

INVESTIGAÇÃO OPERACIONAL

Dezembro 1989

Número 2

Volume 9

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 9 - nº 2 - Dezembro 1989

Publicação semestral

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

Comissão Editorial

| | | | |
|---------------------|----------------------------|-------------------------|----------------------------|
| Mordecai Avriel | (Israel) | Nelson Maculan | (UFRJ - Brasil) |
| Paulo Bárcia | (FE - Univ. Nova Lisboa) | A. Rinnoy Kan | (EU - Holanda) |
| João A. Branco | (IST - Univ. Técn. Lisboa) | A. Simões Monteiro | (NORMA) |
| João Clímaco | (INESC/Univ. Coimbra) | 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) José |
| M. Viegas | (IST - Univ. Técn. Lisboa) | 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

CONTRIBUTO PARA A DEFINIÇÃO DE ESTRATÉGIAS NO DESENVOLVIMENTO DE SADs

Jorge R. G. Freire de Sousa
Gabinete de Gestão e Engenharia Industrial
DEMec – FEUP

José A. Vasconcelos Ferreira
Secção de Planeamento do Território
DEC – FEUP

Resumo

Na última década tem-se assistido a um interesse crescente pelos Sistemas de Apoio à Decisão (SADs). Razões para este facto podem ser encontradas a diferentes níveis. Ao nível das aplicações, esse interesse manifesta-se na análise dos benefícios obtidos pela implementação de SADs específicos. Ao nível metodológico, a sua importância justifica-se pelo contributo que os SADs vieram trazer à discussão de novas formas de abordar problemas de decisão.

Nesta comunicação salienta-se a importância da adequação da atitude do analista ao contexto em que decorre o desenvolvimento do SAD, para o sucesso de um tal sistema.

Factores tão diversos como o hardware e o software disponíveis na organização, a sua gestão e estrutura, o estilo cognitivo e a experiência dos utilizadores, o tipo de problemas a resolver e o nível de intervenção pretendido devem ser explicitamente considerados no desenvolvimento de um SAD. Em resposta a estes factores o analista deve dispor de flexibilidade suficiente para definir uma estratégia apropriada que tenha por base decisões sobre, por exemplo, o nível de participação dos utilizadores na concepção do SAD, o grau de diálogo e compreensão da interface a desenvolver, a maior ou menor integração dos diversos componentes do sistema, a separabilidade dos problemas envolvidos e o investimento necessário à execução do projecto.

Ilustra-se o que foi afirmado com a descrição da experiência que constituiu o processo de concepção, desenvolvimento e implementação de um SAD para o planeamento e controle de um sistema de transportes colectivos urbanos.

1. A IO E OS SADs

Hoje, a uma distância de cerca de duas décadas sobre os primeiros sintomas de crise da Investigação Operacional - IO (Miser (1963), Ackoff (1979a, 1979b) e Guimarães (1986)), está-se numa posição consideravelmente melhor para perceber o que falhou e para procurar reflectir e sintetizar a experiência entretanto adquirida.

Durante largos anos a IO abordou problemas susceptíveis de fácil modelação, com vista à obtenção de soluções óptimas. Alguns desses problemas, por serem mais comuns na prática, mais interessantes do ponto de vista da sua resolução matemática ou simplesmente por terem sido os primeiros a ser identificados, foram objecto de uma atenção especial que os transformou em protótipos perfeitamente definidos e para os quais foram propostos variados métodos de resolução, exactos ou aproximados. O esforço colocado na derivação de novos métodos ou no aperfeiçoamento dos existentes, por forma a obter-se uma maior eficiência, tem constituído a preocupação e a ocupação fundamentais de muitos técnicos e investigadores ligados à IO.

Uma ênfase tão pronunciada no uso de modelos conhecidos e tipificados, com o objectivo de para eles se obter soluções óptimas, conduziu à sua aplicação abusiva em problemas práticos relativamente aos quais eles constituíam uma representação inadequada. Os sistemas reais assumem-se, na maioria dos casos, como entidades complexas no seio das quais se colocam problemas múltiplos e interligados que não permitem o estabelecimento de fronteiras rigorosas e estanques e cujas soluções são avaliadas segundo critérios de decisão conflituosos, evoluindo no tempo. Deste modo, o impacto resultante da utilização de modelos simplistas, incapazes de descrever cabalmente a realidade, foi diminuto e contribuiu para mergulhar a IO numa onda de autocrítica.

O debate interno que entretanto se gerou pôs em causa a metodologia clássica da IO e alertou a comunidade científica e técnica para a necessidade de repensar o recurso à optimização no âmbito dos processos de decisão (Guimarães (1986)). Tornava-se relativamente claro que o papel destes modelos deveria ser o de instrumentos enquadrados numa abordagem mais global que, respeitando as características fundamentais dos problemas a resolver, fossem capazes de propor soluções implementáveis, robustas e de melhor qualidade.

A parte final da década de 70 assistiu ao aparecimento dos primeiros Sistemas de Apoio à Decisão - SADs (Keen e Scott-Morton (1978), Courbon, Grajew e Tolovi (1979) e Alter (1980)). Tratava-se de sistemas criados para apoiar a tomada de decisões, permitindo ao decisor uma intervenção activa na obtenção de soluções alternativas e fornecendo-lhe os elementos necessários à sua avaliação. O objectivo destes sistemas ultrapassava a simples resolução de problemas técnicos (nível operacional) para se concentrar no tratamento da informação que fundamenta decisões de carácter mais alargado e ambicioso (níveis tático e estratégico). O acolhimento dispensado a estes trabalhos pioneiros encorajou as organizações e os analistas a desenvolver abordagens deste tipo, tendo-se assistido, desde 1980, a uma proliferação de SADs com características e destinos diversos.

O momento actual convida a alguma reflexão. Pela análise do trabalho entretanto desenvolvido, é possível, não só compreender as causas que estiveram na origem de resultados menos positivos obtidos por alguns SADs, mas também tomar consciência dos desafios que se lhes colocam de modo a perspectivar a sua evolução futura.

Este artigo procura ser um contributo para este esforço de reflexão. No entender dos autores, há um conjunto de questões fundamentais que, a não serem consideradas, podem afectar a eficácia dos SADs e limitar o seu desenvolvimento. Muitas destas

questões relacionam-se com a adequação da atitude do analista ao contexto em que decorre o desenvolvimento do sistema.

2. A ESTRATÉGIA DO ANALISTA NO DESENVOLVIMENTO DE SADs

A concretização dos benefícios esperados pelo recurso a um SAD depende, em grande parte, da adequação da atitude do analista às condições ambientais, sendo as características do produto final um reflexo dessa atitude. O analista lida com uma envolvente que conhece à partida ou, mais realisticamente, que vai descobrindo ao longo do projecto. A compreensão e consideração pelo analista das condições encontradas, desde que não signifiquem uma atitude conformista, poderão contribuir positivamente e de modo significativo para que sejam alcançados os objectivos estabelecidos. Surge assim a noção de estratégia, entendida como um conjunto de decisões sobre os graus de liberdade de que o analista dispõe na concepção, desenvolvimento e implementação de um SAD, de acordo com o contexto em que o trabalho decorre. Interessa definir uma estratégia que suporte a posição do analista face aos demais intervenientes, favorecendo o aparecimento das virtualidades do sistema.

Apresentam-se, em seguida, os factores que se considera serem mais relevantes na caracterização do ambiente, tentando articulá-los com as decisões a tomar pelo analista. Cada uma destas decisões deverá estar intimamente relacionada com as restantes, mas também com as reacções já conhecidas e com aquelas que se prevê virem a acontecer. A complexidade e o dinamismo de um tal processo dificultam uma exposição sistemática e ordenada das questões que são objecto deste trabalho. Por razões de clareza de exposição, na análise que seguidamente se apresenta, tais questões foram classificadas em quatro grandes grupos (contacto inicial, recursos disponíveis, tipos de problemas e actuação ao nível da organização), tentando-se evitar, desta forma, uma separação demasiado artificial dos factores ambientais considerados.

