $t \in [0, 1]$ " significa que a relação qualificada é válida em subconjuntos de medida de Lebesgue total.

O problema consiste pois em dosear a taxa de consumo de combustível de modo a que o estado seja transferido entre dois pontos dados e, simultaneamente, a quantidade total consumida seja mínima.

Este problema aparentemente razoável deveria ter uma solução. Todavia tal não acontece pois não existe qualquer topologia de $L^1(0, 1)$ para o espaço dos controlos na qual o conjunto destes é compacto, o operador que associa a cada controlo uma trajectória é contínuo e o funcional custo é semicontínuo inferiormente.

Note-se que a sucessão de trajectórias correspondente à sucessão de controlos minimizante, $\{u_n\}$, onde

$$u_n = \begin{cases} \frac{Ae^{-A}x_1}{b(1-e^{-A/n})} & \text{se } t \leq 1/n \\ 0 & \text{se } t > 1/n \end{cases} \quad (2.3)$$

converge para uma trajectória que satisfaz as restrições mas apresenta uma descontinuidade na origem.

Esta trajectória limite pode ser vista como sendo a resposta a um controlo que apresenta um impulso de Dirac no instante inicial que, embora não seja admissível para o problema dado, o é para a respectiva formulação generalizada:

$$\text{Minimizar } \int_{[0,1]} \mu(ds) \tag{2.4}$$

tal que

$$x(t) = x(0) + \int_0^t Ax(s)ds + \int_{[0,t]} b\mu(ds) \text{ q.s. } t \in [0, 1] \tag{2.5}$$

$$x(0) = 0$$

$$x(1) = x_1 \text{ e}$$

$$\mu \in \mathbb{U}^*$$

onde \mathbb{U}^* é o conjunto das medidas regulares de Borel em $[0, 1]$ que se obtém tomando o fecho fraco estrela das medidas associadas com o elemento de \mathbb{U} (ou seja $A \rightarrow \int_A u(t)dt$).

Observe-se que, considerando a topologia indicada, argumentos simples de análise abstracta garantem a existência de solução.

3. Formulação Generalizada do Problema de Controlo Óptimo

O simples exemplo apresentado acima sugere uma nova formulação do problema de controlo óptimo que resulta da adição à equação da dinâmica de um novo termo. Este novo termo poderá dar origem a descontinuidades na trajectória visto que, através dele, o espaço dos controlos é alargado de modo a incluir medidas vectoriais.

Sem qualquer perda de generalidade a nova formulação do problema de controlo óptimo será pois

(P) Minimizar $h(x(1))$

através da escolha de (x, u, μ) tal que

$$dx(t) = f(t, x(t), u(t)) dt + G(t, u(t)) \mu(dt) \quad L, \mu - \text{q.s. em } [0, 1] \tag{3.1}$$

$$(x(0), x(1)) \in C_0 \times C_1 \tag{3.2}$$

$$A(t)x(t) + b(t) \leq 0 \quad \text{para todo o } t \in [0, 1] \tag{3.3}$$

$$u(t) \in \Omega_t \quad L \text{ e } \mu \text{ q.s.} \tag{3.4}$$

$$\mu(B) \in K \quad \text{para todo o subconjunto Borel } B \text{ de } [0, 1] \tag{3.5}$$

onde $h : \mathbb{R}^n \rightarrow \mathbb{R}$, $f : [0, 1] \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$, $G : [0, 1] \times \mathbb{R}^m \rightarrow \mathbb{R}^{n \times k}$,
 $A : [0, 1] \rightarrow \mathbb{R}^n$ e $b : [0, 1] \rightarrow \mathbb{R}$ são funções dadas,
 C_1 e C_0 são subconjuntos fechados de \mathbb{R}^n ,

K é um cone convexo e fechado em \mathbb{R}^k com vértice na origem e

$$\Omega_t := \{u \in \mathbb{R}^m : (t, u) \in \Omega \subset [0, 1] \times \mathbb{R}^m\}.$$

"L e μ q.s." significa que a relação que qualifica se verifica "quase sempre", ou seja, a menos de subconjuntos do intervalo $[0, 1]$ de medidas de Lebesgue e μ nulas. Fazemos notar que (3.1) não é mais do que uma forma abreviada da equação

$$x(t) = x(0) + \int_0^t f(s, x(s), u(s)) ds \int_{[0,t]} G(s, u(s)) \mu(ds).$$

Uma estratégia de controlo consiste num par (u, μ) que, além de satisfazer (3.4) e (3.5), é tal que $t \rightarrow G(t, u(t))$ é integrável em relação à medida μ .

Um processo generalizado (x, u, μ) diz-se admissível se (u, μ) for uma estratégia de controlo e a respectiva trajectória, x , definida como solução de (3.1), satisfaz (3.2) e (3.3).

O alargamento da classe de restrições de estado (3.3) usando os nossos métodos constitui ainda um problema de investigação aliciante.

Em Pereira [8] e Vinter e Pereira [13] constata-se que esta formulação é adequada pois, além de ser possível garantir a existência de solução mediante hipóteses razoáveis, permite o desenvolvimento de condições necessárias de optimalidade e a construção de processos convencionais (envolvendo apenas controlos no sentido usual) arbitrariamente próximos de um dado processo generalizado.

As propriedades dos elementos do nosso problema necessárias para a obtenção das condições de optimalidade são:

- h é localmente Lipschitz contínua;
- Dado $\delta > 0$, $f(t, \cdot, u)$ tem constante de Lipschitz $K_\delta(t)$ em

$$\{Z \in \mathbb{R}^n : \|Z\| < \delta\}$$

para todo $u \in \Omega_t$ e $t \in [0, 1]$;

- Ω é um conjunto $L \times B^m$ - mensurável;

– A função $(t, u) \rightarrow \text{col}(f^T(t, x, u), G^T(t, u))$ é $L \times B^m$ mensurável para todo o $x \in \mathbb{R}^n$;

($L \times B^m$ é a σ -álgebra dos subconjuntos $L \times B$ onde $L \subset [0, 1]$ é de Lebesgue e $B \subset \mathbb{R}^m$ é de Borel.)

- Sendo x^* a trajectória óptima de (P), $\|f(t, x^*(t), u)\|$ é limitada L q.s. por uma função integrável para todo o $u \in \Omega_i$;
- As funções A e b em (3.3) são contínuas; e
- A compactificação do termo singular da dinâmica

$$t \rightarrow \Gamma(t) := \overline{\text{co}} \left\{ \frac{G(t, u) - \omega}{1 + \sum_i |\sum_j g_{ij}(t, u) \omega_j|} : u \in \Omega_i, \omega \in K \cap B_\infty \right\} \subset \mathbb{R}^n$$

é contínua em relação à métrica de Hausdorff (B_∞ denota o cubo unitário de \mathbb{R}^k e, $\overline{\text{co}}A$ o fecho do fecho convexo do conjunto A).

É oportuno observar que as hipóteses acima referidas respeitantes ao termo convencional da dinâmica são apenas ligeiramente mais fortes do que aquelas impostas por Clarke [2] que são de facto as mais fracas utilizadas até hoje para demonstrar um Princípio do Máximo para sistemas com trajectórias absolutamente contínuas.

No caso de m ser uma medida escalar e G uniformemente limitada, então a última hipótese reduz-se à continuidade de Hausdorff de $t \rightarrow \overline{\text{co}} G(t, \Omega_i)$.

4. Condições Necessárias de Optimalidade

As condições necessárias de optimalidade tomam a forma de um princípio do máximo de Pontryagin e são descritas pelo seguinte

Teorema

Seja (x^*, u^*, μ^*) um processo óptimo de (P).

Então existe um terno de elementos, (p, λ, ν) , onde λ é um número não negativo, $p: [0, 1] \rightarrow \mathbb{R}^n$ é absolutamente contínua e ν é uma medida não negativa cujo suporte está contido no conjunto

$$\{t \in [0, 1] : \limsup_{s \rightarrow t} [A(s)x^*(s) + b(s)] = 0\},$$

satisfazendo

$$\lambda + \|v\| + \|p\| \neq 0$$

$$-\dot{p}(t) \in q(t) \cdot \partial_x f(t, x^*(t), u^*(t)) \quad L \text{ q.s.}$$

$$p(0) \in N_{C_0}(x^*(0)), \quad -p(1) + \int_{[0,1]} A(t)v(dt) \in N_{C_1}(x^*(1)) + \lambda \partial_x h(x^*(1))$$

$$u^*(t) \text{ maximiza } u \rightarrow q(t) \cdot f(t, x^*(t), u) \text{ em } \Omega_t \quad L \text{ q.s.}$$

$$\text{Sup} \{ \sigma_K(q(t) \cdot G(t, u)) : u \in \Omega_t \} \leq 0 \quad \text{para todo o } t \in [0, 1]$$

$$e \quad \limsup_{s \rightarrow t} \left\{ \sigma_K(q(s) \cdot G(s, u^*(s))) \right\} = 0 \quad \mu^* \text{ q.s.}$$

Aqui, $q(t) := p(t) + \int_{[0,t]} A(s)v(ds)$ e $\sigma_K(\cdot)$ é a função suporte do cone K , ou seja,

$$\sigma_K(p) = \text{Sup}\{p \cdot k : k \in K\}.$$

$N_A(a)$ e $\partial_x f$ denotam respectivamente o cone normal do conjunto A no ponto a e o gradiente/Jacobiano generalizado da função f na variável x no sentido de Clarke [3].

Se a medida do controlo m for escalar e $K = [0, \infty)$ então as duas últimas relações serão escritas na forma

$$\text{Sup} \{ q(t) \cdot G(t, u) : u \in \Omega_t \} \leq 0 \quad \text{para todo o } t \in [0, 1]$$

$$e \quad \limsup_{s \rightarrow t} \{ q(s) \cdot G(s, u^*(s)) \} = 0 \quad \mu^* \text{ q.s.}$$

Essencialmente, estas condições dizem-nos que a componente convencional do controlo satisfaz o princípio do máximo usual, ou seja, na ausência do termo singular da dinâmica.

Embora as duas últimas condições apenas especificuem directamente o suporte da medida μ^* , esta poderá, mediante condições favoráveis, ser completamente determinada com a ajuda das restantes restrições do problema em conjunção com as relações fornecidas pelas condições necessárias de optimalidade.

A demonstração detalhada deste resultado aparece em Vinter e Pereira [13] e Pereira [9]. Em Pereira [9] mostra-se como as condições necessárias de optimalidade são obtidas a partir de um outro resultado (Teorema 4.1 de Pereira [9]) caracterizando processos cuja trajectória termina na fronteira do conjunto atingível no instante final.

As ideias básicas por detrás da demonstração deste resultado consistem na:

- (i) Utilização de resultados respeitando a aproximação de medidas (derivados em Vinter e Pereira [13] e Pereira [9]) para construir uma sucessão de problemas auxiliares constituindo uma aproximação de (P) e caracterizados por envolverem apenas controlos convencionais, não possuírem restrições de estado e as funções custo reflectirem a violação de (2.3).
- (ii) Aplicação do Teorema de Ekeland (ver Ekeland [4]) para encontrar uma sucessão de processos contendo uma subsucessão convergindo para a solução de (P) e tal que cada um dos seus termos não só quase satisfaz (2.3) como minimiza uma perturbação adequada de cada um dos problemas referidos em (i).
- (iii) Aplicação a cada um dos processos de (ii) do princípio do máximo convencional para problemas não diferenciáveis (ver Clarke [3]) e obtenção no limite das conclusões do Teorema.

5. Comparação com Trabalho Relacionado

Rishel [10] e Warga [14] e [15] também fornecem condições necessárias de optimalidade para problemas de controlo impulsional. Embora tratem problemas diferentes e com diferentes objectivos, ambos têm em comum a aplicação do princípio do máximo usual a um problema convencional obtido do original através de uma mudança da variável independente dependente da estratégia escolhida. Esta é tal que a variável tempo do problema original passa a ser uma das componentes do estado do sistema reparametrizado. Deste modo, ter-se-á que impor a invariância das restrições do controlo e a continuidade de Lipschitz da equação dinâmica em relação ao tempo o que tornam estas condições bastante mais restritivas do que as nossas.

Enquanto que Rishel [10] necessita de impor a "constancy condition" (uma espécie de convexidade direccionada) no termo singular da dinâmica para exprimir as suas condições na parametrização original, Warga [14] e [15] considera apenas a relaxação do problema o que lhe permite considerar uma estrutura mais genérica do que a nossa e a de Rishel.

Questões da adequação da formulação do problema, condições de optimalidade, hipóteses de existência e interpretações de dualidade são endereçadas por

Rockafellar [11] e [12] para o problema de controlo impulsional no contexto do problema generalizado de Bolza. Para isso, embora assumamos apenas a mensurabilidade do integrando em relação ao tempo, necessitamos da sua convexidade conjunta em relação ao estado e à velocidade, hipótese esta que não utilizamos.

6. Referências

- [1] CLARK, C. W., CLARKE, F. H., MUNRO, G. R., (1979), *The optimal exploitation of renewable resource stocks*, *Econometrica*, 47, 25-47.
- [2] CLARKE, F. H., (1976), *The maximum principle under minimal hypotheses*, *SIAM J. Control and Optim.*, 14, 1078-1091.
- [3] CLARKE, F. H., (1983), *Optimization and Nonsmooth Analysis*, John Wiley & Sons Inc., New York.
- [4] EKELAND, I., (1979), *Nonconvex Minimization Problems*, *Bull. Am. Math. Soc.*, 1 (1979), 443-474.
- [5] LAWRENCE, D. F., (1963), *Optimal Trajectories for Space Navigation*, Butterworth, London.
- [6] MAREC, J. P., (1986), *Optimal Space Trajectories*, Elsevier, Amsterdam-Oxford.
- [7] NEUSTADT, L. W., (1965), *A general theory of minimum-fuel space trajectories*, *SIAM J. Control and Optim.*, 3, 317-356.
- [8] PEREIRA, F., (1986), *A Maximum Principle for Impulsive Control Systems*, PhD thesis, Imperial College, University of London, London.
- [9] PEREIRA F., VINTER, R. B., *A Maximum Principle for Impulsive Control Problems with State Constraints*, em preparação.
- [10] RISHEL, R. W., (1976), *An Extended Pontryagin Principle for Control Systems whose Control Laws Contain Measures*, *SIAM J. Control*, 3, 191-205.
- [11] ROCKAFELLAR, R. T., (1976), *Dual Problems for Arcs of Bounded Variation*, in: *Calculus of Variations and Control Theory*, D.L. Russell ed., Academic Press, New York, 155-192.

- [12] ROCKAFELLAR, R. T., (1981), *Optimality Conditions for Convex Control Problems with Nonnegative States and the Possibility of Jumps*, in: *Game Theory and Mathematical Economics*, O. Moeschlin, D. Pallaschke eds., North Holland, Amsterdam, 339-349.
- [13] VINTER, R. B., PEREIRA, F., (1988), *A Maximum Principle for Optimal Processes with Discontinuous Trajectories*, *SIAM J. Control and Optim.*, 26, 205 -229.
- [14] WARGA, J., (1972), *Optimal Control of Differential and Functional Equations*, Academic Press, New York.
- [15] WARGA J., (1966), *Variational Problems with Unbounded Controls*, *SIAM J. of Control and Optim.*, 3, 424-438.

DUALIDADE EM PROGRAMAÇÃO LINEAR FRACCIONÁRIA

Domingos Moreira Cardoso
Departamento de Matemática
Universidade de Aveiro
3800 AVEIRO
INESC - Coimbra
Rua Antero de Quental, 231-cave
3000 COIMBRA

Resumo

Faz-se uma breve revisão dos principais desenvolvimentos obtidos sobre dualidade em programação linear fraccionária e analisam-se algumas das dificuldades levantadas pela abordagem clássica.