2.1. Contacto Inicial

A primeira questão que se coloca é a de saber como surge a decisão de construir um SAD. Esta decisão formaliza-se por um acordo entre o responsável da empresa e o analista, mas o que importa de facto apurar é de quem parte a iniciativa e quem estabelece o contacto. É claro que no caso do analista ser membro da empresa, o que em princípio só acontece em empresas de uma certa dimensão com departamentos de investigação e desenvolvimento próprios, a importância desta questão se atenua. Isto não significa, no entanto, que deva ser esquecida, fundamentalmente se o departamento que concebe o SAD e aquele a que ele se destina gozam de uma grande autonomia.

A situação que parece ser mais comum é aquela em que a empresa manifesta vontade de resolver um determinado problema contactando para esse fim alguém que pensa possuir o perfil, a experiência e os conhecimentos adequados ao trabalho em causa. Neste caso, podem aparecer duas situações limite. Na primeira, a empresa julga saber perfeitamente o que quer, buscando apenas quem seja capaz de o concretizar. Quando assim acontece, o analista começará naturalmente por limitar a sua intervenção ao que lhe é solicitado, clarificando eventualmente um ou outro aspecto que lhe pareça ambíguo ou obscuro. Com as condições bem definidas, pode realizar o seu trabalho sem uma necessidade especial de contactos frequentes com o cliente. Qualquer aditamento relativamente ao que lhe foi solicitado só deverá ser considerado após a aprovação do produto base. O mesmo não se passa, no entanto, quando a empresa não define claramente o que deseja, situação que é a mais frequente. Isso pode suceder por uma manifesta incapacidade de compreensão do problema ou por dificuldades no modo

de o transmitir. Outra hipótese é aquela em que a empresa deseja ver resolvido um problema muito concreto, deixando a porta aberta à introdução de melhoramentos a outros níveis. Em qualquer destes casos, o analista deve procurar definir o problema em colaboração com o cliente, esclarecendo as questões mais problemáticas, aconselhando-o nas opções a tomar e inquirindo sobre os recursos disponíveis, nomeadamente sobre a informação existente e as formas de lhe aceder. Com base nos elementos recolhidos, deve ser elaborada uma proposta que estabeleça a finalidade do trabalho, o modo como se desenvolverá (com particular destaque para os prazos) e as condições que cada uma das partes se obriga a cumprir. A aprovação do plano de trabalho não deve porém evitar que se proceda a contactos frequentes a fim de avaliar as etapas já cumpridas e, se for caso disso, reajustar as que se seguem.

Quando é o analista que contacta um potencial cliente a situação é diferente. Se lhe é apresentado um produto acabado, o cliente pode proceder a testes e julgar objectivamente o que lhe é proposto, responsabilizando-se pela decisão que vier a tomar. Ainda assim há que ter em atenção as condições que devem ser reunidas para uma utilização eficaz do sistema. Se o analista se põe à disposição do cliente para construir um sistema à medida das necessidades deste, precisa de tomar maiores precauções. Deve possuir uma experiência razoável no desenvolvimento de SADs, estar seguro das suas capacidades e possuir um conhecimento adequado do tipo de problemas em causa, caso contrário deverá suportar os custos iniciais de desenvolvimento do projecto. Isso representará um investimento que poderá rentabilizar posteriormente, libertando-o de pressões indesejáveis e prejudiciais. De resto, tudo o que foi dito para o caso em que o contacto parte do cliente aplica-se também aqui, só que os cuidados devem ser redobrados. Quanto mais espaçadas forem as avaliações do sistema, maior é o risco em que o analista incorre de ver rejeitado o seu trabalho e goradas as expectativas criadas pelo utilizador (Andersen et al. (1987)).

Não sendo crítica, a forma como nasce o sistema marca a atitude do cliente durante todo o processo. Isso nota-se particularmente quando surgem dificuldades, situação na qual, tipicamente, o cliente "descarrega" as culpas sobre o analista. Uma atitude cuidadosa por parte do analista durante os primeiros contactos, que conduzem ao estabelecimento de compromissos, reforça significativamente a sua posição e permite-lhe trabalhar em condições mais vantajosas. De início "tudo" pode e deve ser discutido e questionado, até se a construção do SAD constitui a solução apropriada para o problema em causa.

2.2. Recursos Disponíveis

Uma questão importante para a especificação de um sistema deste tipo, seu desenvolvimento e posterior implementação, refere-se aos meios que à partida são colocados à disposição do projecto ou àqueles que poderão vir a ser mobilizados durante a sua realização. No caso específico dos SADs esses meios são, fundamentalmente, a informação a que se tem acesso, o equipamento informático existente, o tipo de utilizadores do sistema, o prazo de execução dos trabalhos e o investimento disponível para os mesmos.

Com base na informação disponível, o sistema deverá ser concebido de modo a tratar e utilizar aquela que é relevante para o fim a que se destina. Por vezes é necessário recorrer a dados de que não existe até ao momento registo, sendo aconselhável definir com rigor a informação que se pretende, o modo como poderá ser obtida e a data em que deverá estar acessível. Mesmo quando os dados existem, é por vezes necessário dar-lhes outra forma ou transferi-los fisicamente. Estas operações envolvem

normalmente pessoal e equipamento, podendo ser demoradas e dispendiosas. Importa estabelecer quem por elas fica responsável.

No que se refere ao equipamento informático, a situação pode apresentar-se com alguma diversidade. Nalguns casos já existe hardware e software disponíveis na organização, havendo que decidir sobre uma eventual ampliação, enquanto noutros é necessário partir do zero. A evolução da eficiência e do preço dos sistemas informáticos aconselha a que os meios não excedam as necessidades previstas. O importante é que as opções tomadas, por razões de compatibilidade, não dificultem a expansão futura do sistema. É preferível que o cliente, após a obtenção de um primeiro conjunto de resultados satisfatórios, sinta as insuficiências do equipamento que possui e a necessidade de o melhorar ou ampliar, do que verifique a falta de aproveitamento do investimento efectuado, agravada pela desactualização do equipamento adquirido. De qualquer modo, a aquisição de equipamento informático deverá ser criteriosa, tendo por base o material existente, as disponibilidades financeiras e o tipo de uso a que vai estar sujeito.

O pessoal constitui o recurso mais delicado de gerir dentro de uma organização, com repercussões que ultrapassam os aspectos meramente financeiros. Acções de despedimento ou de preparação de certos funcionários para determinadas tarefas revelam-se, por vezes, obstáculos intransponíveis. Destas dificuldades tem de ter consciência o analista, que não pode pedir ao cliente aquilo que este não lhe pode dar. O produto a criar deve ter em conta as características cognitivas dos potenciais utilizadores, permitindo uma fácil aprendizagem da sua operação. A comunicação entre o utilizador e o sistema é feita através da interface concebida pelo analista, pelo que esta deverá ter em conta os conhecimentos informáticos e lógicos do pessoal envolvido. O nível de intervenção permitido ao utilizador reflectirá, entre outros aspectos, a importância de que se reveste a experiência desse utilizador na qualidade das soluções a obter. Apesar deste esforço de adequação do sistema ao utilizador ser fundamental, existem situações em que o sucesso do projecto obriga ao recrutamento de novos elementos e/ou à formação específica dos existentes. Nesses casos, interessará ao analista conseguir a garantia de que tal irá de facto acontecer. Não se deve também esquecer que o envolvimento dos utilizadores na concepção e desenvolvimento do sistema os familiariza com o produto, preparando a sua aceitação e utilização. De notar que a atitude dos utilizadores é difícil de julgar, podendo contribuir decisivamente para o êxito ou fracasso do sistema.

A importância da consideração dos prazos deriva, por um lado, do facto de estarem intimamente ligados à avaliação do projecto e, por outro, de fornecer ao analista uma indicação importante sobre a profundidade que aquele deve atingir. O não cumprimento dos prazos previamente estabelecidos lança o descrédito sobre o produto final (Andersen et al. (1987)), mesmo que este se revele superior ao que foi estipulado de início. Daí a necessidade de definir as metas e os respectivos prazos a partir de um conhecimento completo das actividades e dos riscos envolvidos. É indispensável que o analista se aperceba do grau de importância que a obtenção de resultados no curto prazo tem para a empresa. Vale a pena de facto realizar o trabalho por estádios, correspondendo os primeiros a tarefas menos complexas cujo cumprimento dentro dos prazos aumentará o crédito do analista para as etapas seguintes. A ambição de mostrar demasiado serviço em pouco tempo, quando anunciada à partida, representa uma aposta de elevado risco. O analista só será obrigado a cumprir os prazos acordados.

Um último aspecto, com enormes implicações, refere-se aos recursos financeiros envolvidos no projecto. Uma parte significativa das lacunas ao nível dos outros

recursos pode efectivamente ser colmatada com o aumento das verbas disponíveis. No seu todo, a construção de um SAD representa um investimento que, como qualquer outro, o empresário pretende rentabilizar. Mesmo aqueles que são mais sensíveis à necessidade de semear para colher, com a inevitável perda de algumas sementes, não resistem indefinidamente a projectos fracassados, isto é, a projectos cujos custos superam os benefícios. Cabe ao analista avaliar os benefícios que resultarão da introdução do novo sistema a fim de ter uma ideia dos custos que o cliente estará disposto a suportar. Este pode não fazer as contas a priori mas fá-las-á certamente depois. Um primeiro projecto bem sucedido, com uma relação custos-benefícios vantajosa, pode ajudar os clientes mais desconfiados a decidirem-se por trabalhos de maior envergadura. De notar, contudo, que certos aspectos associados à construção de um SAD não podem ser medidos em unidades monetárias nem tão pouco são sentidos de maneira semelhante por todos os intervenientes (é o caso, por exemplo, da imagem da empresa ou da valorização profissional dos utilizadores).

O grau de especificidade que alguns problemas apresentam pode colocar o analista numa situação difícil. Isto acontece, por exemplo, quando as exigências, em termos dos meios financeiros necessários à execução do projecto, são demasiado elevadas para serem suportadas por um só cliente. Neste caso, o analista pode construir um produto base que, pela sua flexibilidade e modularidade, quando sujeito a algumas adaptações, possa interessar a várias empresas. Estas poderão eventualmente entrar em acordo para custear conjuntamente o desenvolvimento do sistema. Caso isso não seja possível, resta ao analista avaliar se o seu interesse na realização do trabalho compensa o sacrifício financeiro que lhe é exigido.

2.3. Tipos de Problemas

Em geral, quanto mais complexo for o problema, mais inovadora for a sua abordagem e maior for o número de intervenientes no desenvolvimento do SAD, maior será a probabilidade de ocorrerem divergências na interpretação daquilo que o SAD deve ser.

Nos casos mais simples em que os problemas são "bem estruturados", pode acontecer que o analista, por definição treinado na observação e análise de factos, se julgue numa posição privilegiada para a formulação do problema, minimizando a importância das diversas interpretações da realidade em estudo pelos vários membros da organização. O método que conduz à definição do problema assemelha-se ao que tradicionalmente se identifica como o método científico. Esta definição é encarada como um passo preliminar à concepção do SAD e não como uma actividade crítica dentro do próprio processo de desenvolvimento do sistema. Nesta situação, a ênfase é posta no que deve ser feito sem se prestar particular atenção a quem o vai fazer. Há uma tentativa de eliminar os juízos de valor e a validade do método é assegurada por uma constante referência à realidade.

Os problemas reais são normalmente complexos e a obtenção de soluções óptimas não é frequentemente possível. Mesmo os mais simples são em geral mal estruturados, o que faz com que modelos simplistas se afastem substancialmente dos problemas reais. Por outro lado, os subproblemas que resultam da decomposição de um problema global não são separáveis, os processos de decisão evoluem no tempo e o conhecimento dos utilizadores é difícil de ser contemplado nos modelos. Estas são algumas das razões que levam a considerar potencialmente mais promissora a atitude de encarar o problema como uma representação que, embora subjectiva, é útil na busca de uma realidade mais satisfatória.

O objectivo da formulação, nesta perspectiva, será o de construir uma representação comum que seja tão compatível quanto possível com os vários pontos de vista das entidades envolvidas, tendo presente que um consenso total raramente é possível quando estão em jogo problemas complexos. Isto implica o desenvolvimento de uma linguagem comum aos vários participantes, bem como a existência de mecanismos que facilitem o debate e a negociação. Quanto maior for a representação comum, maior será a possibilidade de encontrar soluções aceitáveis.

Repare-se que, neste caso, o papel e as características requeridas ao analista são consideravelmente diferentes das referidas previamente. Os agentes de decisão e os utilizadores estão em melhores condições do que o analista para formular o problema e este deve apoiá-los nessa tarefa, facilitando a interacção entre eles e procurando meios de encorajar a expressão dos vários pontos de vista. Deve o analista sugerir métodos e instrumentos de medida de qualidade das soluções apontadas e, ainda, evitar impor a sua própria definição do problema aos membros da organização.

Entre as características desta metodologia (Landry, Pascot e Briolat (1985)), apresentam-se as seguintes:

- papel central atribuído aos elementos da organização, tendo o analista uma função predominantemente de apoio;
- necessidade de especificar não só as etapas a cumprir, mas também quem deve ser responsável por elas;
- validade do método assegurada por mecanismos que facilitem a comunicação e o debate;
- aceitação de juízos de valor como parte integrante e normal do processo;
- ênfase na fase de definição do problema.

As dificuldades de formulação são facilmente ultrapassadas quando é uma única pessoa a definir e a resolver os seus próprios problemas, isto é, quando existe um só agente de decisão. O mesmo acontece quando há modelos de referência para o problema em causa. Mas pode ocorrer uma outra situação que, não provocando dificuldades na fase de definição, as transfere para a fase de implementação. Trata-se do caso em que o agente de decisão impõe a sua definição aos restantes participantes, em particular aos futuros utilizadores do sistema, que não partilham da sua interpretação dos factos. É uma situação que se deve evitar desde início, sob pena de se detectarem as dificuldades daí resultantes numa fase já avançada do projecto.

Para além da abordagem referida, existem outras que podem ser eficazes. Por exemplo, uma delas propõe a construção de um protótipo que permita, pelo menos parcialmente, a resolução das contradições detectadas na interpretação do problema (Keen (1981)). Esta abordagem sugere também uma reorientação metodológica, aumentando a preocupação com o conhecimento do funcionamento da organização e com a evolução das suas estruturas, bem como com a redefinição do papel do analista. Este passa a ser visto mais como um condutor dos processos de concepção, projecto e implementação do SAD (Courbon (1981)).

A um nível mais operacional, a investigação orientada para os geradores de SADs tem permitido a aplicação de diferentes abordagens. Estas requerem várias iterações e são acompanhadas pela necessidade de "ferramentas" flexíveis e integradas para apoiar os processos de decisão, em que os principais componentes são o utilizador, os dados e os modelos (Sprague (1987)). Mais recentemente, Angehrn (1989) sintetiza da seguinte forma aquelas que, na sua opinião, constituem as três principais tendências actuais na área dos SADs:

- (1) O desenvolvimento de sistemas centrados no utilizador.
- (2) O fornecimento de um apoio activo através da integração "suave" de técnicas analíticas (as clássicas, da IO, bem como outras, mais recentes, desenvolvidas no âmbito da Inteligência Artificial e dos Sistemas Periciais).
- (3) A exploração de novas ideias e técnicas, com origem na investigação na área da Inteligência Artificial, para construir sistemas reactivos, simbióticos, que sejam capazes de adaptar dinamicamente as suas contribuições às necessidades do utilizador.

Finalmente, o analista deverá ter sempre presente que, se é importante que o SAD forneça soluções automáticas para o problema, não se deve retirar ao utilizador a possibilidade de definir, ele próprio, soluções parciais ou completas, de ser flexível no que respeita ao processo de decisão (isto é, ser capaz de adaptar soluções a pequenas modificações nos dados do problema), de verificar a admissibilidade das soluções e de avaliar as medidas de eficiência.

Mais ainda, o sistema deve fornecer ao utilizador explicações sobre os procedimentos adoptados, bem como sugestões que lhe dêem a última palavra na definição do problema e na técnica a usar para a sua resolução (o que pode ser facilitado por uma interface amigável e pelo uso das capacidades gráficas do computador).

É sabido que, de qualquer projecto deste tipo que seja bem sucedido, se esperam mudanças técnicas e de procedimento. Uma mudança técnica pode resultar, por exemplo, de um novo sistema informático. As mudanças de procedimento traduzir-se-ão, por exemplo, em novas posições profissionais a ocupar por cada participante na empresa ou novos elementos que sejam necessários para operar com o novo sistema.