Tomando por base um método dirigido à análise paramétrica de programas lineares fraccionários, apresenta-se um dual para este tipo de problemas, cuja resolução não apresenta grande dificuldade e com o qual, não só se verificam as relações primal-dual mais comuns, como também, com facilidade, se pode proceder à análise de sensibilidade.

Finalmente, fazem-se algumas interpretações geométricas num espaço determinado pelas imagens da região admissível obtidas a partir do primal e do dual respectivamente.

Abstract

Firstly we present an overview of the main developments on duality in linear fractional programming and we analyse some of the difficulties raised by the classical approach.

Taking as basis a method directed to the parametric analysis of linear fractional programs, we put forward a dual formulation for this type of problems, the resolution of which does not present great difficulty. The proposed formulation verifies the more common dual-primal type of relationships and allows for an easy sensitivity analysis.

Finally, geometric interpretations in a space determined by the images of the admissible region (obtained from the primal and dual problems respectively) are presented.

Keywords

Dual Problem, Duality Theory, Optimization, Fractional Programming.

1. Introdução

Designa-se por programa fraccionário o seguinte programa matemático:

$$(P) \quad \min \left\{ f(x) = \frac{f_1(x)}{f_2(x)} : x \in X \right\}$$

relativamente ao qual vamos considerar que X é uma parte não vazia compacta de \mathbb{R}^n , que f_1 e f_2 são funções de \mathbb{R}^n em \mathbb{R} contínuas num aberto A que contém X , e ainda que $f_2(x) > 0 \forall x \in A$.

Obviamente que este problema tem solução uma vez que se trata de encontrar, num compacto, o mínimo de uma função que, sendo o quociente de duas funções contínuas e sem pontos de singularidade, é contínua.

Associados a estes problemas e com o intuito de os resolver, Dinkelbach [7] propõe a seguinte função paramétrica:

$$\begin{aligned} \eta : \mathbb{R} &\longrightarrow \mathbb{R} \\ \alpha &\longmapsto \eta(\alpha) = \min \{ f_1(x) - \alpha f_2(x) : x \in X \} \end{aligned}$$

que tem a propriedade de ser côncava, monótona estritamente decrescente e contínua.

Das propriedades enunciadas, apenas vamos provar a que se referê à continuidade de η , por ser a única que não é demonstrada em Dinkelbach [7].

Com efeito, seja $\alpha \in \mathbb{R}$, $\varepsilon \in \mathbb{R}^+$ e $I \equiv [\alpha - \varepsilon, \alpha + \varepsilon]$. Então $\forall \delta \in \mathbb{R}$ tal que $\alpha + \delta \in I$ e $\forall \lambda \in [0, 1]$, tem-se que $\alpha + \lambda\delta \in I$ e consequentemente pela concavidade de η vem:

$$\begin{aligned} \eta(\lambda(\alpha + \delta) + (1 - \lambda)\alpha) &\geq \lambda\eta(\alpha + \delta) + (1 - \lambda)\eta(\alpha) \Leftrightarrow \\ \eta(\alpha) - \eta(\alpha + \lambda\delta) &\leq \lambda(\eta(\alpha) - \eta(\alpha + \delta)) \end{aligned}$$

Uma vez que η é monótona estritamente decrescente, fazendo $M = \eta(\alpha - \varepsilon) - \eta(\alpha + \varepsilon)$ conclui-se que $\eta(\alpha) - \eta(\alpha + \delta) \leq M \forall \delta \in \mathbb{R}$ com $\alpha + \delta \in I$, pelo que $\eta(\alpha) - \eta(\alpha + \lambda\delta) \leq \lambda M$. De modo análogo, substituindo $\alpha + \delta$ por $\alpha - \delta$, se obtém a inequação $\eta(\alpha + \lambda\delta) - \eta(\alpha) \leq \lambda M$. Combinando estas duas inequações, vem finalmente, $|\eta(\alpha + \lambda\delta) - \eta(\alpha)| \leq \lambda M$, o que completa a demonstração.

Como consequência imediata de η ser monótona estritamente decrescente

e contínua, e do facto de $\lim_{\alpha \rightarrow +\infty} \eta(\alpha) = -\infty$ e $\lim_{\alpha \rightarrow -\infty} \eta(\alpha) = +\infty$, tem-se que a equação $\eta(\alpha) = 0$, tem uma única raiz.

Com base na função paramétrica $\eta(\alpha)$, e tendo presente o teorema a seguir, Dinkelbach [7] propõe um algoritmo muito simples que na prática corresponde à aplicação do método de Newton à determinação da raiz da equação $\eta(\alpha) = 0$. Deve observar-se porém que, antes da apresentação deste método, já Isbell e Marlow [10] haviam proposto uma abordagem perfeitamente semelhante a esta, embora com vista unicamente à resolução de programas lineares fraccionários.

Teorema 1

$x^0 \in X$ é solução óptima para (P) sse sendo α^0 raiz da equação $\eta(\alpha) = 0$, $\eta(\alpha^0) = f_1(x^0) - \alpha^0 f_2(x^0)$ e $\alpha^0 = f(x^0)$.

Prova:

$x^0 \in X$ é solução óptima para (P)

$$\begin{aligned} &\Leftrightarrow f(x^0) \leq f(x) \quad \forall x \in X \\ &\Leftrightarrow f_1(x) - f(x^0) f_2(x) \geq 0 \quad \forall x \in X \\ &\Leftrightarrow \eta(f(x^0)) = \min \{ f_1(x) - f(x^0) f_2(x) : x \in X \} \geq 0 \\ &\Leftrightarrow \eta(f(x^0)) = 0, \text{ uma vez que } f_1(x^0) - f(x^0) f_2(x^0) = 0. \end{aligned}$$

O facto da equação $\eta(\alpha) = 0$ ter uma única raiz, obriga a que se tenha $\alpha^0 = f(x^0)$.

◆

Ao problema (P), tem-se assim associada uma curva que corresponde ao gráfico de $\eta(\alpha)$, para a qual, sendo x_α^* tal que $\eta(\alpha) = f_1(x_\alpha^*) - \alpha f_2(x_\alpha^*)$, se verifica que a recta tangente em $(\alpha, \eta(\alpha))$ intersecta o eixo dos $\alpha\alpha$ em $\alpha = f(x_\alpha^*)$ e o eixo dos $\eta\eta$ em $\eta = f_1(x_\alpha^*)$, sendo o seu declive

$$\frac{d\eta(\alpha)}{d\alpha} = -f_2(x_\alpha^*),$$

conforme se exemplifica na figura 1.

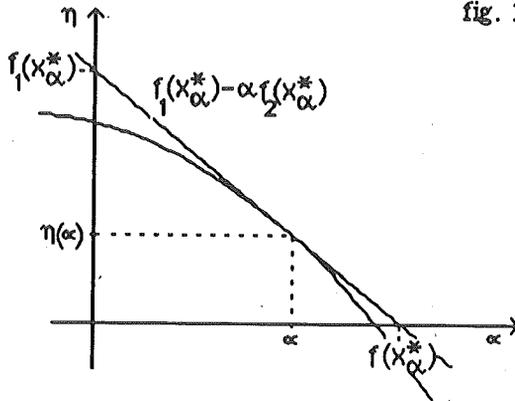


fig. 1

Por outro lado, sendo x^1 e x^2 , respectivamente, soluções dos problemas $\min\{f_2(x) : x \in X\}$ e $\max\{f_2(x) : x \in X\}$, $\eta(\alpha)$ converge assintoticamente à esquerda para a recta com declive $-f_2(x^1)$, que intersecta o eixo dos $\alpha\alpha$ em $\alpha = f(x^1)$ e o eixo dos $\eta\eta$ em $\eta = f_1(x^1)$, e à direita para a recta com declive $-f_2(x^2)$, que intersecta o eixo dos $\alpha\alpha$ em $\alpha = f(x^2)$ e o eixo dos $\eta\eta$ em $\eta = f_1(x^2)$, conforme se ilustra na figura a seguir.

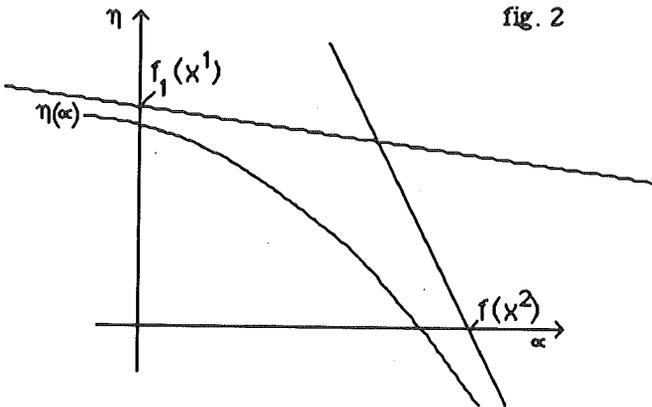


fig. 2

Quando em (P), f_1 e f_2 são aplicações afins e X é definido por um conjunto de restrições lineares, este problema designa-se por programa linear fraccionário.

No que se seguirá, f_1 e f_2 são definidas por $f_i(x) = c_i^T x + b_i$, com $c_i, x \in \mathbb{R}^n$ para $i = 1, 2$, e X por $X = \{x \in \mathbb{R}^n : Ax \leq b, x \geq 0\}$, com $A \in \mathbb{R}^{m \times n}$ e

$b \in \mathbb{R}^m$, e $F(X)$ corresponderá à parte compacta convexa de \mathbb{R}^2 , $F(X) = \{(f_2(x), f_1(x)) : x \in X\}$ (deve observar-se que a imagem por uma função contínua de um compacto é compacta e que a imagem por uma aplicação afim de um convexo é um convexo). Nestas condições, a curva $\eta(\alpha)$ associada a (P) tomará o aspecto da figura 3, ou seja, trata-se de uma curva linear por partes onde, a cada intervalo de valores de α no qual $\eta(\alpha)$ é afim, corresponde um vértice em $F(X)$ e, a cada ponto onde a derivada de $\eta(\alpha)$ é descontínua, corresponde uma aresta de $F(X)$.

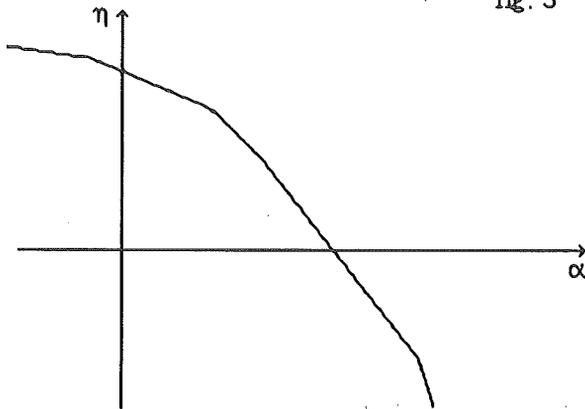


fig. 3

Dos métodos desenvolvidos para a resolução deste tipo de problemas (programas lineares fraccionários) destacamos ainda:

- os apresentados por Ibaraki [9], que se baseiam na aplicação do método das bissecções à determinação da raiz da equação $\eta(\alpha) = 0$;
- o proposto por Clímaco e Cardoso [5] que, embora fazendo parte da família de métodos que têm por base a função paramétrica $\eta(\alpha)$, apresenta um critério de escolha do parâmetro α , que se fundamenta na transformação do problema inicial num problema bicritério;
- o método de Martos [14], que embora lide com funções objectivo do tipo $f_1(x) - \alpha f_2(x)$, as vai modificando de vértice para vértice, de acordo com as sucessivas actualizações do parâmetro α , independentemente de ter ou não obtido o correspondente valor óptimo, $\eta(\alpha)$;

- o método de Charnes e Cooper [4], que consiste na aplicação de uma mudança adequada de variáveis, de modo a transformar o programa linear fraccionário num programa linear, cuja resolução determina a solução óptima do problema inicial.

2. Análise dos Principais Desenvolvimentos obtidos sobre a Dualidade em Programação Linear Fraccionária.

Vamos começar por recordar os conceitos de pseudo-convexidade (respectivamente pseudo-concavidade) e quase-convexidade (respectivamente quase-concavidade).

Dada uma parte C de \mathbb{R}^p , uma função $\psi : \mathbb{R}^p \rightarrow \mathbb{R}$ diz-se pseudo-convexa (respectivamente pseudo-côncava) em C se:

$$\begin{aligned} \forall x^1, x^2 \in C \quad (x^2 - x^1) \nabla_x \psi(x^1) \geq 0 &\Rightarrow \psi(x^2) \geq \psi(x^1) \\ ((x^2 - x^1) \nabla_x \psi(x^1) \leq 0 &\Rightarrow \psi(x^2) \leq \psi(x^1)) \end{aligned}$$

e diz-se quase-convexa (respectivamente quase-côncava) em C se C é convexo e se verifica que:

$$\begin{aligned} \forall x^1, x^2 \in C \quad \psi(x^2) \leq \psi(x^1) &\Rightarrow (x^2 - x^1) \nabla_x \psi(x^1) \leq 0 \\ (\psi(x^2) \geq \psi(x^1) &\Rightarrow (x^2 - x^1) \nabla_x \psi(x^1) \geq 0) \end{aligned}$$

Como é óbvio, de acordo com esta definição, uma função ψ diz-se pseudo-convexa, sse $-\psi$ é pseudo-côncava e quase-convexa sse $-\psi$ é quase-côncava.

De acordo com Mangasarian [13], considerando-se como problema primal:

$$(PW) \quad \min_{x \in \mathbb{R}^n} \{ \theta(x) : g(x) \geq 0 \}$$

em que a função objectivo ($\theta : \mathbb{R}^n \rightarrow \mathbb{R}$) e as componentes da função que define a região de admissibilidade ($g : \mathbb{R}^n \rightarrow \mathbb{R}^m$) são diferenciáveis num aberto A de \mathbb{R}^n , o bem conhecido resultado da suficiência das condições de Kuhn-Tucker para a solução óptima de (PW), quando θ e as componentes de

-g são convexas em A, pode extender-se aos casos em que θ é pseudo-convexa e as componentes de g são quase-côncavas, conforme se enuncia no teorema a seguir (a prova deste teorema pode encontrar-se em [12] ou [13] pag. 151-153).

Teorema 2

Seja $\theta : \mathbb{R}^n \rightarrow \mathbb{R}$ e as componentes de $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$, funções diferenciáveis num aberto A de \mathbb{R}^n , C uma parte convexa de A e considere-se ainda que θ é pseudo-convexa em C e que as componentes de g são quase-côncavas em C.

Se $\exists x^0 \in C$ e $y^0 \in \mathbb{R}^m$ satisfazendo as condições de Kuhn-Tucker, ou seja:

$$\begin{aligned} \text{(a)} \quad & \nabla_x \theta(x^0) - \nabla_x \sum_{i=1}^m y_i^0 g_i(x^0) = 0, \\ \text{(b)} \quad & y^0 \geq 0, \\ \text{(c)} \quad & g(x^0) \geq 0, \\ \text{(d)} \quad & \sum_{i=1}^m y_i^0 g_i(x^0) = 0, \end{aligned}$$

então, $\theta(x^0) = \{\theta(x) : g(x) \geq 0\}$.

Aplicando a (PW) o dual proposto por Wolfe [19] para a programação convexa, ou seja, tomando-se para dual de (PW) o programa matemático:

$$\text{(DW)} \quad \max_{x \in \mathbb{R}^n, y \in \mathbb{R}^m} \{ \varphi(x, y) : \nabla_x \varphi(x, y) = 0, y \geq 0 \}$$

com $\varphi(x, y) = \theta(x) - y^t g(x)$, e supondo que as condições do teorema 2 são satisfeitas, dado que (a) e (b) correspondem às condições de admissibilidade do dual e (c) corresponde às de admissibilidade do primal, se (x^0, y^0) é admissível para este dual e x^0 é admissível para o primal, para que x^0 seja uma solução óptima de (PW), basta que a condição (d) se verifique (trata-se obviamente da conhecida condição de complementaridade de "slacks").

Este resultado é aplicável a problemas de programação linear fraccionária, dado que nestes programas matemáticos as restrições são lineares, logo são quase-côncavas (e quase-convexas) e a função objectivo é pseudo-convexa (e

pseudo-côncava).

Com efeito, se uma dada função ψ é linear num dado convexo C , então é côncava (e convexa) em C , e nestas condições:

$$\begin{aligned} \forall x^1, x^2 \in C \quad \psi(x^2) - \psi(x^1) &\leq (x^2 - x^1) \nabla_x \psi(x^1) \\ (\psi(x^2) - \psi(x^1)) &\geq (x^2 - x^1) \nabla_x \psi(x^1) \end{aligned}$$

donde se tira que $(x^2 - x^1) \nabla_x \psi(x^1) \leq 0$ ($(x^2 - x^1) \nabla_x \psi(x^1) \geq 0$) \Rightarrow
 $\psi(x^2) - \psi(x^1) \leq 0$ ($\psi(x^2) - \psi(x^1) \geq 0$).

Por outro lado, $\nabla f(x) = \frac{1}{f_2(x)} (c_1^t - f(x)c_2^t)$ e

$$(x^2 - x^1) \nabla f(x^1) = \frac{f_2(x^2)}{f_2(x^1)} (f(x^2) - f(x^1)),$$

donde se conclui que

$$(x^2 - x^1) \nabla f(x^1) \geq 0 \Rightarrow f(x^2) - f(x^1) \geq 0$$

$$((x^2 - x^1) \nabla f(x^1) \leq 0 \Rightarrow f(x^2) - f(x^1) \leq 0).$$

Assim, considerando (PW) como sendo o programa linear fraccionário, (P), e tomando (DW) para seu dual, embora, de acordo com Mangasarian [12], se (DW) tem solução óptima (x^*, y^*) e a matriz Hessiana da respectiva f.o., em relação a x , é não singular em x^* , x^* seja solução óptima para (PW), não se verifica nem a validade do teorema da dualidade directa, nem a validade do teorema da dualidade fraca de Wolfe. Segue-se um exemplo em que se evidenciam estas dificuldades.

Seja o programa linear fraccionário $\min \left\{ \frac{x-3}{x} \mid x > 0, 1 \leq x \leq 2 \right\}$.

O dual de Wolfe correspondente será o programa matemático:

$$\max \left\{ \frac{x-3}{x} - y_1(x-1) - y_2(-x+2) \mid x > 0, y_1 - y_2 = \frac{3}{x^2}, y_1, y_2 \geq 0 \right\}$$

Fazendo $y_2 = 0$ e $y_1 = \frac{3}{x^2}$, vem:

$$\max \left\{ \frac{x-3}{x} - \frac{3x-3}{x^2} \mid x > 0 \right\} \Leftrightarrow \max \left\{ \frac{3(1-2x)}{x^2} + 1 \mid x > 0 \right\}$$

e, enquanto este dual não tem óptimo finito (dado que quando x tende para zero, a f.o. tende para infinito) o primal tem como solução óptima $x = 1$.

Conclui-se assim que este tipo de abordagem, para a obtenção do dual de um programa linear fraccionário, é pouco vantajosa, não só pelos inconvenientes da ausência das relações primal-dual mais comuns, como ainda pelo facto do dual assim obtido em geral vir mais "complicado" do que o primal, pelo que vamos considerar outras abordagens.

Considere-se novamente a curva associada a $\eta(\alpha)$, e tenha-se em conta que para cada α , a determinação de $\eta(\alpha)$ corresponde à resolução do P.L.:

$$(P_\alpha) \min \{ (c_1 - \alpha c_2)^t x + \beta_1 - \alpha \beta_2 : x \in \mathbb{R}^n, Ax \geq b, x \geq 0 \}$$

ao qual corresponde o dual de Wolfe:

$$(D_\alpha) \max \{ y^t b + \beta_1 - \alpha \beta_2 : y \in \mathbb{R}^m, y^t A \leq (c_1 - \alpha c_2)^t, y \geq 0 \}$$

Deste modo, para cada $\alpha \in \mathbb{R}$, $\eta(\alpha)$ determina também o valor óptimo de (D_α) . A figura 4 indica o aspecto das respectivas gamas de valores das f.o de (P_α) e (D_α) nas correspondentes regiões admissíveis.

Pelas características de $\eta(\alpha)$, com facilidade se conclui que, se acrescentarmos às restrições de (D_α) a respectiva f.o., obrigando-a a ser não negativa, e maximizarmos α , obtemos a raiz da equação $\eta(\alpha) = 0$, ou seja, o valor óptimo de (P) . Assim, tomando-se para dual de (P) , o P.M.:

$$(DCC) \max \{ \alpha : [y^t, \alpha] \in \mathbb{R}^{m+1}, [y^t, \alpha] \begin{bmatrix} A & -b \\ c_2^t & \beta_2 \end{bmatrix} \leq [c_1^t, \beta_1], [y^t, \alpha] \geq 0 \}$$

verifica-se a validade dos teoremas da dualidade fraca, dualidade directa e dualidade recíproca (deve observar-se que, embora em (DCC) apareça $\alpha \geq 0$, estando deste modo a supor-se que o valor óptimo de (P) é não negativo, no caso geral esta variável deverá aparecer como variável livre).

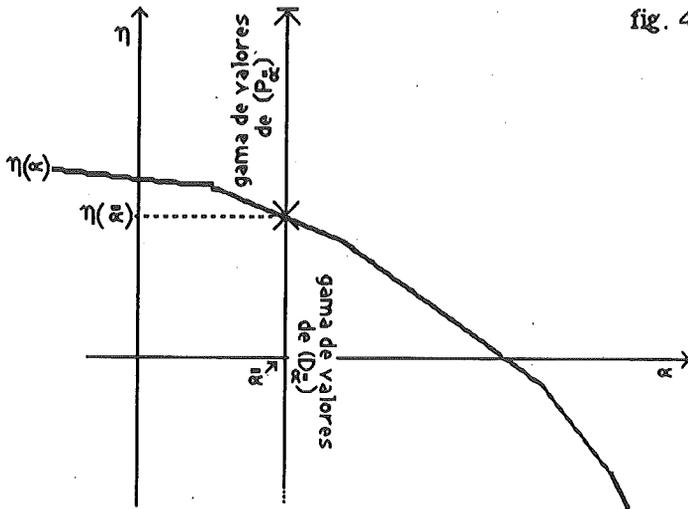


fig. 4

O dual (DCC) é semelhante ao apresentado por Chada [3], que embora não obrigue a região admissível a ser compacta, impõe a condição $Ax \leq 0 \Rightarrow x = 0$ e é uma particularização, para programas lineares fraccionários, do dual proposto por Jaganathan em [11] e por Schaible [15] (neste último não se obriga a região admissível a ser compacta).

Qualquer destes duais corresponde ao dual do P.L. que se obtém fazendo uso da transformação de variáveis de Charnes e Cooper [4], aplicada a (P). No entanto, os resultados que se apresentam em [3], [11] e [15] são mais fracos que o resultado enunciado no teorema a seguir.

Teorema 3

$\exists (y^*, \alpha^*)$ pertencente ao conjunto de soluções óptimas de (DCC) e $\exists x^*$ pertencente ao conjunto de soluções óptimas de (P), tal que $y^{*t}(Ax^* - b) = 0$ e α^* é o valor óptimo para ambos os problemas.

Prova:

Sabe-se que $\forall \alpha \in \mathbb{R}$, (P_α) tem solução óptima, x_α^* , e conseqüentemente pelo teorema da dualidade directa $\exists y_\alpha^*$ tal que y_α^* é solução óptima de (D_α) .

Para $\alpha = \alpha^*$ tal que $\eta(\alpha^*) = 0$, tem-se então que $\exists x^*$ solução óptima de (P_{α^*}) e $\exists y^*$ solução óptima de (D_{α^*}) e

$$y^{*t}b + \beta_1 - \alpha\beta_2 = \eta(\alpha^*) = (c_1 - \alpha c_2)^t x^* + \beta_1 - \alpha\beta_2.$$

Nestas condições vem que $y^{*t}b = (c_1 - \alpha c_2)^t x^*$ e como $y^*, x^* \geq 0$, de $y^{*t}A \leq (c_1 - \alpha c_2)$ e de $Ax^* \geq b$, tira-se que $y^{*t}b \leq y^{*t}Ax^* \leq (c_1 - \alpha c_2)^t x^* \Rightarrow 0 \leq y^{*t}(Ax^* - b) \leq (c_1 - \alpha c_2)^t x^* - y^{*t}b = 0$.

Por outro lado, sendo v_α^* o valor óptimo de (D_α) , $v_\alpha^* \leq \eta(\alpha)$, e como $\eta(\alpha) < 0$ para $\alpha > \alpha^*$, podemos concluir que $\forall (y, \alpha)$, solução admissível para (DCC), $\alpha \leq \alpha^*$ (visto que $\eta(\alpha) \geq y^t b + \beta_1 - \alpha\beta_2$ e, pelas condições de admissibilidade de (DCC), $y^t b + \beta_1 - \alpha\beta_2 \geq 0$).

◆

Apesar de "poderoso", dada a validade das relações primal-dual mais comuns, este dual apresenta o inconveniente de não permitir a análise de sensibilidade, a não ser a perturbações nos coeficientes da função numerador da função objectivo. Com efeito, a partir de (DCC) é muito difícil analisar a função de perturbação de (P), que determina a variação dos valores óptimos para diferentes vectores de termos independentes e a consequente análise da estabilidade de (P), sugerida por Geoffrion [8].

Existem ainda outras abordagens para a obtenção do dual de um programa linear fraccionário, como seja a de Craven e Mond [6], que partindo de uma certa relação de equivalência entre P.Ms., propõe dois duais possíveis para (P), os quais designa de duais fortes, por respeitarem os teoremas da dualidade fraca, dualidade directa e dualidade recíproca, sendo um idêntico a (DCC), e o outro, tal como os propostos nas abordagens de Swarup [17] e de Sharma e Swarup [16], um programa linear fraccionário. No entanto, todos apresentam os inconvenientes já referidos para (DCC), em alguns casos agravados ainda por outras dificuldades. Com efeito, no que se refere ao apresentado em [17], verifica-se que não satisfaz o teorema da dualidade recíproca, de acordo com um exemplo dado por Abrham e Luthra [1].

3. Nova Abordagem à Dualidade em Programação Linear Fraccionária

O dual que se pretende propor tem por base um método paramétrico de resolução e análise de programas lineares fraccionários, apresentado por Wolf [18], relativamente ao qual procuraremos, de um modo breve, apresentar os principais resultados (em [2] procede-se ao estudo deste método, seguindo-se uma abordagem mais geométrica em contraste com a abordagem mais analítica seguida por Wolf [18]).

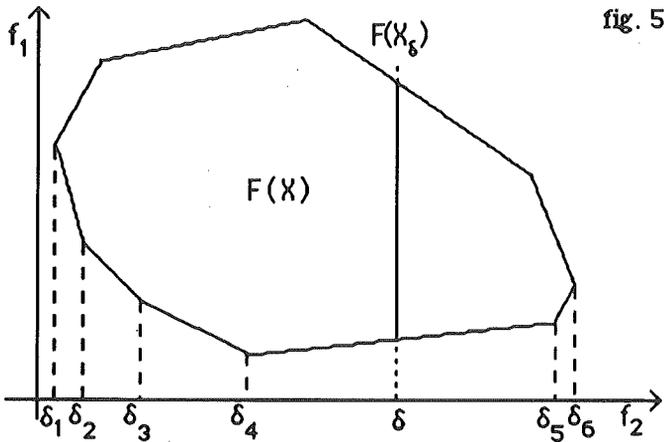


fig. 5

Tendo presente (P) identificaremos por Δ o conjunto

$$\{\delta \in \mathbb{R} : \exists x \in X, f_2(x) = \delta\},$$

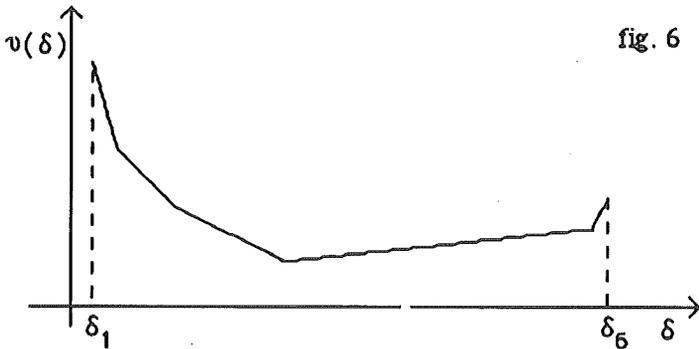
por X_δ o conjunto $\{x \in X : f_2(x) = \delta\}$ (note-se que $X \equiv \bigcup_{\delta \in \Delta} X_\delta$) e por $v(\delta)$ a seguinte aplicação:

$$v: \Delta \longrightarrow \mathbb{R}$$

$$\delta \longmapsto v(\delta) = \min \{f_1(x) : x \in X_\delta\}$$

De acordo com Wolf [18] verifica-se que Δ é um intervalo fechado e $v(\delta)$ é uma função convexa, contínua e afim por partes, i. e., $\exists p \in \mathbb{N}$ tal que $v(\delta) = \alpha_i \delta + \tau_i$ se $\delta \in [\delta_i, \delta_{i+1}]$, com $\alpha_i, \tau_i \in \mathbb{R}$ para $i = 1, 2, \dots, p$ e $\Delta \equiv \bigcup_{i=1}^p [\delta_i, \delta_{i+1}]$.

$\delta_{i+1}]$. A partir da figura 5, onde se esquematiza $F(X)$, com facilidade se conclui que o gráfico de $v(\delta)$ corresponde à parte "inferior" de $F(X)$, conforme a figura 6 exemplifica.



Identificando-se por Δ^* o subintervalo de Δ , no qual se verifica que:

$$\forall \delta^* \in \Delta^* \quad \frac{v(\delta^*)}{\delta^*} = \min \left\{ \frac{v(\delta)}{\delta} : \delta \in \Delta \right\},$$

ainda de acordo com Wolf [18], podemos enunciar os seguintes resultados:

Teorema 4

$x^* \in X$ é solução óptima para (P) sse $\exists \delta^* \in \Delta^*$ tal que $\frac{v(\delta^*)}{\delta^*} = f(x^*)$.

Prova:

Seja $x^* \in X$ uma solução óptima para (P); então, fazendo $\delta^* = f_2(x^*)$, vem que:

$$\frac{v(\delta^*)}{\delta^*} = \min_{\delta \in X_{\delta^*}} \frac{f_1(x)}{f_2(x^*)} \leq f(x^*) \leq \min_{\delta \in X_{\delta}} \frac{f_1(x)}{\delta} = \frac{v(\delta)}{\delta} \quad \forall \delta \in \Delta$$

donde se conclui que $\delta^* \in \Delta^*$ e $\frac{v(\delta^*)}{\delta^*} = f(x^*)$.