Na especificação de um projecto, há frequentemente tendência para sobrevalorizar os aspectos técnicos e ignorar os comportamentais. O analista deve assegurar que a tecnologia introduzida é a resposta certa às necessidades da empresa e que esta tem a gente certa, motivada e com os conhecimentos adequados, bem como sistemas organizacionais compatíveis com o uso dessa tecnologia.

2.4. Actuação ao Nível da Organização

Problemas de decisão surgem, numa organização, a vários níveis, embora se reconheça que os mais complexos se colocam habitualmente aos elementos que ocupam o topo da estrutura hierárquica. Desde o momento em que se admite o interesse no recurso a um SAD até à altura em que este entra em funcionamento, o processo cumpre dentro da empresa um percurso longo que engloba, normalmente, as seguintes fases:

- (i) reconhecimento da necessidade de um SAD;
- (ii) decisão sobre o seu desenvolvimento;
- (iii) estabelecimento de contacto com o analista;
- (iv) colaboração na concepção, desenvolvimento e implementação do sistema;
- (v) avaliação do trabalho realizado;
- (vi) utilização plena do SAD.

Durante este percurso, o SAD convive com diferentes pessoas que vão desde administradores até simples executantes, passando pelos responsáveis pelos vários departamentos e secções. Muitas das vezes acontece que o responsável pela decisão de construir o SAD não acompanha o seu desenvolvimento nem integra o grupo de potenciais utilizadores do sistema. Assim, em diferentes fases do processo, o analista tem de contactar indivíduos com diferentes percepções do problema, que manifestam posições divergentes sobre o interesse do projecto e que, frequentemente, não

estabelecem uma plataforma de entendimento entre si. Este facto dificulta a tarefa do analista, que se vê empurrado para um jogo em que o reclamar autoridade e competência para julgar o trabalho e o declinar responsabilidades são atitudes tomadas consoante os interesses de momento. As dificuldades crescem à medida que os utilizadores do sistema descem na hierarquia da empresa, afastando-se portanto de quem tem autoridade para decidir quanto a problemas orçamentais e pronunciar a última palavra sobre as opções a tomar.

Quando o sistema se destina a ser utilizado ao nível operacional, o papel das chefias intermédias, responsáveis imediatos pelo sector, é muito importante, uma vez que dentro da empresa essas chefias fazem a ponte entre a administração (poderosa, distante e sempre muito ocupada) e os utilizadores (conhecedores dos problemas mas com dificuldades em exprimir as suas preocupações). São interlocutores privilegiados dos analistas que devem recorrer a eles para defender as suas propostas junto da administração e obter a colaboração dos seus subordinados. Para evitar que se gere à volta do projecto um ambiente de certa instabilidade, concretizada em permanentes indefinições e alterações sucessivas, o analista deve procurar que sejam estabelecidos por escrito os objectivos a atingir pelo sistema, definindo quem contribui para a sua especificação e a quem cabe avaliar o trabalho.

Não podendo ouvir toda a gente, o analista deve contactar aqueles que pensa poderem ajudá-lo na sua tarefa de levantar o sistema e também aqueles cuja posição possa obstruir ou facilitar a sua aceitação. No que respeita aos primeiros, deve o analista lembrar-se que a sua auscultação pode contribuir para o enriquecimento do sistema, conduzindo-o aos objectivos desejados. De notar que algumas das metas a perseguir pelo sistema surgem na fase da sua concepção devido a estes contactos. Quanto aos segundos, devem ser alvo de uma atenção especial. A aceitabilidade de um projecto resulta directamente da avaliação que cada um faz das previsíveis mudanças que esse projecto irá provocar (Klein e Hirschheim (1985)). Uma vez que o desenvolvimento e a implementação de um SAD podem alterar a qualidade das condições de trabalho de um indivíduo, afectar a sua carreira, modificar a sua posição perante os colegas ou mesmo conduzir a uma nova relação de forças dentro da organização, é de esperar que cada um dos seus membros procure adivinhar o modo como poderá ser afectado. E como a forma de uma pessoa ver a situação é necessariamente influenciada pelos seus receios (March e Simon (1958)), corre-se o risco de se verem desencadear resistências que favoreçam uma atitude de boicote ao sistema. Para que isso não aconteça, é conveniente envolver os intervenientes na construção do sistema, fazendo-os sentir que as suas opiniões serão tidas em consideração e explicando-lhes a razão de ser de certas opções tomadas. Note-se no entanto que as justificações baseadas numa racionalidade completa e irrefutável podem não agradar àqueles agentes de decisão cuja autoridade repousa fundamentalmente no carisma ou no poder formal.

Tendo em conta a estrutura da empresa e os equilíbrios de poder estabelecidos, cabe ao analista criar as condições para que a introdução do SAD constitua um momento particularmente motivante para os elementos envolvidos, consolidando uma atitude de consenso que potencie uma mudança de cultura na organização baseada no desenvolvimento dos níveis de conhecimento individual. Importa por isso garantir um ambiente de trabalho adequado, em que a evolução das estruturas se faça de uma forma harmoniosa e os impactos humanos e organizacionais sejam permanentemente tomados em consideração.

3. DESCRIÇÃO DE UM CASO

3.1. Contexto

Desde Outubro de 1984, o Gabinete de Gestão e Engenharia Industrial (GEIN) da FEUP tem vindo a colaborar com o Serviço de Transportes Colectivos do Porto (STCP) no desenvolvimento de um Sistema de Apoio à Decisão para Elaboração de Horários das Viaturas e do Pessoal Tripulante (Sousa (1986), Freire de Sousa (1988) e Vasconcelos Ferreira (1988)).

Apesar do processo de elaboração dos horários ser executado por elementos das Secções de Movimento e de Escalamento, partiu da Direcção dos Serviços de Exploração, responsável por aquelas secções, o reconhecimento das insuficiências associadas aos procedimentos manuais então adoptados. Daí resultou a decisão de informatizar, no todo ou em parte, esse processo, tendo em vista os seguintes objectivos básicos:

- (i) aumentar a velocidade de resposta;
- (ii) simplificar a geração de horários, libertando os elementos responsáveis pela sua elaboração das tarefas mais rotineiras e auxiliando-os na obtenção de soluções alternativas para os problemas mais complexos;
- (iii) aumentar a disponibilidade daqueles elementos para a análise e comparação de decisões alternativas, fazendo uso da sua longa experiência;
- (iv) facilitar, melhorar e tornar mais rápida a produção e impressão de documentos para uso interno ou externo à empresa.

Não possuindo o STCP internamente os meios necessários para levar a cabo a informatização desejada, a Direcção dos Serviços de Exploração optou por delegar essa responsabilidade em especialistas exteriores à empresa. Com essa finalidade, foi contactado o GEIN que, manifestando interesse pelo projecto, iniciou uma fase de reuniões com os responsáveis de modo a clarificar o que se pretendia e as condições em que o mesmo seria realizado. Embora essas reuniões tenham sido quase sempre conduzidas, por parte da empresa, pela Direcção dos Serviços de Exploração, esta encontrava-se mandatada para o efeito pelo Conselho de Gerência, a quem competia a aprovação do projecto. Ao mesmo tempo, consultava regularmente elementos das secções envolvidas, que forneciam explicações detalhadas sobre os procedimentos utilizados e avançavam algumas questões que gostariam de ver contempladas no novo sistema.

O orçamento e os prazos disponíveis para a execução do projecto não foram impostos à partida, mas era evidente a vontade do STCP de, não efectuando um investimento elevado (até porque não eram claros os benefícios financeiros resultantes do projecto), dispor do sistema o mais rapidamente possível. A proposta apresentada pelo GEIN, e que foi posteriormente aprovada pela empresa, tentou acautelar estes dois aspectos, embora as características pioneiras do trabalho tornassem a questão dos prazos algo delicada.

De início, o STCP colocava como prioridade a automatização das tarefas de elaboração dos serviços das tripulações, consideradas como difíceis e envolvendo custos elevados, sobretudo em recursos humanos. No entanto, após uma primeira análise, constatou-se que, mais do que possuir uma ferramenta que resolvesse aquele problema específico, se tornava indispensável dispor de um sistema integrado de

informação, fortemente interactivo, que abrangesse um processo mais alargado de tomada de decisões. Esta constatação fundamentou-se essencialmente em duas razões:

- (i) a elaboração dos horários das linhas, das viaturas e das tripulações processa-se num conjunto de fases entre as quais as interacções são frequentes;
- (ii) a reduzida ou nula formação informática dos potenciais utilizadores, profundos conhecedores dos problemas, aconselhava a adopção de um sistema interactivo em que todo o tratamento da informação se processasse de uma forma progressiva, construtiva e transparente, simples do ponto de vista dos utilizadores.