Seja $x^* \in X$ tal que $\frac{v(\delta^*)}{\delta^*} = f(x^*)$, com $\delta^* = f_2(x^*) \in \Delta^*$, e suponha-se

que $\exists x' \in X$ tal que x' é solução óptima para (P); então, de acordo com a primeira parte da demonstração, $\exists \delta' \in \Delta^*$ tal que $f(x') = \frac{v(\delta')}{\delta'}$, concluindo-se que $f(x^*) = \frac{v(\delta^*)}{\delta^*} \leq \frac{v(\delta')}{\delta'} = f(x') \leq f(x) \forall x \in X$, donde x^* é solução óptima para (P).

♦

Teorema 5

- (a) Se $\exists k \in \{1, \dots, p\}$ tal que para $\delta \in [\delta_k, \delta_{k+1}]$ $v(\delta) = \alpha_k \delta$ então $\Delta^* \equiv [\delta_k, \delta_{k+1}]$;
- (b) Se $\tau_1 < 0$, $\Delta^* \equiv \{\delta_1\}$;
- (c) Se $\tau_{p+1} > 0$, $\Delta^* \equiv \{\delta_{p+1}\}$;
- (d) Se $\exists k \in \{2, \dots, p\}$ tal que $\tau_{k+1} > 0$ e $\tau_k < 0$, $\Delta^* \equiv \{\delta_k\}$.

Prova:

Dado que $v(\delta)$ é convexa, podemos concluir que $\tau_j > \tau_{j+1} \forall j \in \{1, \dots, p\}$, conforme o exemplo que se tem vindo a esquematizar documenta (ver figura 7).

Por outro lado, a primeira derivada de $\frac{v(\delta)}{\delta}$ em $]\delta_j, \delta_{j+1}[$, para $j = 1, \dots,$

p é dada por $(\frac{v(\delta)}{\delta})' = \frac{-\tau_j}{\delta^2}$, donde se conclui que em $[\delta_j, \delta_{j+1}]$:

- se $\tau_j < 0$, então $\frac{v(\delta)}{\delta}$ é monótona crescente;
- se $\tau_j = 0$, então $\frac{v(\delta)}{\delta}$ é constante;
- se $\tau_j > 0$, então $\frac{v(\delta)}{\delta}$ é monótona decrescente.

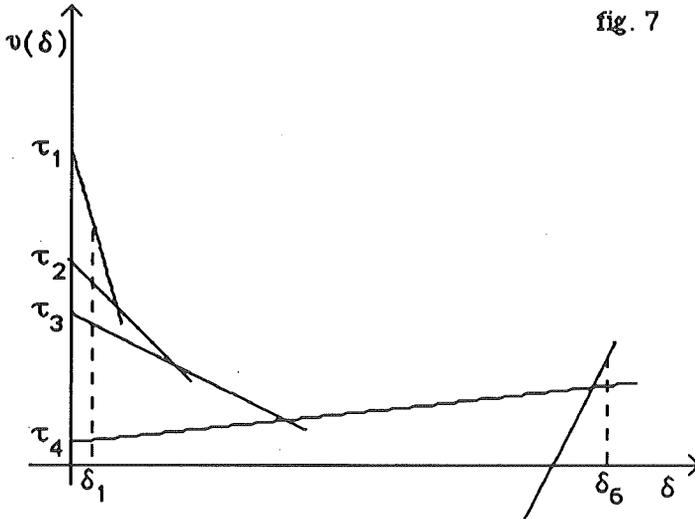


fig. 7

Em (a) vem que, se $\exists j \in \{1, 2, \dots, p\}$ tal que $j < k$, então, em $[\delta_j, \delta_{j+1}]$, $\frac{v(\delta)}{\delta}$ é monótona decrescente e, se $\exists i \in \{1, 2, \dots, p\}$ tal que $i > k$, então, em $[\delta_i, \delta_{i+1}]$, $\frac{v(\delta)}{\delta}$ é monótona crescente; nestas condições, $\Delta^* \equiv [\delta_k, \delta_{k+1}]$.

Em (b), uma vez que $\tau_1 < 0$, vem que $\tau_j < 0 \forall j \in \{1, 2, \dots, p\}$, logo $\frac{v(\delta)}{\delta}$ é monótona crescente em Δ e nestas condições $\Delta^* \equiv \{\delta_1\}$.

Em (c), $\tau_{p+1} > 0$, o que implica $\tau_j > 0 \forall j \in \{1, 2, \dots, p\}$, donde se conclui que $\frac{v(\delta)}{\delta}$ é monótona decrescente em Δ pelo que $\Delta^* \equiv \{\delta_{p+1}\}$.

Finalmente em (d), vem que $\tau_j > 0 \forall j \leq k-1$ e $\tau_i < 0 \forall i \geq k$, pelo que, $\frac{v(\delta)}{\delta}$ é monótona decrescente em $[\delta_1, \delta_k]$ e monótona crescente em $[\delta_k, \delta_{p+1}]$; nestas condições, pode concluir-se que, $\Delta^* \equiv \{\delta_k\}$.

◆

Com base nestes dois últimos teoremas, estamos agora em condições de introduzir o dual que se pretende propor.

Para cada $\delta \in \Delta$, considere-se o P.M., $(P^\delta) v(\delta) = \min \{f_1(x) : x \in X_\delta\}$, ou seja:

$$v(\delta) = \min f_1(x)$$

$$(P^\delta) \quad \text{s. a.} \quad \begin{bmatrix} A \\ c_2^t \\ -c_2^t \end{bmatrix} x \geq \begin{bmatrix} b \\ \delta - \beta_2 \\ -\delta + \beta_2 \end{bmatrix}$$

$$x \geq 0$$

ao qual corresponde o programa dual:

$$\psi(\delta) = \max y^t b + (\delta - \beta_2)z + \beta_2$$

$$(D^\delta) \quad \text{s. a.} \quad [y^t z] \begin{bmatrix} A \\ c_2^t \end{bmatrix} \leq c_1^t$$

$$y \geq 0,$$

$$z \in \mathbb{R}$$

Assim, com δ percorrendo Δ , para a família de pares de programas (P^δ) - (D^δ) , tem-se que $v(\delta) = \psi(\delta)$. Por outro lado, $\forall x$ admissível para (P^δ) e $\forall [y^t, z]$ admissível para (D^δ) , tem-se que $y^t b + (\delta - \beta_2)z \leq c_1^t x$ donde, $\frac{y^t b}{\delta} + (1 - \frac{\beta_2}{\delta})z + \frac{\beta_1}{\delta} \leq \frac{\psi(\delta)}{\delta} = \frac{v(\delta)}{\delta} \leq \frac{c_1^t x + \beta_1}{\delta} \quad \forall \delta \in \Delta$, logo em particular, $\min\{\frac{\psi(\delta)}{\delta} : \delta \in \Delta\} = \min\{\frac{v(\delta)}{\delta} : \delta \in \Delta\}$. Nestas condições, se Δ^* é o intervalo de valores óptimos para $\min\{\frac{v(\delta)}{\delta} : \delta \in \Delta\}$, podemos concluir que Δ^* é também o intervalo de valores óptimos para $\min\{\frac{\psi(\delta)}{\delta} : \delta \in \Delta\}$.

Tomando-se para dual de (P) o P.M.:

$$\min_{\delta \in \Delta} \max \quad \frac{y^t b}{\delta} + (1 - \frac{\beta_2}{\delta})z + \frac{\beta_1}{\delta}$$

$$(D) \quad \text{s.a.} \quad [y^t z] \begin{bmatrix} A \\ c_2^t \end{bmatrix} \leq c_1^t$$

$$y \geq 0,$$

$$z \in R$$

pode-se concluir que $\exists x^* \in X$, solução óptima para (P), e $\exists [y^*, z^*]$, solução óptima para (D), tais que os respectivos valores óptimos coincidem. Deve observar-se ainda que, tendo em conta as propriedades de $v(\delta)$ enunciadas no teorema 5, e o facto de se ter $v(\delta) = \psi(\delta) \forall \delta \in \Delta$, a determinação da solução óptima de (D) pode ser feita resolvendo-se um P.M. arbitrário, (D^δ) , e procedendo-se, a partir dele, a uma análise paramétrica adequada (incrementando ou decrementando δ , consoante se obtenha o termo independente de $\psi(\delta)$, para o subintervalo de Δ em estudo, respectivamente positivo ou negativo). Em [2] faz-se um estudo detalhado desta análise paramétrica.

A partir de (D), pode com facilidade proceder-se à análise de sensibilidade, nomeadamente dos efeitos, sobre a solução óptima encontrada, de perturbações provocadas no vector dos termos independentes do sistema de equações definidor da região admissível de (P), e nos termos independentes das funções numerador e denominador definidoras de $f(x)$, bem como nas componentes do gradiente da função numerador da f.o. do primal.

Voltando-se aos gráficos relativos, respectivamente a $v(\delta)$ (ou $\psi(\delta)$) e $\eta(\alpha)$, vem que para $x^* \in X$, $\exists \delta \in \Delta$ tal que $v(\delta) = f_1(x^*)$ sse $\exists \alpha \in R$ tal que $\eta(\alpha) = f_1(x^*) - \alpha f_2(x^*)$ e $f_2(x^*) = \delta$. Deste modo, conclui-se que:

$$\forall \delta \in \Delta \exists \alpha \in R$$

tal que

$$\eta(\alpha) = v(\delta) - \alpha \delta \quad \text{e} \quad \forall \alpha \in R \exists \delta \in \Delta$$

tal que

$$\eta(\alpha) = v(\delta) - \alpha\delta$$

Assim, se num certo subintervalo de Δ , $[\delta_i, \delta_{i+1}]$, $\forall \delta \in [\delta_i, \delta_{i+1}] \psi(\delta) = \alpha_i\delta + \tau_i$, poderá concluir-se que $\eta(\alpha_i) = \tau_i$, donde, pelas propriedades de $v(\delta) = \psi(\delta)$ e sendo v^* o valor óptimo correspondente ao par de problemas primal dual (P)–(D) vem que, se $\tau_i > 0$, então $\alpha_i < v^*$, e se $\tau_i < 0$, então $\alpha_i > v^*$. Deste modo, embora para o dual proposto, (D), não seja válido o teorema da dualidade fraca de Wolfe, esta contrariedade não invalida que, ao longo da resolução de (D), não possamos minorar, de modo adequado, os valores da f.o. de primal (P).

Identificando-se por Δ_α o intervalo de valores de δ para os quais se verifica a equação $\eta(\alpha) = v(\delta) - \alpha\delta$, e por α_δ o intervalo de valores de α para os quais se verifica $\eta(\alpha) = v(\delta) - \alpha\delta$, uma análise cuidada dos gráficos de $v(\delta)$ e de $\eta(\delta)$, permite-nos ainda afirmar o seguinte:

- a cada intervalo de valores de α , α_δ , onde $\eta(\alpha)$ é afim, corresponde um ponto de $v(\delta)$ onde a derivada é descontínua;
- a cada intervalo de valores de δ , Δ_α , onde $v(\delta)$ é afim, corresponde um ponto de $\eta(\alpha)$ onde a derivada é descontínua;
- $\bigcup_{\alpha \in \mathbb{R}} \Delta_\alpha = \Delta$ e $\bigcup_{\delta \in \Delta} \alpha_\delta = \mathbb{R}$.

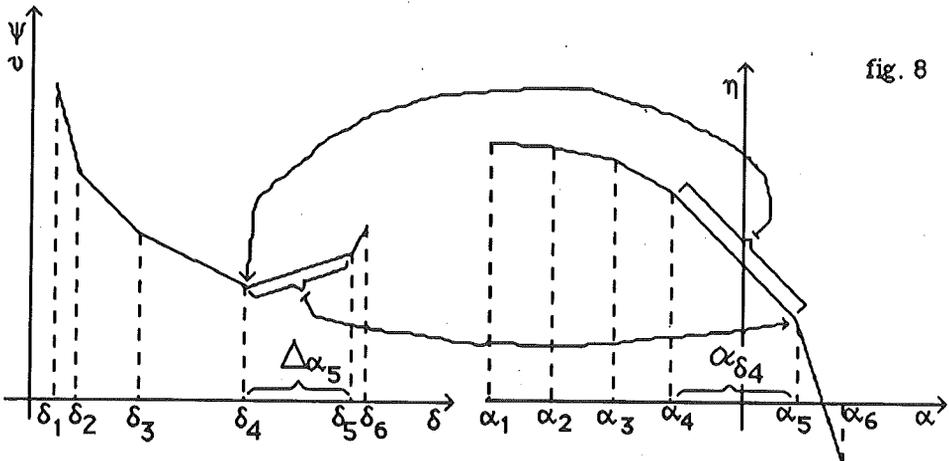


fig. 8

Referências

- [1] ABRHAM, J. and LUTTHRA, S., *Comparison of Duality Models in Fractional Linear Programming*, Zeitschrift für Operations Research, 21, pp. 125-130, (1977).
- [2] CARDOSO, D.M., *Programação Linear Fraccionária: Análise Comparativa das Principais Abordagens*, Dissertação de Mestrado em Ciências da Computação, Universidade de Coimbra, (1988).
- [3] CHADA, S.S., *A Dual Fractional Program*, ZAMM, 51, pp. 560-561, (1971).
- [4] CHARNES, A. and COPPER, W.W., *Programming with Linear Fractional Functionals*, Naval Research Logistics Quarterly, 9, pp. 181-186, (1962).
- [5] CLÍMACO, J.N. and CARDOSO, D.M., *Linear Fractional Programming: A New Bicriteria Approach*, Belgian Journal of Operations Research, Statistics and Computer Science, vol. 29, nº 3, pp. 4-24, (1990).
- [6] CRAVEN, B.D. and MOND, B., *The Dual of a Fractional Linear Program*, Journal of Mathematical Analysis and Applications, 47, pp. 507-512, (1973).
- [7] DINKELBACH, W., *On Nonlinear Fractional Programming*, Management Science, 13, pp. 492-498, (1967).
- [8] GEOFFRION, A.M., *Duality in Nonlinear Programming A Simplified Applications-Oriented Development*, SIAM, 13, pp. 1-37, (1971).
- [9] IBARAKI, T., *Parametric Approaches to Fractional Programs*, Mathematical Programming, 26, pp. 345-362, (1983).
- [10] ISBELL, J.R. and MARLOW, W.H., *Attrition Games*, Naval Research Logistics Quarterly, 3, pp. 71-93, (1956).
- [11] JAGANATHAN, R., *Duality for Nonlinear Fractional Programs*, Zeitschrift für Operations Research, 17, pp. 1-3, (1973).
- [12] MANGASARIAN, O.L., *Pseudo-Convex Functions*, J. SIAM Control, 3, pp. 281-290.
- [13] MANGASARIAN, O.L., *Nonlinear Programming*, Mc. Graw-Hill, (1969).

- [14] MARTOS B., *Hiperbolic Programming*, Naval Research Logistics Quarterly, 11, pp. 135-155, (1964).
- [15] SCHAIBLE, S., *Fractional Programming I, Duality*, Management Science, 22, pp. 858-867, (1976).
- [16] SHARMA, I.C. and SWARUP, K., *On Duality in Linear Fractional Functionals Programming*, Zeitschrift für Operations Research, 16, pp. 91-100, (1972).
- [17] SWARUP, K., *Some Aspects of Duality for Linear Fractional Functionals Programming*, ZAMM, 47, pp. 204, (1967).
- [18] WOLFE, H., *A Parametric Method for Solving the Linear Fractional Programming Problem*, Operations Research, 33, pp. 835-841, (1985).
- [19] WOLFE, P., *A Duality Theorem for Non-Linear Programming*, Quarterly of Applied Mathematics, 19, pp. 230-244, (1961).