Assim, pareceu que se deveria investir fundamentalmente na concepção de um sistema global com estrutura modular, passando a geração automática dos serviços das tripulações a constituir um dos vários módulos do sistema.

Em conformidade, o sistema foi desenvolvido por forma a permitir fazer, em etapas sucessivas, a descrição topológica da rede, a definição dos horários das linhas, a geração dos horários das viaturas e dos serviços das tripulações e, finalmente, a elaboração das escalas do pessoal tripulante.

3.2. Caracterização Geral do Processo de Decisão

O planeamento e controle de um sistema de transportes colectivos urbanos é um processo dinâmico de decisão envolvendo objectivos múltiplos. De facto, quer os custos de operação do sistema de transportes quer a qualidade do serviço prestado ao público devem ser considerados na avaliação de esquemas alternativos. A qualidade do serviço pode ser avaliada de muitas maneiras: em termos do número de pessoas servidas pela rede de transportes, da frequência de viagens a diferentes horas do dia e em diferentes dias da semana ou do tipo e qualidade das viaturas utilizadas no serviço, por exemplo. Algumas destas medidas são de índole subjectiva, sendo por isso dificilmente quantificáveis.

Este problema pode ser formulado como uma minimização dos custos totais de operação, satisfazendo um conjunto de restrições que traduzem a necessidade de se garantir níveis aceitáveis da qualidade dos serviços prestados ou que reflectem limitações de natureza tecnológica e financeira.

Na prática, a complexidade do problema leva a que este seja decomposto em problemas parcelares e abordado por fases. De um modo geral, aqueles problemas não são completamente separáveis e, embora havendo a preocupação de estabelecer as suas fronteiras por forma a minimizar as interacções entre eles, não é possível anulá-las por completo. Assim, as fases que correspondem à abordagem dos diferentes problemas não se apresentam numa sequência rígida, existindo frequentemente "feedbacks" entre elas.

Para cada um dos problemas considerados isoladamente, podem ser encontradas, na bibliografia existente sobre o tema, diferentes formulações e modelos de optimização, com referência aos métodos de resolução mais apropriados a cada um deles. A difícil adaptabilidade de alguns destes modelos aos problemas concretos do STCP, acrescida do tipo de resposta pretendida e da necessidade de satisfação de diálogo com o planeador, tornou a sua utilização não desejável, pelo menos numa primeira fase de desenvolvimento.

Por exemplo, o problema da geração dos serviços das tripulações ("Bus Crew Scheduling Problem"), que pode ser formulado como um problema de cobertura de conjuntos, é resolvido de um modo heurístico por aplicação do método das três fases, baseado num procedimento proposto por Parker e Smith (Freire de Sousa (1988)).

Os dois problemas fundamentais relativos à elaboração das escalas do pessoal tripulante ("Rostering Problem") mereceram um tratamento distinto no que diz respeito à sua resolução (Vasconcelos Ferreira (1988)). O primeiro, respeitante à ordenação dos serviços assegurados em regime rotativo, é formulado como um problema de caixeiro viajante e resolvido através de uma heurística composta. Quanto à distribuição diária dos serviços a realizar pelos tripulantes disponíveis, trata-se de um problema de afectação para o qual se desenvolveu um algoritmo baseado no método húngaro.

3.3. Selecção do Equipamento

O sistema foi implementado num microcomputador. Esta opção teve em conta, entre outras, a vantagem de conferir aos serviços de exploração a autonomia desejável, com um investimento relativamente baixo. Adicionalmente, tendo-se optado por um sistema monoposto, garantiu-se que o ritmo de processamento não sofreria flutuações significativas ao longo de cada sessão de trabalho (tal característica é fundamental em sistemas que se pretendem fortemente interactivos).

Quanto à informação necessária à operação do sistema, a que existia revelou-se suficiente, tendo sido preciso, contudo, digitalizá-la. Num ou noutro caso, aproveitou-se a transferência dos dados para lhes dar uma forma mais adequada ao tratamento informático.

3.4. Caracterização e Envolvimento dos Utilizadores

Procurou-se, numa fase inicial do projecto, identificar as principais características cognitivas dos potenciais utilizadores do sistema. Na sua concepção e desenvolvimento foi explicitamente tomado em consideração que tais utilizadores possuem uma reduzida formação informática e lógica, mas que conhecem profundamente os problemas a resolver.

Por forma a tornar o sistema mais atractivo e mais fácil de operar, foram realizadas numerosas sessões de trabalho com os futuros utilizadores, visando integrar as suas sugestões e, sobretudo, detectar os pontos em que as interações com o sistema se revelavam mais críticas. Através da projecção de transparentes e posteriormente no próprio computador, eram apresentados e discutidos os ecrãs que deveriam constituir a interface com o utilizador. Deste modo, para além de se conseguir o envolvimento dos utilizadores na concepção do sistema (o que se considera uma condição indispensável para o sucesso de um projecto desta natureza), preparou-se a sua aceitação, eliminando algumas desconfianças e receios latentes.

A automatização de procedimentos permite uma redução substancial dos indivíduos envolvidos na elaboração de horários. No entanto, dado que a empresa não pretende efectuar despedimentos, a eventualidade dessa redução nunca provocou mal estar entre o pessoal. Os mais aptos para trabalhar com o novo sistema mantêm-se na sua secção, recebendo formação adequada; os restantes mudam de funções.

Por outro lado, a utilização do sistema contraria a adopção de critérios arbitrários, ao introduzir uma maior racionalidade nas decisões a tomar. O que poderia ser sentido como uma ameaça à liberdade de decisão é contudo compensado pelo aumento de controle resultante de uma melhoria na informação disponível e pela transferência de algumas responsabilidades para o sistema. O balanço geral é, do ponto de vista dos utilizadores, francamente positivo.

3.5. Caracterização Geral do SAD

O sistema foi concebido por forma a explorar o conhecimento dos utilizadores, não tentando substituí-los e assumindo-se antes como um auxiliar do processo de decisão. A sua construção visou obter níveis tão elevados quanto possível de:

- (i) flexibilidade;
- (ii) integração da informação;
- (iii) interactividade.

O desenvolvimento de um sistema, com a amplitude e a complexidade que este possui, constitui um trabalho demorado que não pode ser reiniciado sempre que nele se pretendam introduzir refinamentos ou se verifiquem alterações nas condições de operação. Daqui resulta a necessidade de o dotar de uma estrutura flexível que favoreça o seu crescimento por etapas, preparando-o para futuras expansões. A estrutura adoptada, baseada numa partição do sistema num conjunto de módulos relativamente autónomos e com um mínimo de interdependências entre si, torna significativamente mais simples proceder à manutenção do sistema bem como à integração de novos elementos.

Cada um dos componentes do sistema recorre a um volume de informação que, uma vez processada, volta a ser guardada em ficheiro. Frequentemente, verifica-se que os dados de saída de alguns componentes coincidem com os dados de entrada de outros. A integração de toda esta informação é uma questão crucial, não sendo aconselháveis duplicações nem o registo de informação irrelevante. Com as estruturas de informação criadas, tentou-se reduzir a informação armazenada e torná-la facilmente acessível. A interface desenvolvida dissimula as diferenças entre estas estruturas e os antigos arquivos, permitindo ao utilizador manipular os dados de um modo que lhe é familiar.

Pretendendo-se dotar o sistema de um elevado grau de interactividade, o que se traduz na possibilidade de o utilizador intervir em todas as fases do processo de decisão, o desenvolvimento de uma interface adequada teria que ser uma das preocupações principais do analista. Essa interface é constituída por um vasto conjunto de ecrãs organizados segundo uma estrutura arborescente. Alguns, os menus de selecção, servem para subir ou descer na árvore, localizando o operador na área de trabalho que lhe interessa. Os outros permitem a entrada e saída de informação e o desencadear de processamentos. O objectivo foi o de criar níveis de opção progressivamente mais desagregados, permitindo ao utilizador, a partir de qualquer ecrã, progredir numa dada sequência ou regressar ao menu imediatamente anterior. Pela repetição sistemática deste procedimento, poderá o utilizador aceder a qualquer dos ecrãs disponíveis.