PROGRAMAÇÃO QUADRÁTICA NÃO CONVEXA NA GESTÃO INTEGRADA DE MINI-HÍDRICAS

Victor Sousa e Ana M. Faustino
Faculdade de Engenharia
Universidade do Porto

Resumo: A actual solicitação intensiva de estudos e projectos para a instalação de mini-hídricas em Portugal sugere a eminência de um número elevado deste tipo de aproveitamentos numa mesma linha de água.

O modelo de gestão integrada de mini-hídricas agora implementado procura constituir-se num instrumento de análise à viabilidade e interesse de se gerir a capacidade de regularização conjunta de uma série de pequenos reservatórios em cascata.

O desenvolvimento deste modelo gera um problema quadrático não convexo, com uma condição adicional de complementaridade, que é resolvido iterativamente por um algoritmo Sequencial LCP (SLCP). Nesse processo o óptimo global é obtido a partir da resolução de um número finito de problemas lineares complementares (LCP), utilizando um método enumerativo híbrido.

1 - Introdução

O interesse que o projecto de mini-hídricas vem suscitando em Portugal potencia a ocorrência de vários daqueles empreendimentos numa mesma linha de água, o que conduzirá necessariamente à interdependência do respectivo funcionamento.

A reduzida capacidade de regularização, geralmente associada a pequenos aproveitamentos hidroeléctricos, conduz a modelos de exploração do tipo "fio-de-água" ou a modelos que procuram reproduzir as possibilidades de turbinamento concentrado de pequenos caudais, com arranque e paragem das turbinas regulados pelo nível da albufeira.

Em alternativa, procura formular-se um modelo que se aplique a sistemas de mini-hídricas no mesmo curso de água, aceitando a sua interdependência e tirando partido da capacidade de regularização conjunta com optimização da produção energética do sistema.

O modelo de gestão integrada de mini-hídricas foi desenvolvido no sentido de permitir a avaliação do interesse daquele tipo de gestão relativamente à gestão independente convencional.

A configuração-base do sistema de mini-hídricas que foi objecto de modelação prevê mini-hídricas implantadas na mesma linha de água e encerra as seguintes características fundamentais:

- açude galgável
- central no pé do açude com restituição próxima
- descarga de fundo, como solução mais corrente de descarga auxiliar

Com esta configuração, designadamente na consideração de vazão sobre o açude, procura encontrar-se uma situação abrangente relativamente a outras configurações de mini-hídricas, onde estruturas tais como câmaras de carga retiram a importância daquela vazão na definição da queda útil do aproveitamento.

Foi assim criado um modelo matemático que aproxima suficientemente bem a realidade e que consiste num problema de programação quadrática não convexa por segmentos com uma condição adicional de complementaridade. Para a resolução desse problema foi desenvolvido um algoritmo que se baseia na solução sequencial de problemas lineares complementares.

Este modelo, tendo como objectivo primeiro a optimização da energia produzida por um sistema numa dada sequência de unidades de tempo, pode também ser aplicado a uma mini-hídrica funcionando isoladamente, o que permite a comparação dos respectivos resultados com os modelos de exploração mais vulgarmente utilizados.

A estrutura do artigo é a seguinte. Na Secção 2 é descrita a construção do modelo de optimização. O algoritmo para a resolução desse problema é discutido na Secção 3. Finalmente os resultados de alguns ensaios e consequentes conclusões são apresentados na última secção.

2. Construção do Modelo de Optimização

2.1. - Variáveis do problema

O estudo de um sistema de mini-hídricas envolve os seguintes caudais integrais, com i e j designando respectivamente a posição da mini-hídrica no sistema e o número de ordem do intervalo de tempo em estudo:

- volumes afluentes $Q(i, j)$, previstos com base na caracterização do estado da bacia hidrográfica e no conhecimento das solicitações hídricas mais recentes;
- volumes $DS(i, j)$ descarregados pelo açude, associados à respectiva curva de vazão;
- volumes $DA(i, j)$ descarregados pela descarga auxiliar;
- volumes $T(i, j)$ turbinados;

- volumes $V(i, j)$ na albufeira i no final do período de tempo j .

Sendo $Q(i, j)$ um dado do problema, as restantes quatro variáveis contêm um grau de dependência traduzido pela equação de balanço hídrico

$$V(i, j) = V_{i,0} + \sum_{k=1}^j Q(i, k) - \sum_{k=1}^j [DS(i, k) + DA(i, k) + T(i, k)] + \sum_{k=1}^j [DS(i-1, k) + DA(i-1, k) + T(i-1, k)] \quad (1)$$

onde $V_{i,0}$ representa o volume inicial no reservatório i .

As variáveis do problema, para um estudo que envolva n mini-hídricas e m períodos de tempo, serão em número de $3n + m$ e dadas por $DS(i, j)$, $DA(i, j)$ e $T(i, j)$.

2.2. - Função Objectivo

A energia produzida por n mini-hídricas é dada por

$$E = \sum_{i=1}^n ETA_i \mu g h_i T_i \quad (2)$$

com μ e g valores constantes, ETA_i o rendimento do conjunto turbina-alternador, h_i a queda útil média explorável e T_i o volume turbinado.

Dado que h_i é função do volume médio da albufeira no intervalo de tempo em estudo, da carga sobre o açude devida a DS e do nível de água na restituição dependente de $DS + DA + T$, e como o rendimento é também função dos três tipos de variáveis, a equação (2) pode apresentar a formulação geral

$$E = \text{const. } f(DS, DA, T) \quad (3)$$

Sendo perfeitamente definida a contribuição do volume turbinado T como factor, optou-se por linearizar toda a restante expressão para que a função objectivo pudesse ser quadrática. Para tal, adoptaram-se aproximações da curva do rendimento (por patamares) e das curvas que contribuem para a queda útil (com interpolações lineares segmentadas), consideradas preferíveis às simplificações de natureza matemática em que se incorreria se se optasse pela resolução de um problema de programação de grau superior ao quadrático.

No que diz respeito à curva de rendimentos $f(DS, DA, T)$, modificada para introdução das perdas de carga no circuito hidráulico, o intervalo de discretização adoptado determina o número de patamares que são convenientemente indexados.

Quanto à queda útil $H(DS, DA, T)$, cada uma das rectas de interpolação das curvas $h(V)$ (de capacidade do reservatório), $h(Q)$ (da vazão sobre o açude) e $y(A)$ (da altura de água na restituição) foi assimilada a uma expressão do tipo

$$y = k_1 + k_2 x \quad (4)$$

sendo os parâmetros k_1 e k_2 indexados sob a forma $k(c, i, p)$, o que permite referenciá-los ao tipo de curva e ordem do parâmetro, ao reservatório a que se aplica, e à sua localização na sequência de rectas de cada interpolação [5].

A expressão (3), que constitui a função objectivo do problema, pode ser apresentada na forma

$$c^T x + 1/2 x^T Q x \quad (5)$$

em que o vector x terá $3 n m$ variáveis, correspondentes a um valor de $DS(i, j)$, $DA(i, j)$ e $T(i, j)$ para cada reservatório e cada mini-hídrica. O seu desenvolvimento, conforme a metodologia indicada, permite gerar o vector c dos termos independentes da função objectivo com $3 n m$ componentes e a matriz Q quadrada de ordem $3 n m$ [5]. O vector c tem apenas componentes associadas a $T(i, j)$, função do desnível geométrico $A(i)$ entre o coroamento do açude e o leito do rio na restituição, do volume inicial da albufeira, dos caudais próprios afluentes e de combinações dos $k(c, i, p)$. A matriz Q terá elementos associados a $T(i, j)^2$ e aos produtos cruzados de $T(i, j)$ com $DS(i, j)$ e $DA(i, j)$.

2.3. - Condições de Restrição

As condições de restrição exigidas pelo problema são do tipo $x \geq 0$ (todas as variáveis só podem tomar valores não negativos) e ainda da forma $Ax \geq b$, resultantes de outros limites a impor. Assim

$$V_{m,i} \leq V(i, j) \leq V_{M,i} \quad (6)$$

determinam os volumes mínimo e máximo no reservatório i . O volume máximo de turbinamento, para o intervalo de tempo em estudo, é imposto pela condição

$$T(i, j) \leq T_{M,i} \quad (7)$$

e as condições de volume máximo ou mínimo a ser produzido pela descarga auxiliar no mesmo intervalo de tempo são dadas pelas expressões

$$\begin{aligned} DA(i, j) &\leq DA_{M,i} \\ &\text{e/ou} \\ DA(i, j) &\geq DA_{m,i} \end{aligned} \quad (8)$$

Com uma destas últimas condições (8), a condição (7) e ainda as condições (6), desenvolvidas de acordo com a expressão (1), o modelo gera a matriz das restrições A de ordem $4 n m$ por $3 n m$ e o vector b com $4 n m$ componentes, função dos limites das condições de restrição.

2.4. - Condição Adicional de Complementaridade

Podem ocorrer situações em que a transferência de água de uma mini-hídrica para a situada imediatamente a jusante seja exigida pelas condições de optimização energética. Nos casos em que a indisponibilidade da central não seja total, o modelo poderia recorrer à descarga no açude, por ser mais favorável em termos de queda útil, sem que o reservatório estivesse cheio e violando por isso a realidade física.

Porque havia de remeter essa transferência, em tais casos, para a descarga de fundo, houve que recorrer a uma condição adicional de complementaridade, entre o volume descarregado pelo açude e o volume livre da albufeira, do tipo

$$s(i, j)^T DS(i, j) = 0 \quad (9)$$

onde $s(i, j)$ é uma variável de afastamento dada por

$$s(i, j) = V_{M,i} - V(i, j) \quad (10)$$

Impõe-se assim que só haja descarga pelo açude ($DS(i, j) > 0$) quando o reservatório estiver cheio ($s(i, j) = 0$) e que, por outro lado, estando este abaixo do nível máximo ($s(i, j) > 0$) terá a descarga superior de ser nula ($DS(i, j) = 0$).

Passaremos então a ter um problema quadrático com uma condição de complementaridade cuja resolução é feita à custa de um problema equivalente do tipo

$$\begin{aligned} &\text{minimizar} - (c^T x + 1/2 x^T Q x) + p s^T x_D \\ &\text{sujeito a} \quad A x \geq b \\ &\quad \quad \quad x \geq 0 \end{aligned} \quad (11)$$

onde x_D é um sub-vector de x correspondendo às componentes do tipo $DS(i, j)$, s o vector das variáveis $s(i, j)$ definidas em (10) e p um escalar positivo e suficientemente grande para que a minimização de $s^T x_D$ ocorra para o valor zero.

Caso o problema quadrático seja resolvido satisfazendo em si mesmo a condição de complementaridade, torna-se desnecessária a intervenção de p . No entanto, no caso contrário, este parâmetro terá um efeito de penalização sobre a nova parcela da função objectivo e será incrementado, tanto quanto necessário, até que a contribuição dessa parcela se torne nula.

O desenvolvimento de $s^T x_D$ nas variáveis do problema conduz a uma expressão com a forma

$$s^T x_D = v^T x + 1/2 x^T C x \quad (12)$$

onde C é uma nova matriz quadrada de ordem $3 n m$ e v um vector de $3 n m$ componentes [5].

3 - Resolução do Problema de Optimização

Como se infere dos pontos anteriores, o problema quadrático global passa a ser do tipo

$$\begin{aligned} &\text{minimizar} && - (c^T x + 1/2 x^T Q x) + p (v^T x + 1/2 x^T C x) \\ &\text{sujeito a} && A x \geq b \\ &&& x \geq 0 \end{aligned} \quad (13)$$

onde nenhuma das matrizes Q e C é positiva semi-definida. A resolução do problema é então equivalente a um MLCP do tipo [2]

$$\begin{aligned} &\text{minimizar} && (-c + p v)^T x + b^T \phi \\ &\text{sujeito a} && \mu = (-c + p v) + (-Q + p C) x - A^T \phi \\ &&& \theta = -b + A x \\ &&& x, \phi, \mu, \theta \geq 0 \\ &&& \mu^T x = \phi^T \theta = 0 \end{aligned} \quad (14)$$

onde Q não é constante, dependendo da combinação dos intervalos relativos às interpolações lineares das curvas que definem a queda útil. O problema quadrático terá por isso de ser resolvido iterativamente até haver coincidência entre todos os intervalos de entrada (cada um deles função de variáveis que designamos por VS) e os intervalos correspondentes à última solução do problema.

Considerando os novos vector e matriz

$$\bar{c} = -c + p v \quad (15)$$

$$\bar{Q} = -Q + p C$$

o algoritmo do problema é então como se segue:

Passo 1

Fixe \bar{Q} inicialmente, de acordo com o conjunto de arranque das variáveis VS (abscissas das curvas que definem H) e com $p=0$ (não impondo a condição adicional de complementaridade).

Passo 2

Resolva o MLCP. Se a solução obtida satisfizer a condição adicional de complementaridade vá para o Passo 4. Caso contrário vá para o Passo 3.

Passo 3

Altere \bar{Q} , com incremento do tipo

$$p = p_i = p_{i-1} + p_{inc} \quad (16)$$

(onde p_{inc} é o incremento de p) para forçar a complementaridade a ser satisfeita, e volte ao Passo 2.

Passo 4

Construa um novo conjunto de VS a partir da solução obtida, e verifique a correspondência entre os intervalos de interpolação ligados àquele conjunto e os referentes aos valores de VS que intervieram na construção de \bar{Q} . Se tal se verificar, o processo termina e obteve-se a solução do problema. Caso contrário, \bar{Q} é redefinida com as VS actualizadas e regressa-se ao Passo 2, tomando-se assim como solução inicial uma solução básica cujas variáveis básicas e não básicas têm índices iguais às variáveis básicas e não básicas da última solução obtida.

Para a resolução do MLCP utiliza-se um algoritmo SLCP [3], que consiste na resolução de um sequência de Problemas Lineares Complementares (LCP). Cada LCP é resolvido por um método enumerativo híbrido [4], que incorpora o algoritmo de gradiente reduzido de Al-Khayyal [1] e algumas regras heurísticas. Nesse algoritmo é introduzido um parâmetro w , sendo a função objectivo substituída por uma restrição do tipo

$$\bar{c}^T x + b^T \phi \leq w \quad (17)$$

O óptimo global x do problema quadrático é a solução do LCP (w), onde w é o último valor de uma sucessão decrescente w_k para o qual o LCP (w_k) tem solução. Os valores daquela sucessão são obtidos com a utilização de um incremento e_k na respectiva expressão, que no caso foi tomado igual $0.01 w_k$.

O método enumerativo híbrido aplicado na resolução da sequência dos LCP (w_k) procura encontrar uma solução complementar analisando apenas soluções básicas realizáveis. Para tal explora uma árvore binária que na pior das hipóteses terá 2^n ramos, onde n é o número de pares de variáveis complementares. O método incorpora algumas regras heurísticas que o tornam eficiente. Além disso, em cada nodo usa-se uma modificação do método de Al-Khayyal [1] em que, a partir da solução básica associada a

cada nodo, se passa sucessivamente para soluções básicas adjacentes, diminuindo o somatório de $z_i y_i$. Devido a essas modificações o algoritmo encontra a solução complementar normalmente nos primeiros nodos.

Como a solução do LCP (w_k) pouco difere da solução do LCP (w_{k-1}), a base associada à solução deste último é utilizada para solução inicial do LCP (w_k), o que normalmente ainda vai reduzir mais o esforço computacional do método enumerativo híbrido.

4. Ensaios e Conclusões

Os ensaios iniciaram-se sobre um sistema de duas mini-hídricas actualmente em exploração no rio Vizela, nos lugares de Negrelos e Caniços, e que se ajustam no essencial à configuração-base apresentada no ponto 1. Foram ensaiadas várias combinações de solicitações, diferentes curvas de capacidades das albufeiras e de turbinamento, comparando-se sempre os resultados energéticos obtidos com os que decorriam da exploração isolada de ambos os aproveitamentos, fazendo-se afluir ao reservatório de jusante (Caniços) o conjunto de efluências determinadas pela exploração de Negrelos.

As comparações efectuadas com base no mesmo programa de optimização conduziram a variações mais sensíveis de acordo com o número de intervalos de tempo (para a sequência de volumes afluentes), o grau de discretização no tempo e, sobretudo, com o valor médio e distribuição daqueles volumes. Assim, com base numa série histórica de caudais médios diários, foi possível encontrar acréscimos energéticos de 1.9% em meses secos e valores não significativos para meses húmidos. Entretanto, invertendo-se a posição das suas mini-hídricas (o volume do reservatório de Caniços é consideravelmente superior) encontraram-se produções superiores para o sistema, que se traduziram num acréscimo de 3.3% nos referidos meses secos.

Havia no entanto que generalizar a validade dos resultados, pelo que se optou por ensaiar sistemas com número variável de aproveitamentos iguais, com sequências de caudais iguais com diferente extensão. Para que a interpretação dos resultados pudesse traduzir a referida generalização, introduziram-se dois parâmetros:

- tempo de retenção t_r definido pelo quociente entre o volume útil médio dos reservatórios e o caudal médio afluente, com o que se pretende traduzir o significado das efluências independentemente do reservatório que as recebe;
- eficiência da regularização conjunta e_{rc} que indica o aumento percentual relativamente à gestão independente.

Conforme evidencia a figura 1, estes ensaios permitiram a construção de curvas de igual eficiência para um número variável de reservatórios e sequências de caudais iguais correspondentes a diferentes tempos de retenção. Os resultados correspondem ao que seria de esperar, ou seja, a eficiência de regularização aumenta com o número de reservatórios do sistema e com os caudais aos quais correspondem maiores tempos de retenção.

De referir no entanto que valores superiores de eficiência foram encontrados para sequências de caudais não uniformes e quando o volume do reservatório de montante, V_m , era consideravelmente superior ao de jusante, V_j , o que é evidenciado pela figura 2, em que se apresentam resultados para sistemas de duas mini-hídricas com o mesmo volume de encaixe total, mas diferentes valores de $n=V_m/V_j$.

Interessará também referir uma característica constatada na generalidade dos ensaios realizados: a eficiência da gestão integrada nunca é obtida à custa de perdas para qualquer das mini-hídricas dos sistema que se possam considerar significantes. O que valerá por dizer que a repartição de propriedade e usufruto dos vários aproveitamentos poderá não constituir obstáculo deste processo de gestão.

O modelo implementado revelou-se consistente com tempos de execução que tornam possível a operação em tempo real e, embora sem ter sido demonstrado, sempre convergente no processo iterativo resultante da utilização de interpolações segmentadas. Além disso está adequado à variação de todos os parâmetros envolvidos no problema, podendo ser simulados graus de discretização no tempo sem grandes limitações. Os reduzidos tempos de execução resultam em parte do facto de se ter deixado de exigir a resolução do LCP (w_k) inadmissível sempre que $p = 0$, já que se verificou em todos os ensaios realizados que a solução óptima ocorria para o LCP (w_0). Na verdade, a intervenção da condição adicional de complementaridade só em condições excepcionais de exploração poderá ser exigida e conduzir portanto à extensão desses tempos de execução. Poderá isto querer dizer que, nas situações correntes de funcionamento do sistema, a solução óptima global é única.

Por outro lado, com este modelo, é possível simular situações de indisponibilidade de uma das centrais, produzindo efluências através da descarga auxiliar, rentabilizando o resto do sistema e deixando a central indisponível de funcionar como simples agente passivo de retenção.

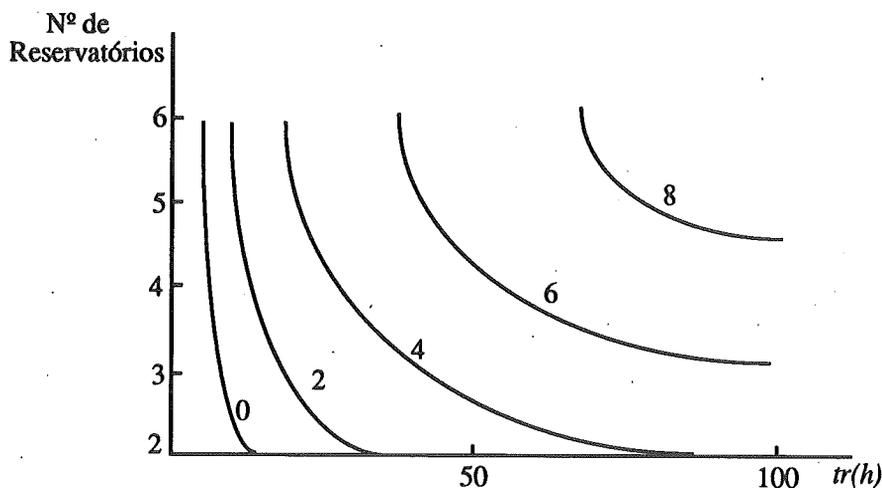


Fig. 1 – Curvas de igual eficiência para número de reservatórios variável e sequências de três dias

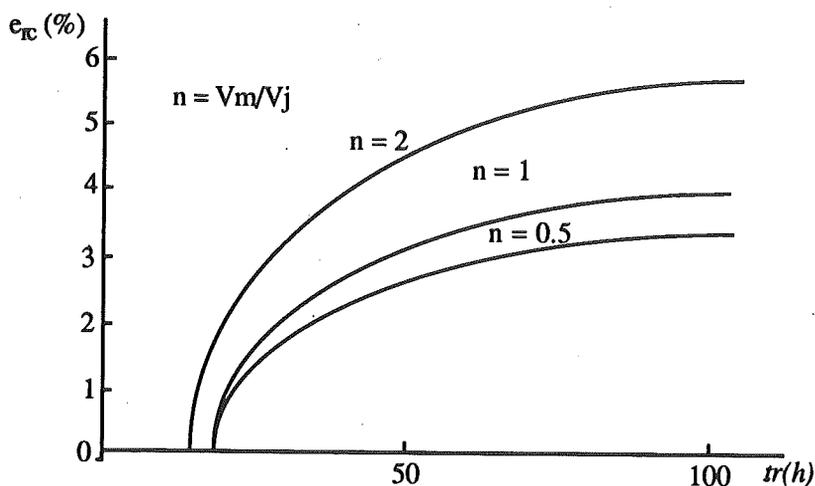


Fig. 2 – Influência da relação montante-jusante n entre volumes de albufeira

O modelo, adaptando-se a todas as configurações que envolvem mini-hídricas, inclusive a sua interligação com grandes aproveitamentos, dispõe da possibilidade de se aplicar à gestão de aproveitamentos isolados com vantagens sensíveis relativamente aos modelos de exploração normalmente utilizados.

Os resultados obtidos ao longo dos ensaios sugerem algumas cautelas quanto ao interesse da gestão integrada das mini-hídricas, até porque os valores de eficiência com

algum significado aparecem para sistemas com número de reservatórios elevado, para previsões de vários dias e para caudais reduzidos.

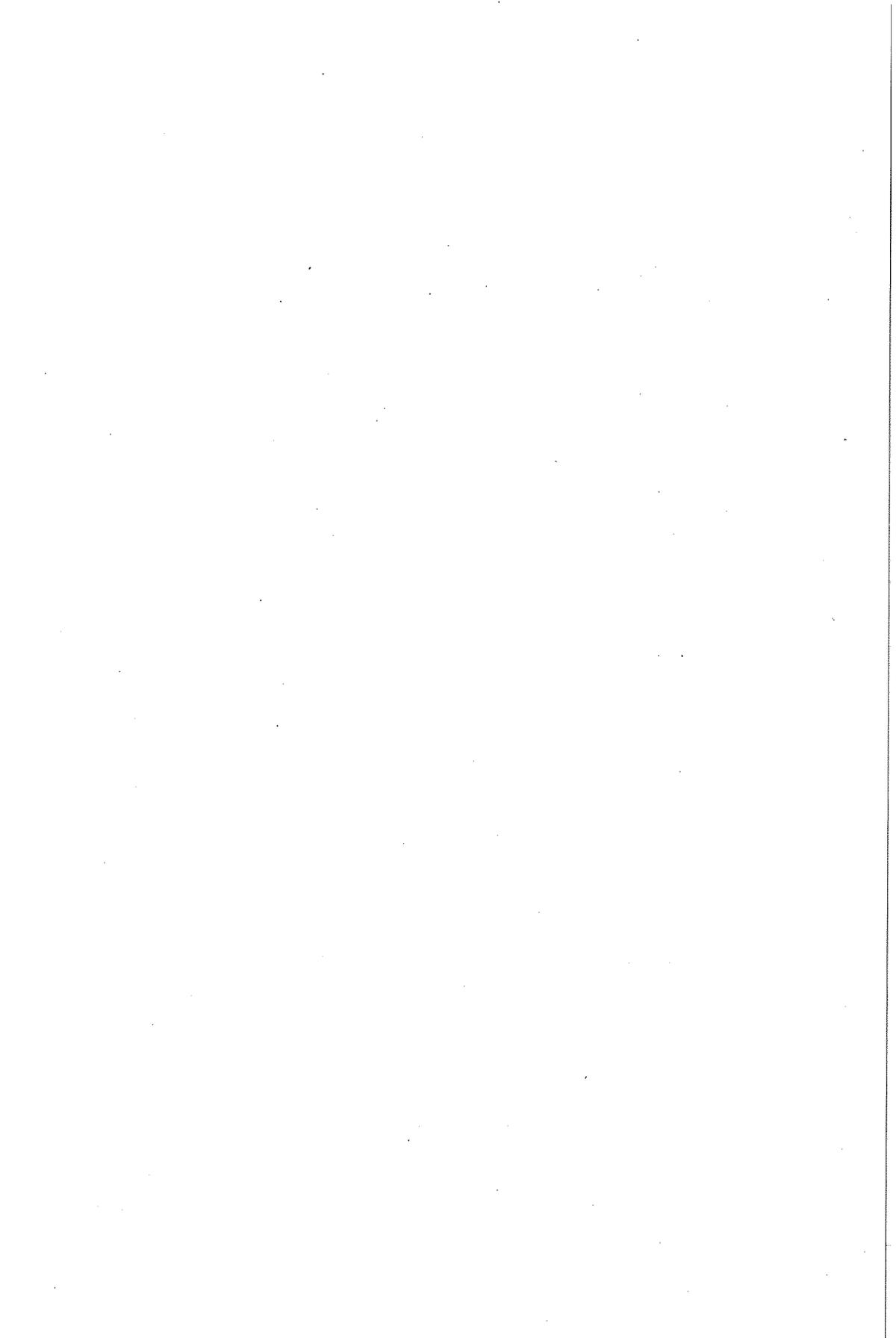
Há no entanto um lado positivo, que o modelo confirmou em plenitude e que assume alguma relevância. Trata-se do facto de os ganhos energéticos mais importantes se encontrarem em alturas de estiagem, quando é maior o défice energético.

Uma reflexão merece também o facto de se estar a trabalhar fundamentalmente à escala diária, no que se poderia chamar um segundo nível da exploração das mini-hídricas. Sobre esta optimização a nível diário poderá ser aplicada a mesma metodologia, conseguindo-se então a este primeiro nível de gestão sub-diária resultados ainda mais favoráveis. Muito residirá, ao fim ao cabo, na transmissão e tratamento imediatos de informações periódicas sobre o estado da bacia hidrográfica e das suas linhas de água, para que a operação em tempo real possa ser uma realidade.

Ainda uma referência final ao possível interesse da interligação de mini-hídricas com grandes aproveitamentos, que resultou reforçado do desenvolvimento deste trabalho. Um modelo do tipo que se implementou estará adequado a uma nova definição do regime de descargas das grandes albufeiras, que normalmente se fazem de modo aleatório, satisfazendo necessidades pontuais de apoio à agricultura ou simples preocupações ecológicas.

BIBLIOGRAFIA

- [1] F. A. Al-Khayyal – *An implicit enumeration procedure for the general linear complementarity problem*, Mathematical Programming Study 31 (1987) 1-20.
- [2] J. J. Júdice – *A study of the linear complementarity problems*, Tese de Doutoramento, Brunel University, 1982.
- [3] J. J. Júdice, A. M. Faustino – *A sequential LCP algorithm for bilinear and concave quadratic programs*, Investigação Operacional 8 (2) (1988) 67-87.
- [4] J. J. Júdice, A. M. Faustino – *An experimental investigation of enumerative methods for linear complementarity problems*, Computers and Operations Research, 15 (1988) 417-426.
- [5] V. Sousa – *Modelação da Gestão Integrada de Mini-Hídricas*, Dissertação de Mestrado, FEUP, 1989.



SISTEMA DE APOIO A GESTÃO E PROGRAMAÇÃO FINANCEIRA DE PROJECTOS

José Filipe Rafael
Universidade Católica Portuguesa
FCEE/CEPI

Resumo

Apresenta-se um sistema com o objectivo de permitir a construção e avaliação financeira de calendários para projectos definidos como um conjunto de actividades para as quais é conhecida uma relação de precedências imediatas.

1. Apresentação do Problema

São bem conhecidos os modelos de gestão de projecto construídos a partir do Método do Caminho Crítico (CPM). Este assume que um projecto é constituído por um conjunto de m actividades no qual é definida uma relação de precedência imediata, e por isto representável sob a forma de uma rede acíclica. Cada actividade é caracterizada pela sua duração, custos (ou consumo de recursos) e, enquanto arco na rede que representa o projecto, pelos nós inicial e final.

Quando se admite a duração de cada actividade como determinística, o problema que tradicionalmente é resolvido pelo modelo é o da minimização da duração total do projecto sujeita ou não a restrições na disponibilidade dos recursos. Ao ser resolvido fica identificada a sequência de actividades com maior duração e por isso designada de caminho crítico.

Matematicamente o problema pode ser formulado como um modelo de programação linear tomando como referencial de tempo o início do projecto. O seu dual é resolúvel por inspecção por ordem progressiva dos nós, permitindo, para cada nó i da rede que descreve o projecto, a identificação do tempo mais cedo de realização do nó TC_i . Formulação análoga para

determinação do caminho mais longo entre o nó n (nó final que representa a conclusão do projecto) e o nó 1 (início do projecto) permite determinar, por ordem regressiva dos nós, o tempo mais tarde de cada nó TT_i , condicionados à duração mínima do projecto. A formulação dos problemas aqui mencionados é revista em [6] sendo sugeridos procedimentos para a sua resolução.

O problema que aqui abordamos é o da calendarização das actividades do projecto, ou seja o da fixação de um momento para início de cada uma das actividades a do projecto, I_a , com o objectivo de maximizar indicadores financeiros para o projecto. Como do início de cada actividade depende a fixação no tempo dos fluxos financeiros associados à actividade (custos e benefícios), do calendário a estabelecer dependerão os indicadores financeiros para o projecto.

Se as durações das actividades não forem consideradas determinísticas, mas sim variáveis aleatórias (independentes ou não) com função de distribuição conhecida, então a duração do projecto será também aleatória. Podemos então definir o risco de atraso do projecto como a probabilidade de a sua duração ultrapassar um limite pré-definido, por exemplo, a duração mínima do projecto quando todas as actividades tiverem duração igual ao respectivo valor esperado.