Todas as opções do utilizador, quanto à sequência a prosseguir, foram explicitamente separadas do processo de entrada de dados pela afectação de teclas especiais de função. Quando o sistema sugere uma determinada linha de acção com base num modelo nele incluído, o utilizador poderá sempre contrariar aquela sugestão, se entender que existem critérios ou dados que não são tomados em devida conta no modelo. Numa situação limite, o utilizador poderá inclusivamente construir os horários e definir os serviços tal como o fazia manualmente (ainda que de uma forma mais cómoda), ignorando por completo as sugestões que o sistema está potencialmente habilitado a fornecer. Na prática, esta capacidade revela-se de grande interesse pois possibilita uma aproximação natural e progressiva do utilizador ao sistema. A aprendizagem torna-se assim uma tarefa simples.

Num dos módulos do sistema, desenvolveu-se um conjunto de ecrãs recorrendo ao modo gráfico do microcomputador. Conseguiu-se assim uma representação visual mais expressiva das entidades em jogo e a apresentação simultânea, em cada ecrã, de uma quantidade de informação cuja representação como texto se tornava impossível. A informação numérica, por exemplo, é apresentada através de caracteres de dimensões

reduzidas especialmente desenvolvidos para o efeito. Foi ainda concebido um cursor especial que permite ao utilizador, sobre o próprio ecrã, seleccionar as entidades que pretende manipular. Alguns inconvenientes que o recurso ao modo gráfico possa apresentar, nomeadamente no que respeita ao tempo de produção dos ecrãs e à redução do número de cores disponíveis, são compensados por uma qualidade de representação muito superior.

De um modo geral, os documentos produzidos pelo sistema constituem instrumentos de trabalho quotidiano para os elementos da empresa envolvidos na operação dos serviços. A sua concepção exigiu que se estabelecessem compromissos entre as limitações impostas pela sua produção em computador e a sua legibilidade. Nalguns casos, aproveitou-se a oportunidade para se alterar radicalmente esses documentos, tornando-os mais racionais. No entanto, procurou-se sempre envolver os utilizadores no seu desenho, só procedendo às alterações mais profundas após uma aceitação clara por parte daqueles.

Em resumo, procurou-se, tal como Little (1970) preconiza, que o sistema fosse simples, robusto, fácil de controlar, adaptativo, exaustivo nas questões fundamentais e de fácil comunicação.

Tal como foi concebido, o sistema constitui um instrumento que permite o ensaio expedito de cenários alternativos de exploração da rede, facilitando o estabelecimento de critérios de decisão que considerem o carácter eminentemente multi-objectivo do processo.

As soluções que se obtêm para os problemas da geração dos serviços e do escalamento das tripulações são bastante satisfatórias, representando ganhos significativos no tempo necessário para os resolver.

A versão actual do sistema cobre apenas um subconjunto das tarefas que constituem o processo global de planeamento e controle dos serviços de transportes. No entanto, dada a sua estrutura modular, o sistema poderá facilmente vir a ser expandido e melhorado. O seu aperfeiçoamento é um processo dinâmico que exige uma continuada colaboração com os utilizadores, por forma a serem detectados os aspectos menos bem tratados. No caso do STCP, acredita-se que esta colaboração será fácil de prosseguir, dada a cooperação e o interesse manifestados pelos elementos da empresa directamente envolvidos no projecto.

O facto de se terem alargado os objectivos iniciais do projecto, tornando-o mais ambicioso, teve como contrapartida um aumento substancial do custo e dos prazos de execução. Crê-se, todavia, que a opção tomada, procurando obter um sistema integrado de informação para apoio à decisão, em detrimento de um sistema que processasse exclusivamente os serviços das viaturas e das tripulações, foi correcta. Para o GEIN, as características inovadoras de que se revestia o problema justificaram uma atitude de investimento numa área de grande interesse estratégico, acreditando-se que o produto desenvolvido possa vir a interessar outras empresas do sector. Para o STCP, a espera que suportou foi recompensada pela aquisição de um sistema reconhecidamente eficaz como auxiliar na tomada de decisões, a um preço inferior aos custos de desenvolvimento.

4. CONCLUSÃO

O caso que se acaba de descrever não pretende ser um exemplo da estratégia ideal no desenvolvimento de um SAD. Constituiu, no entanto, uma experiência enriquecedora que inspirou aos autores a reflexão sistematizada neste artigo. O conjunto dos factores ambientais considerados, bem como das correspondentes atitudes por parte

do analista, pode e deve ser substancialmente alargado.e discutido. Espera-se de algum modo ter contribuído para lançar essa discussão.

Contudo, esta questão é apenas uma das que se colocam a quem concebe, desenvolve e implementa SADs. Ultrapassada uma fase em que a sua divulgação e aplicação constituíram moda, para que estes assumam todas as suas potencialidades, é necessário:

- (i) Caracterizar de forma precisa o que se entende por um SAD. A definição de SAD deverá ser suficientemente ampla para não restringir o aparecimento de contribuições com diferentes tónicas nem a evolução do próprio conceito, mas simultaneamente permitir a delimitação de fronteiras que operacionalizem o debate de ideias e a integração de novas experiências.
- (ii) Estabelecer, ainda que em traços largos, uma metodologia para a construção dos SADs, orientando o trabalho do analista sem lhe retirar a liberdade de proceder de acordo com a sua própria sensibilidade e com as características particulares do problema.
- (iii) Assumir que o desenvolvimento de um SAD representa uma oportunidade privilegiada para criar mudança na organização a que se destina, refira-se ela à competência e especialização do pessoal envolvido, ao modo de trabalho e uso da tecnologia ou à própria estrutura organizacional.
- (iv) Procurar a integração nos SADs de novos instrumentos que os tornem mais eficazes ou eficientes. É o caso, por exemplo, das técnicas que permitem dotar os sistemas com a capacidade de aprendizagem e de auto-adaptação.
- (v) Valorizar a adequação da atitude do analista ao contexto em que decorre o desenvolvimento de um SAD. Para tal, importa identificar os factores relevantes que caracterizam o ambiente, permitindo ao analista seleccionar a estratégia mais aconselhável.
- (vi) Tentar provar o impacto positivo dos SADs no funcionamento das organizações, notando embora as dificuldades deste processo.

Apesar da multidisciplinaridade reconhecida aos SADs, a sua génese e a formação de base dos que os desenvolvem e utilizam apontam a comunidade científica associada à IO como uma das que está em melhores condições de responder aos desafios atrás enunciados. Pela qualidade dessas respostas passa em grande parte o futuro da própria IO.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ACKOFF, R.L. (1979a), The Future of Operational Research is Past, Journal of the Operational Research Society, vol. 30, 93-104.
- [2] ACKOFF, R.L. (1979b), Resurrecting the Future of Operational Research, Journal of the Operational Research Society, vol. 30, 189-199.
- [3] ALTER, S. (1980), Decision Support Systems: Current Practice and Continuing Challenges, Addison-Wesley, Reading MA.
- [4] ANDERSEN, E.S. , GRUDE, K.V. , HAUG, T. e TURNER, J.R. (1987), Goal Directed Project Management, eds. Kogan Page e Coopers & Lybrand, Great Britain.
- [5] ANGEHRN, A.A. (1989), Modelling by Example - New Ideas for Interactive Decision Support, VI EURO Summer Institute, Madeira, Portugal.

- [6] COURBON, J.C. (1981), *L' Approche Systémique et la Conception Évolutive*, Informatique et Gestion, 122.
- [7] COURBON, J.C. , GRAJEW, J. e TOLOVI, J.R. (1979), *Conception et Mise en Oeuvre des Systèmes Interactifs d' Aide à la Décision - L' Approche Évolutive*, Informatique et Gestion, 103, 51-59.
- [8] FREIRE DE SOUSA, J. (1988), *Desenvolvimento de um Módulo Interactivo para Geração de Serviços de Tripulações em Transportes Colectivos Urbanos*, Tese de Mestrado, Instituto Superior Técnico - UTL.
- [9] GUIMARÃES, R.C. (1986), *A Optimização no Contexto dos Problemas de Decisão*, 1º Ciclo de Conferências em Análise Numérica e Optimização, Departamento de Matemática, Universidade de Coimbra.
- [10] KEEN, P.G.W. (1981), *Value Analysis: Justifying Decision Support Systems*, MIS Quarterly, 5.
- [11] KEEN, P.G.W. e SCOTT-MORTON, M.S. (1978), *Decision Support Systems: An Organizational Perspective*, Addison-Wesley, Reading MA.
- [12] KLEIN, H.K. e HIRSCHHEIM, R. (1985), *Fundamental Issues of Decision Support Systems: A Consequentialist Perspective*, Decision Support Systems, vol. 1, nº 1, North-Holland.
- [13] LANDRY, M. , PASCOT, D. e BRIOLAT, D. (1985), *Can DSS Evolve Without Changing Our View of the Concept of "Problem"?*, Decision Support Systems, vol. 1, nº1, North-Holland.
- [14] LITTLE, J. (1970), *Models and Managers: The Concept of a Decision Calculus*, Management Science, vol. 16, nº 8.
- [15] MARCH, J. e SIMON, H. (1958), *Organizations*, Wiley, New York.
- [16] MISER, H. (1963), *Operations Research in Perspective*, Operations Research, vol. 11, 669-677.
- [17] SOUSA, J.P. (1986), *Um Sistema de Apoio à Decisão para a Elaboração de Horários de Viaturas e Tripulações em Serviços de Transportes Colectivos Urbanos*, Tese de Mestrado, Instituto Superior Técnico - UTL.
- [18] SPRAGUE, R.H. (1987), *DSS in Context*, Decision Support Systems, vol. 3, nº 3, 197-202, North-Holland.
- [19] VASCONCELOS FERREIRA, J. (1988), *Concepção de um Módulo Interactivo para Geração de Escalas de Pessoal Tripulante em Transportes Colectivos Urbanos*, Tese de Mestrado, Instituto Superior Técnico - UTL.

ALGUMAS REFLEXÕES SOBRE UMA BASE DE MÉTODOS DE PROGRAMAÇÃO LINEAR MULTICRITÉRIO

C. Henggeler Antunes^(1,3)
Maria J. Alves^(2,3)
Ana L. Silva
João N. Clímaco^(2,3)

Resumo

Este artigo pretende apresentar uma síntese dos resultados das investigações em métodos interactivos de programação linear multiobjectivo que temos vindo a desenvolver.

Será apresentada uma implementação computacional de uma base de métodos de programação linear multiobjectivo, onde é dada particular atenção ao diálogo com o agente de decisão, em torno da qual serão feitas algumas reflexões sobre as características desejáveis em métodos interactivos para constituírem um sistema de apoio à decisão eficaz. A base de métodos integra alguns dos mais representativos métodos interactivos, permitindo tirar partido da sua combinação e da transferência de informação recolhida em cada um deles. Trata-se de um trabalho pioneiro nesta área, ao implementar um conjunto de métodos e conceber mecanismos de integração entre eles, permitindo grande flexibilidade de utilização.

Esta base de métodos resulta de um trabalho que vem sendo desenvolvido há alguns anos na área da programação matemática com objectivos múltiplos, incluindo a inovação ao nível metodológico, a realização de implementações computacionais e a aplicação a problemas de planeamento energético. A base de métodos foi apresentada pela primeira vez no "IIASA Workshop on Multiple Criteria Decision Support" [7], tendo a demonstração do programa então efectuada suscitado uma interessante discussão sobre as possibilidades de desenvolvimento futuro.

1 - INTRODUÇÃO

A complexidade dos problemas reais é essencialmente caracterizada pela existência de múltiplos critérios em que o agente de decisão (AD) é confrontado com a necessidade de ponderar os conflitos entre os critérios, com vista a encontrar uma solução de compromisso satisfatória. De uma forma geral, subdividem-se os problemas multicritério em problemas multiobjectivo e multiatributo. O problema de optimização multiobjectivo refere-se aos casos em que as alternativas formam um contínuo, definidas implicitamente por um conjunto de restrições. O problema multiatributo refere-se aos casos em que as alternativas são explicitamente conhecidas, num número finito.

Dada a existência de múltiplos objectivos, a noção de solução óptima (que conduz ao melhor valor admissível para uma única função objectivo) cede lugar à noção de solução não dominada. Uma solução não dominada (também designada por óptima de Pareto, eficiente ou não inferior) caracteriza-se por não existir outra solução admissível que melhore simultaneamente todos os objectivos, isto é, a melhoria num objectivo é alcançada à custa de piorar pelo menos um dos outros.

(1) Deptº de Engª Electrotécnica - Universidade de Coimbra, Largo Marquês de Pombal, 3000 Coimbra

(2) Faculdade de Economia - Universidade de Coimbra, Av. Dias da Silva, 3000 Coimbra

(3) INESC - Núcleo de Coimbra, Rua Antero de Quental, 231 cave, 3000 Coimbra

Enquanto o conceito de não dominância é geralmente usado para pontos no espaço dos objectivos, o conceito de eficiência é usado para a respectiva imagem no espaço das variáveis de decisão. Uma solução de compromisso para o problema multiobjectivo deverá ser não dominada.

Os métodos de programação matemática multiobjectivo destinam-se a calcular soluções não dominadas. Contudo, para a maior parte dos problemas o cálculo exaustivo do conjunto de soluções não dominadas obriga a um elevado esforço computacional e não dá conta da subjectividade humana inerente a qualquer problema multiobjectivo. Surgiu então a necessidade de desenvolver métodos interactivos que, tirando partido da intervenção do AD, reduzam o âmbito da pesquisa e minimizem o esforço computacional. Os métodos interactivos alternam fases de cálculo e fases de decisão, nas quais o AD é chamado a expressar as suas preferências de modo a conduzir, usando diferentes estratégias, o processo de cálculo para a zona da região admissível onde se localizam as soluções que melhor correspondam ao seu sistema de preferências.

Nas últimas duas décadas observou-se um desenvolvimento considerável de métodos interactivos de programação linear multiobjectivo (PLMO) (para uma descrição e análise comparativa, ver por exemplo, Steuer [1], Vanderpooten e Vincke [2] ou Zionts [3]). Estes métodos têm vindo a constituir meios de apoio à tomada de decisões numa enorme gama de problemas em áreas como a engenharia, a gestão, ou o planeamento nos sectores público e privado (ver Weisstrofer et al [4] para uma recente descrição de aplicações).

O estudo de problemas de planeamento energético usando vários métodos interactivos de PLMO permitiu-nos compreender as razões pelas quais não se pode concluir a superioridade de um método sobre os outros em todas as circunstâncias (ver Clímaco e Antunes [5]). De facto, mesmo no estudo de um dado problema, existem métodos que são mais adequados do que outros em diferentes fases do processo de decisão. Por exemplo, o AD pode não estar disposto a comparar pares de soluções antes de ter mais algum conhecimento do problema, fazendo previamente uma pesquisa estratégica mais global. E, raramente o mesmo método permite ao AD realizar estas duas acções para construir o seu sistema de preferências.

A experiência adquirida com o desenvolvimento das implementações computacionais e com as aplicações de métodos interactivos de PLMO, bem como a consideração das suas características conceptuais, fizeram surgir a necessidade de dispôr de uma ferramenta mais flexível capaz de constituir uma valiosa base de experimentação. Esta avaliação conduziu ao desenvolvimento, a partir de implementações isoladas de cada método, de uma base integrada de métodos que permite tirar partido da combinação de diferentes tipos de métodos interactivos de PLMO [6]. Actualmente a base de métodos inclui os métodos STEM, Zionts-Wallenius, Interval Criterion Weights, Pareto Race e TRIMAP, tornando possível a transferência de informação entre cada um deles e usando um "interface" comum com o utilizador.