2. Formulação

O problema, admitindo durações determinísticas, foi já formulado e resolvido com recurso a Programação Binária por Doerch e Patterson ([2] e [5]). Trata-se de uma formulação que exige, mesmo para exemplos de projectos simples, o recurso a um grande número de variáveis pelo que a sua resolução é morosa.

Aqui propomos uma formulação mais simples, porque admitimos as variáveis contínuas, e a partir dela sugerimos um procedimento interactivo para construção e avaliação de calendários que satisfaçam as diferentes restrições impostas ao problema.

O objectivo de minimização da duração do projecto é considerado impondo a I_a a restrição de se encontrar entre um mínimo e um máximo conformes ao tempo mais cedo para o seu nó de início e ao tempo mais tarde para o seu nó de fim, respectivamente $TC_{I(a)}$ e $TT_{F(a)} - t_a$, onde $I(a)$ e $F(a)$ denotam os nós de início e fim de a e t_a a sua duração.

A função objectivo do problema será o indicador financeiro a escolher. Pela sua generalização escolhemos o valor actual líquido (VAL). O único recurso considerado é o capital e não se consideram limitações na sua disponibilidade.

Se fôr v_a o valor líquido da actividade actualizado para o seu início (ignorando assim a distribuição ao longo da realização da actividade dos custos e benefícios que lhe estão associados, o que é irrelevante para o problema em discussão) e se fôr tx a taxa de actualização a usar, o VAL do projecto é a soma, para as m actividades que o constituem, do valor de cada uma actualizado do respectivo início para o início do projecto (momento 0). O objectivo, em função das variáveis de decisão I_a , vem:

$$\text{MAX VAL} = \sum_{a=1}^m v_a \cdot (1 + tx)^{-I_a}$$

Passamos de seguida à formulação das restrições impostas a I_a . Lembremos que, para cada actividade a individualmente considerada

$$TC_{I(a)} \leq I_a \leq TT_{F(a)} - t_a$$

Não podemos, porém esquecer a inter-dependência entre os inícios das actividades:

- $TC_{I(a)}$ só é limite inferior para o valor a atribuir ao início da actividade a enquanto admitirmos que o início das actividades que constituem o caminho mais comprido da origem ao nó $I(a)$ se fez tão cedo quanto possível e que o início de todas as outras actividades precedentes (imediate ou mediatamente) de a se fixou por forma a não atrasarem estas.
- $TT_{F(a)} - t_a$ só é limite superior para I_a enquanto admitirmos que as actividades que constituem o caminho mais longo de $F(a)$ a n se

iniciam tão tarde quanto possível e que todas as outras actividades com precedência de a (imediate ou mediate) são iniciadas por forma a necessitarem a antecipação destas.

Há, então que rever os limites para I_a . Se os admitirmos associados ao nó s inicial da actividade, podemos escrever

$$l_{I(a)} \leq I_a \leq L_{F(a)} - t_a$$

onde, l_i e L_i definidas para cada nó i , são também variáveis do problema e definidas como

$$l_i = \text{MAX}_{a \in J(i)} \{ I_a + t_a \} \quad , \quad i = 2, \dots, n$$

e

$$L_i = \text{MAX}_{a \in L(i)} I_a \quad , \quad i = 1, \dots, n-1$$

onde $L(i)$ e $J(i)$ são os conjuntos de actividades com, respectivamente, início e fim no nó i . Assim continuamos a ter o problema formulado em termos das variáveis I_a ($a = 1, \dots, m$):

$$\text{MAX}_{b \in J(I(a))} \{ I_b + t_b \} \leq I_a \leq \text{MIN}_{c \in L(F(a))} I_c - t_a$$

Admitimos para início do projecto $l_1 = TC_1 = 0$ e concluímos a minimização da duração do projecto impondo $L_n = TT_n = TC_n$.

As restrições podem desdobrar-se, a de minoração em

$$\left\{ \begin{array}{l} I_{b_1} + t_{b_1} \leq I_a \\ I_{b_2} + t_{b_2} \leq I_a \\ \cdot \quad \cdot \quad \cdot \\ I_{b_R} + t_{b_R} \leq I_a \end{array} \right.$$

onde b_1, b_2, \dots, b_R são os R elementos de $J(I(a))$.

A condição de majoração pode desdobrar-se em

$$\left\{ \begin{array}{l} I_a \leq I_{c_1} - t_a \\ I_a \leq I_{c_2} - t_a \\ \cdot \cdot \cdot \\ I_a \leq I_{c_S} - t_a \end{array} \right.$$

onde c_1, c_2, \dots, c_S são os S elementos de $L(F(a))$.

Uma vez que cada actividade a faz parte de $L(F(b_r))$ e de $J(I(c_s))$, na construção dos limites ao início de b_r e de c_s aparecerão as condições

$$I_{b_r} \leq I_a - t_{b_r}$$

e

$$I_a + t_a \leq I_{c_s}$$

o que significa que temos duplicação de restrições. Exceptuam-se as actividades para as quais $J(I(a))$ é vazio porque $I(a) = 1$, ou para as quais $L(F(a))$ é vazio porque $F(a) = n$. Para estas actividades, que designamos por iniciais e finais, vem

$$I(a) = 1 \Rightarrow TC_1 = 0 \leq I_a$$

$$F(a) = n \Rightarrow I_a \leq TT_n - t_a$$

O problema fica assim formulado em programação mista:

$$\text{MAX VAL} = \sum_{a=1}^m v_a \cdot (1 + tx)^{-I_a}$$

s.a., para cada actividade a

$$\left\{ \begin{array}{ll} 0 \leq I_a & \Leftarrow I(a) = 1 \\ I_a \leq TT_n - t_a & \Leftarrow F(a) = n \\ I_a \leq I_c - t_a, c \in L(F(a)) & \Leftarrow F(a) \neq n \end{array} \right.$$

3. Decomposição do Problema

Note-se, porém, que uma duração para o projecto igual a TT_n implica que para as actividades que fazem parte do caminho crítico a folga (definida como $L_F(a) - t_a - I_I(a)$) seja 0. Se fôr a sequência de P actividades $k_1, k_2, k_3, \dots, k_p$ então para qualquer actividade p na sequência vem

$$I_{I(k_p)} = \text{MAX}_{a \in J(I(k_p))} \{I_a + t_a\} = I_{k_{p-1}} + t_{k_{p-1}}$$

e

$$L_{F(k_p)} = \text{MIN}_{c \in L(F(k_p))} I_c = I_{k_{p+1}}$$

pelo que o seu início de k_p deixou de ser variável de decisão no problema:

$$I_{k_{p+1}} = I_{k_{p-1}} + t_{k_{p-1}} + t_{k_p}$$

Atendendo a que a função objectivo é aditiva nos termos em I_a , o problema pode decompor-se numa série de problemas do tipo

$$\text{MAX} \quad \text{VAL}_\epsilon = \sum_{a=1}^m v_a \cdot (1 + tx)^{-I_a}$$

s.a., para cada actividade $a \in \epsilon$

$$\begin{cases} 0 \leq I_a & \Leftarrow I(a) = 1 \\ I_a \leq TT_n - t_a & \Leftarrow F(a) = n \\ I_a \leq I_c - t_a, c \in L(F(a) \cap \epsilon) & \Leftarrow F(a) \neq n \end{cases}$$

onde ϵ é um conjunto de actividades com incícios interdependentes entre si, mas independentes de todas as outras actividades (críticas ou não). A estes grupos chamamos grupos de partilhas de folgas.

Porque vamos estar interessados em tirar partido desta partição das actividades sem ser para resolver os problemas que acima formulamos, propomos o seguinte procedimento para identificação dos subgrupos de partilha de folgas:

- Passo 1:** encontrar uma actividade não crítica ainda não examinada. Esta é a actividade em exame.
- Passo 2:** se o nó inicial da actividade em exame não é crítico, incluir no mesmo grupo de partilha todas as actividades que têm fim ou início nesse nó e que não tenham ainda sido examinadas.
- Passo 3:** se o nó final da actividade em exame não é crítico, incluir no mesmo grupo todas as actividades que tenham fim ou início nesse nó e que não tenham ainda sido examinadas.
- Passo 4:** repetir os passos 2 e 3 para todas as actividades incluídas no grupo de partilha ainda não examinadas. O fim deste passo dá-nos um grupo de partilha.
- Passo 5:** re-iniciar o processo a partir do passo 1 até que todas as actividades não críticas tenham sido examinadas. O fim deste passo garante-nos que todos os grupos de partilha foram identificados.

Para cada grupo de partilha pode estudar-se a sub-rede definida: como se verá nos pontos seguintes têm interesse a enumeração dos caminhos que a constituem e, muito especialmente, a identificação, para cada actividade, do número de actividades que constituem o mais complexo dos carinhos a que pertence, n_a .

Para tanto recorreu-se a uma estrutura de apontadores indexados por forma que cada índice corresponda a um caminho possível. Encontrar novos caminhos é, numa exploração progressiva, encontrar um nó não crítico de semi-grau positivo maior que 1. Tal procedimento dá origem a uma estrutura em árvore com células que vão sendo partilhadas por diversas que nela ficam penduradas. Há assim um tronco comum que vai sendo ramificado até terminar em tantos ramos quantos os caminhos. Pela leitura das sequências construídas é depois fácil determinar n_a .

4. Implementação de um Sistema de Apoio à Programação Financeira de Projectos

A figura 1 apresenta uma síntese do diagrama de fluxos de dados (DFD) do sistema: sobre a base de dados operam quatro rotinas que passamos a descrever.

- 4.1 A primeira rotina permite o manuseamento de dados (introdução, modificação, cópia, eliminação, etc.)
- 4.2 A segunda rotina propõe ao utilizador a resolução do modelo CPM, produzindo os resultados referidos em 1 (tempos mais cedo e tempos mais tarde para cada nó de rede, caminho crítico e inícios mais cedo e mais tarde para cada actividade). Com a determinação destes ficam definidos dois calendários que designamos de automáticos: o que tem os inícios mais cedo e o que tem os inícios mais tarde para todas as actividades. Estes dois calendários correspondem a duas situações extremas em termos de risco e de desempenho financeiro: o primeiro é o de menor risco de atraso mas, porque implica uma antecipação de custos, tem piores indicadores financeiros. O segundo, porque se adiaram o mais possível os custos, corresponde a melhores indicadores financeiros mas também ao risco de atraso máximo.
- 4.3 Soluções intermédias são desejáveis mas não automáticas e a respectiva construção é possível com o auxílio da terceira rotina que permite uma inter-acção modelo/utilizador para construção de calendários em alternativa à programação matemática apresentada no ponto 2 e tirando partido da decomposição proposta no ponto 3.

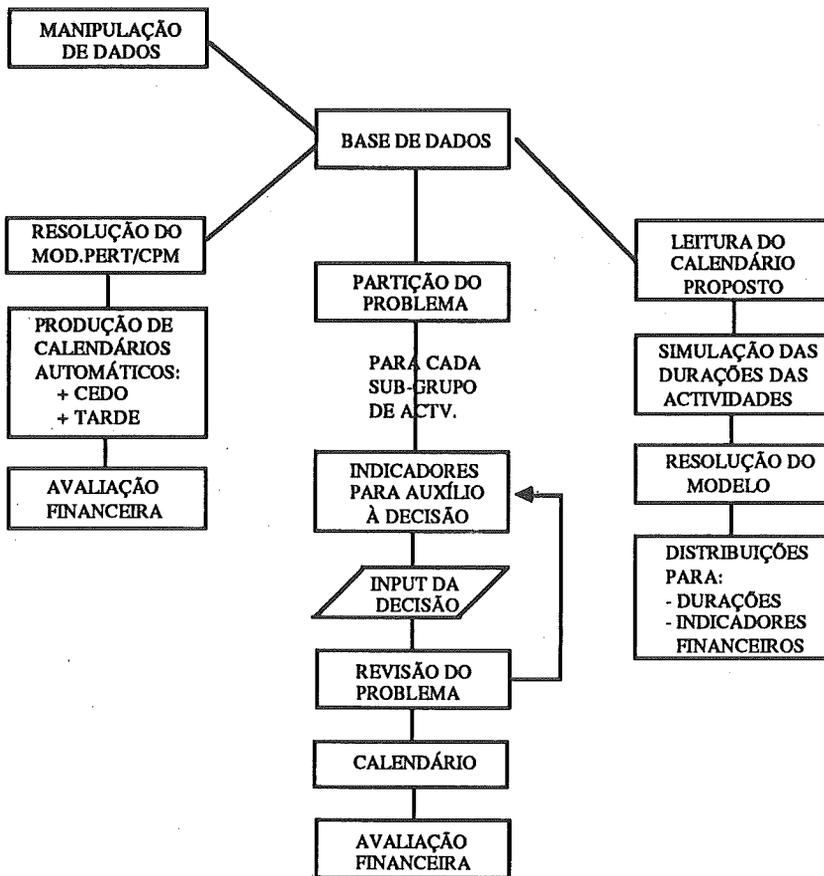


Figura 1 – Síntese do DFD do Sistema

Para cada sub-rede de partilha de folgas e para a decisão do início a fixar para cada actividade, o sistema fornecerá ao utilizador elementos perante os quais ele decide o início de uma actividade não crítica. Então o modelo revê a informação sobre as restantes actividades e o utilizador é de novo chamado a decidir. E a tomada de decisões é sequencial até que os inícios de todas as actividades não críticas estejam definidos e então avaliado o calendário construído.

Apresenta-se de seguida a rotina que é utilizada para gerir as folgas de cada sub-rede de partilha em separado uma vez que em cada uma o problema é independente do das outras. Porém, a avaliação faz-se sempre para todo o projecto, justamente porque sendo sub-redes paralelas, isto é, ocorrendo

simultaneamente, exigem investimento em simultâneo e como tal perderíamos as vantagens que se apontam para o sistema se as avaliássemos em separado.

Para cada sub-rede de partilha:

Passo 1: apresenta-se ao utilizador uma lista das actividades que a constituem e para cada actividade a dão-se indicadores para auxílio à decisão.

Os indicadores sobre as actividades com início por decidir que se julgam convenientes são os seguintes:

- os limites para o início que, para além de evitarem decisões impossíveis, dão uma ideia, consoante a amplitude do intervalo, do 'espaço de manobra' do decisor.
- u_a → número de actividades (da sub-rede de que a faz parte) sem início definido e em sequência a . É um bom indicador do número de minorantes ou majorantes para os inícios de outras actividades (a jusante ou a montante) que poderá ser revisto devido à fixação do início da actividade a e que é facilmente determinado por um procedimento semelhante ao que no ponto 3 se mencionou para calcular n_a .
- v_a → valor actualizado para o início da actividade do seu benefício líquido. Dá-nos uma ideia da importância da actividade tanto na recuperação do investimento se for positivo como na distribuição do investimento se for negativo.
- w_a que procura dar uma ideia da grandeza dos benefícios associados aos nós da sub-rede com realização dependente do início da actividade a , uma vez que é definido como a sua soma, actualizada para esse momento:

$$w_a = \sum_{j \in M_a} \frac{B_j}{(1+tx)^{-(TC_j - L_a)}}$$

onde M_a é o conjunto dos nós atingíveis por actividade não calendarizadas a partir de a , B_j o benefício líquido do nó j , tx a taxa de actualização e $(TC_j - L_a)$ é o tempos mais cedo do nó j admitindo a iniciada em L_a (minorante para o início).

- Passo 2:** pede-se ao utilizador que fixe o início de uma destas actividades. Os atributos apresentados para as actividades b , ainda não calendarizadas, são, sempre que se fixa o início de a , revistos como em seguida se descreve:
- a) revêm-se os das actividades a jusante e os (majorantes para o início) das actividades a montante por utilização de um procedimento semelhante ao CPM mencionado no ponto 1, modificado porque se separa a aplicação da parte progressiva da regressiva e se altera a condição de paragem que é conjugada com a condição de se tratar de uma alternativa não calendarizada.
 - b) revêm-se os valores de μ_b por um procedimento como o mencionado no passo 1, mas da mesma forma revisto na condição de inclusão de uma actividade.
 - c) recalculam-se os w_b .

Passo 3: revêm-se os parâmetros apresentados e repetem-se os passos 2 e 3 até que estejam calendarizadas todas as actividades do grupo. Proceder-se-á então à avaliação financeira do calendário proposto.

4.4 A quarta rotina permite a construção por simulação das distribuições de probabilidade para durações e indicadores financeiros dos calendários construídos. Por isso permite quantificar o risco de atraso do projecto, tal como definido no ponto 1 e o risco de um VAL negativo por construção do respectivo perfil de risco.

5. Conclusão

Um procedimento deste tipo não garante, obviamente, uma solução óptima para o problema proposto. Mas tem vantagens:

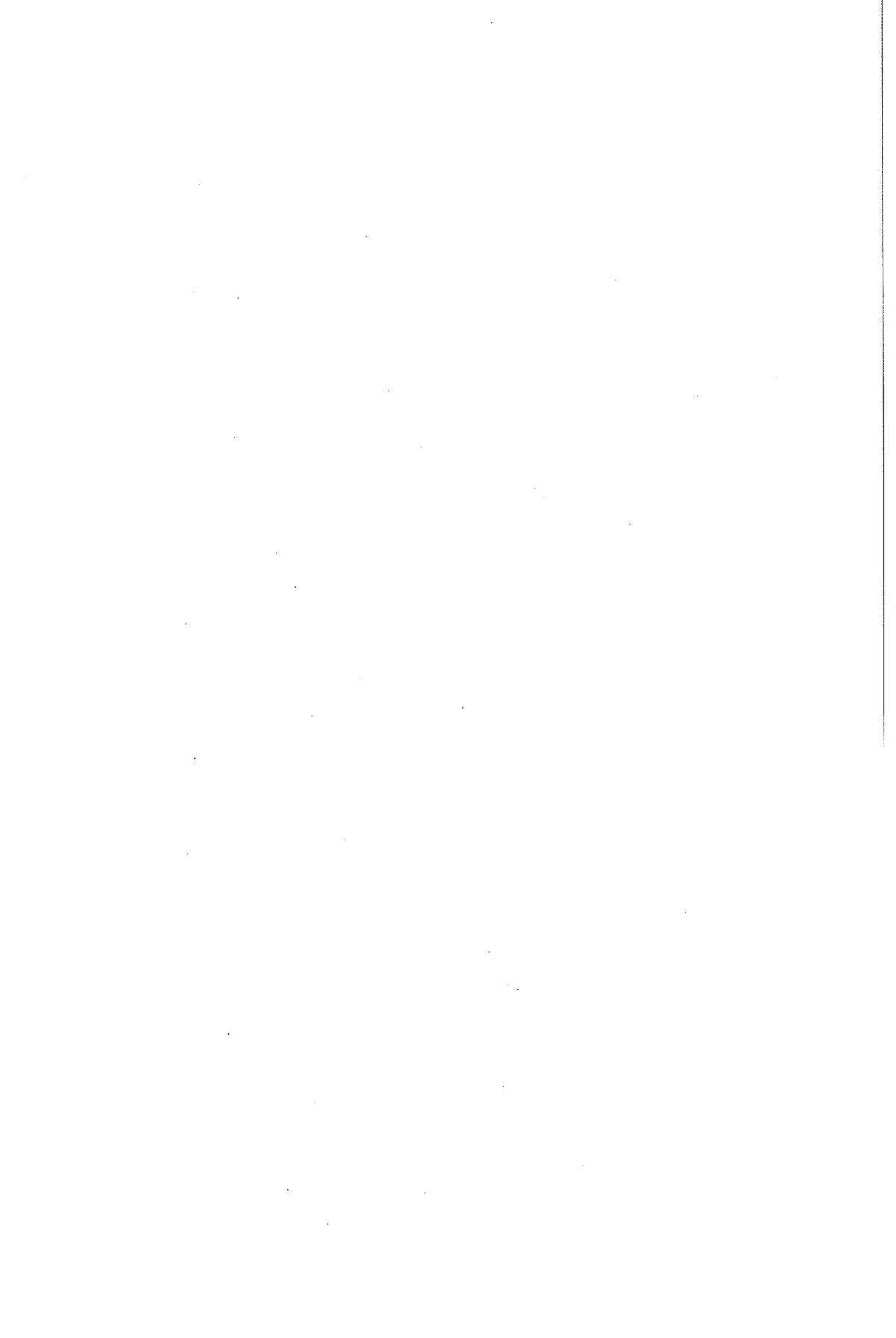
- 1º Se no problema existirem restrições orçamentais, podemos procurar o calendário com melhor VAL de entre os que não ultrapassam um determinado investimento por período.
- 2º porque permite trabalhar, não só com um mas com vários critérios: VAL, taxa interna de rentabilidade, fluxo de caixa (acumulado ou

não) ao longo da realização do projecto, período de retorno, etc. Podemos então elaborar um quadro onde se apresentam as avaliações de cada calendário segundo os vários critérios para que sirva de base a posterior análise multicritério.

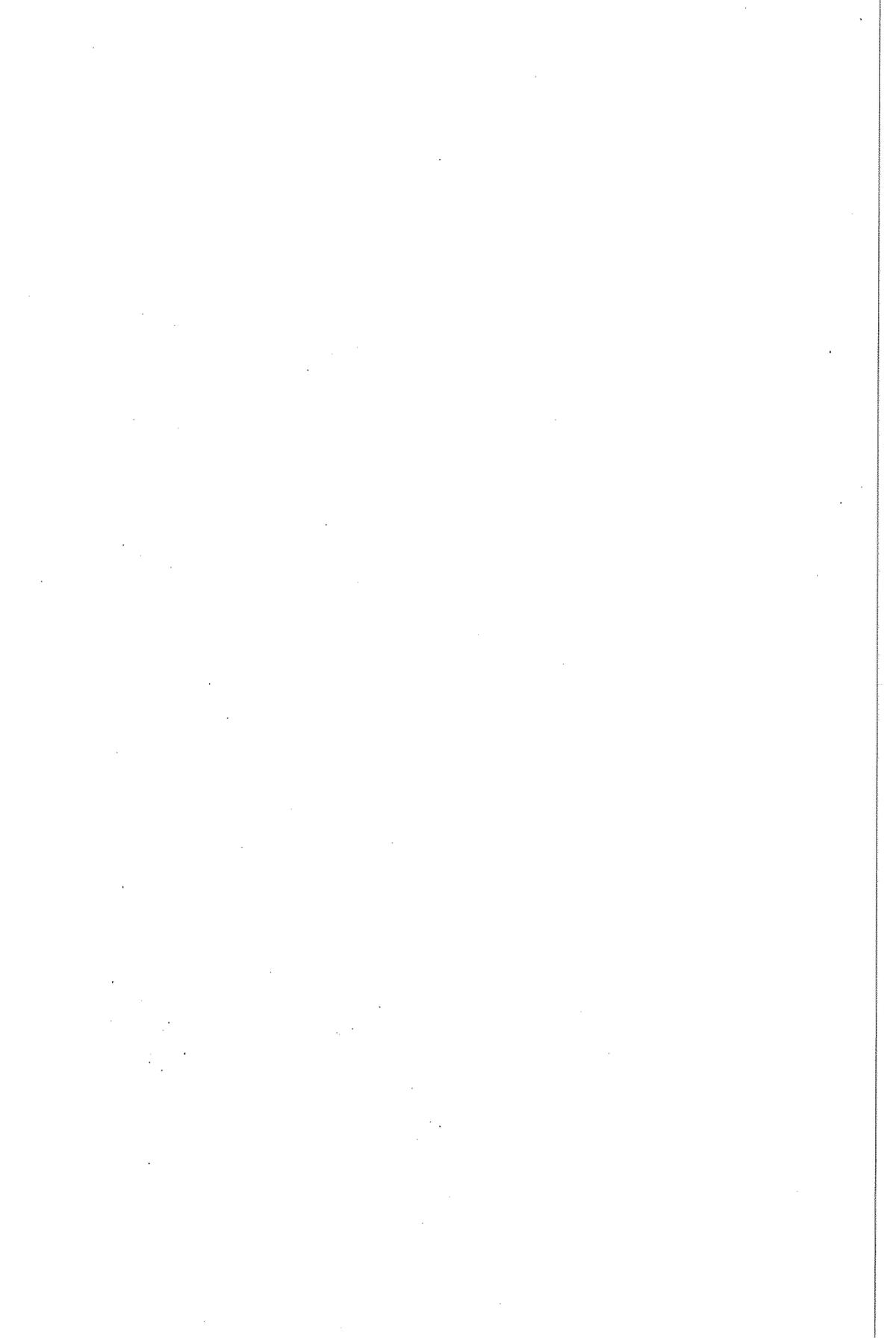
- 3^a porque proporciona ao utilizador uma percepção importante das restrições que condicionam as suas decisões bem como das consequências de cada decisão isoladamente ou integrada na sua estratégia de calendarização, ou seja, o modelo, enquanto o orienta e lhe permite ensaiar diferentes estratégias, está a ensinar o utilizador e a desenvolver nele uma capacidade e uma intuição dos quais tirará benefício quando tomar decisões definitivas.

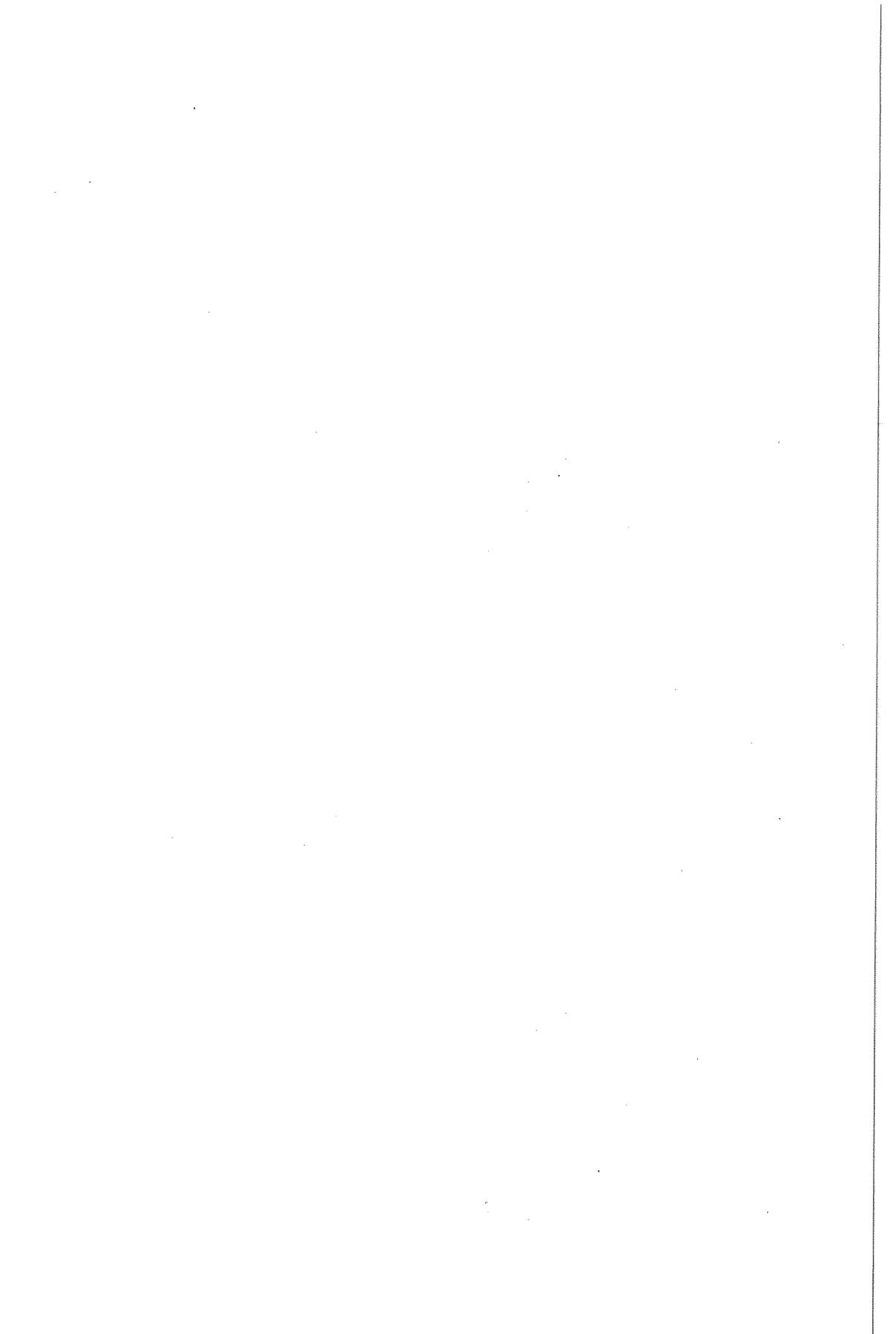
Referências

- [1] BAKER, K., *Introduction to Sequencing and Scheduling*, J.Wiley & Sons, 1974.
- [2] BARNETT, M., *Role of the Merchant Banker in Projects*, Int.J.of Proj.Management, Vol. 5 (1987), pp.197-203.
- [3] DOERSCH, R.H. e PATTERSON, J.H., *Scheduling a Project to Maximize its Net Present Value: a zero-one programming approach*, Manag.Science, Vol. 23 (1977), pp.882-889.
- [4] ELMAGHRABY, S.E., *Activity Networks: Project Planning and Control by Network Models*, J.Wiley & Sons, 1977.
- [5] PATTERSON, J.H., *A Decision Support System for Maximizing Project Net Present Value under Resource Constraints*, comunicação apresentada no EURO IX (Paris, Julho 88) e no 1st Int.Meeting on Project Management and Scheduling (Lisboa, Julho 88).
- [6] RAFAEL, J.F., *Sistema de Apoio à Programação Financeira em Gestão de Projectos*, dissertação para a obtenção do grau de mestre em Investigação Oepracional e Engenharia de Sistemas no IST, Lisboa, Julho 1988.
- [7] TAVARES, L.V., *Multicriteria Scheduling of a Railway Renewal Program*, EJOR, Vol. 25 (1986), pp.395-405.



Fotografia, Montagem
Impressão e Acabamentos
Tip.Nocamil
COIMBRA





ÍNDICE

<i>L. Amado</i> Apoio á Decisão num Sistema de Transportes	3
<i>C.A. Bana e Costa et al</i> MEMSOR - Método Multicritério	19
<i>R. Alves</i> Estimação da distribuição da taxa de produção	29
<i>F.M.F. Lobo Pereira</i> Condições de optimalidade em processos descontínuos	55
<i>D.M. Cardoso</i> Dualidade em Programação Linear Fraccionária	65
<i>V. Sousa et al</i> Programação Quadrática na gestão de mini-hídricas	85
<i>J.F. Rafael</i> Sistema de Apoio á Gestão de Projectos	97



Associação Portuguesa para o Desenvolvimento
da Investigação Operacional

CÉSUR - Instituto Superior Técnico - Avenida Rovisco Pais
1000 Lisboa - Telef. 80 74 55