O sistema de apoio à decisão global inclui a base de métodos (associada a um conjunto de procedimentos auxiliares), uma base de diálogo e um módulo de gestão de dados. O princípio subjacente é apoiar interactivamente o AD a diminuir progressivamente o âmbito da pesquisa. À medida que vai sendo reunido maior conhecimento sobre o problema em cada interacção, o sistema de preferências do AD vai evoluindo, obrigando-o a reflectir sobre as opções manifestadas anteriormente ou mesmo a rever o seu sistema de preferências, nomeadamente se ocorrerem incoerências entre as preferências manifestadas em diferentes interacções [7]. O AD é considerado uma componente activa do sistema de apoio à decisão. A condição de paragem do processo interactivo é a satisfação do AD com o

conhecimento adquirido sobre o problema, e não o teste de convergência de alguma função utilidade (que transforme numa única dimensão todos os aspectos do problema).

A abordagem de integração de diferentes métodos interactivos de PLMO subjacente a TOMMIX tem uma filosofia diferente da "consolidação" proposta por Steuer e Whisman [8]. Esta aproximação pretende criar um método interactivo global de PLMO que englobe a maioria dos métodos mais conhecidos. Na base desta abordagem está o chamado "meta-programa" que inclui os diferentes tipos de métodos interactivos, que passam a ser casos particulares do "algoritmo consolidado", através da concretização de variáveis e da consideração ou não de conjuntos de restrições. Em cada transição de método, em qualquer interacção com o AD, a preocupação fundamental é preservar os níveis de convergência já atingidos.

Por seu lado, a base de métodos de PLMO tem como objectivo a criação de uma ferramenta flexível, respeitando as características conceptuais dos métodos. A preocupação principal é a possibilidade de combinar os métodos garantindo a transferência de informação útil, que possa ser usada nas fases seguintes do processo de decisão, e não a convergência para o óptimo de uma função utilidade implícita do AD. Actualmente, este trabalho está especialmente adaptado para problemas de PLMO com três objectivos (o que permite o uso de meios gráficos particularmente úteis no diálogo com o AD), integrando alguns dos métodos representativos de diferentes estratégias. Por esta razão foi-lhe dado o nome de TOMMIX ("three-objective methods mixed") [9].

A base integrada de métodos não tem a intenção de substituir o AD, mas de o apoiar no estudo do problema e na construção do seu sistema de preferências, fornecendo-lhe argumentos no sentido de reforçar ou enfraquecer as suas convicções que evoluem ao longo do processo interactivo. Desta forma, no processo de decisão, o conceito de "produzir decisões" dá lugar ao conceito de "apoio à decisão". Tal como foi claramente afirmado por French [10], "um bom apoio à decisão deve ajudar o AD a explorar não apenas o problema mas também ele próprio, chamando a sua atenção para possíveis conflitos e incoerências nas suas preferências de modo a que ele possa pensar na respectiva resolução".

Este artigo pretende fazer uma síntese em língua portuguesa do trabalho desenvolvido nesta área, já divulgado sob diferentes perspectivas nos textos referenciadas em [6, 7, 9]. Em [6] é apresentada a base de métodos, salientadas as principais características da aplicação computacional e realizada uma aplicação em planeamento energético. Em [7] são discutidas as vantagens de dispôr de um utensílio informático deste tipo e as características de interacção homem-máquina desejáveis. A implementação computacional, incluindo um manual de utilização do programa, e um exemplo detalhado de utilização de TOMMIX são apresentados em [9].

Na secção 2 do presente artigo são descritos os aspectos gerais da base de métodos. A descrição das principais características do interface homem-máquina é feita na secção 3. Na secção 4 apresenta-se um resumo de cada um dos métodos, descrevendo-se o funcionamento da base de métodos na secção 5. Na secção 6 é apresentado um exemplo ilustrativo. Algumas aplicações são referidas na secção 7, enquanto as conclusões são apresentadas na secção 8.

2 - PRINCIPAIS CARACTERÍSTICAS DE TOMMIX

O sistema de apoio à decisão é constituído por uma base de métodos de PLMO (associada a uma "caixa de ferramentas" auxiliar que inclui um conjunto de procedimentos que podem ser usados em qualquer fase do processo de decisão), uma base de diálogo e um módulo de gestão de dados.

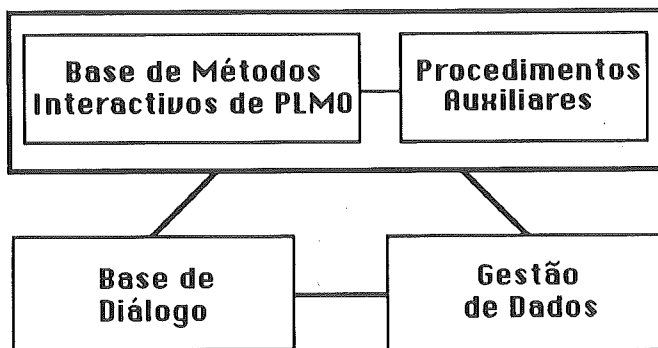


Fig. 1 - Diagrama de blocos de TOMMIX

A base de métodos inclui presentemente cinco métodos interactivos de PLMO: STEM, Zions-Wallenius (ZW), "Interval Criterion Weights" (ICW), "Pareto Race" (PR) e TRIMAP. Estes métodos são representativos de diferentes estratégias: redução da região admissível, redução do espaço dos pesos, contracção do cone dos critérios e pesquisa direccional.

TOMMIX permite o uso isolado de cada método oferecendo também a possibilidade de trocar de método em qualquer interacção com o AD. É, assim, possível tirar melhor partido das potencialidades de cada método em diferentes fases do processo de decisão, bem como avaliar a eficiência da aplicação de cada um dos métodos em problemas de decisão particulares. A base de métodos de PLMO permite total flexibilidade nas transições entre métodos, dado que as limitações adicionais introduzidas usando um dado método não interferem com as características fundamentais dos outros métodos.

Numa fase inicial do processo de pesquisa, revela-se geralmente mais adequado o uso dos métodos que permitam a limitação da pesquisa futura a uma região mais restrita do poliedro admissível (por ex., STEM, ICW, TRIMAP). Numa fase final do processo de decisão privilegiam-se geralmente os métodos que permitam uma pesquisa local (por ex., ZW, PR). Apesar da dificuldade em realizar estudos experimentais nesta área, esta forma de proceder está de acordo com os resultados de experiências divulgados na literatura científica. Por exemplo, Buchanan e Daellenbach [11] concluíram que os participantes nas suas experiências de laboratório faziam inicialmente uma apreciação global a partir de uma visão geral, prosseguindo depois para uma pesquisa local numa vizinhança próxima da solução preferida. De facto, as preferências do AD são geralmente vagas e não explicitamente conhecidas. Torna-se assim necessário levar a cabo uma fase inicial de exploração em que o AD possa conhecer melhor o problema e rever as suas preferências *a priori* (geralmente relativas a problemas semelhantes que se lhe depararam anteriormente). Apenas quando o AD estiver suficientemente seguro quanto ao seu sistema de preferências, podendo responder a eventuais questões formuladas pelos métodos (por ex., comparação entre pares de soluções), é então útil prosseguir uma fase de procura onde o AD tenta encontrar uma solução de compromisso ou um conjunto de alternativas para posterior análise. Esta estrutura

acomoda a componente "errática" inerente a qualquer processo humano de decisão, tentando tirar partido da criatividade deste processo, por forma a proporcionar ao AD o aumento do seu conhecimento sobre o problema, que será utilizado em fases seguintes do processo de decisão.

A utilização adequada deste conhecimento permite o aparecimento de novas intuições sobre o problema que podem ser usadas como ensaios para expressar novas preferências e/ou especificar novas direcções de pesquisa. O princípio subjacente a este sistema é ajudar o AD a, progressiva e interactivamente, diminuir o âmbito da pesquisa, graças à aprendizagem do problema que o sistema lhe proporciona. O AD é assim considerado como uma componente activa do processo de decisão, estando o seu sistema de preferências em construção progressiva em cada interacção.

As acções do AD são sempre revogáveis permitindo-lhe fazer tentativas sem consequências posteriores, se a direcção de pesquisa for considerada sem interesse. Consegue-se assim acomodar julgamentos subjectivos do AD que poderão ou não ser confirmados em experiências posteriores.

A base de métodos está associado uma "caixa de ferramentas", constituída por um conjunto adicional de procedimentos que permitem calcular todos os pontos extremos não dominados, calcular todos os pontos extremos não dominados adjacentes a uma dada solução extrema e "filtrar" um conjunto de soluções que serão posteriormente apresentadas ao AD. Por exemplo, o cálculo de todos os pontos extremos não dominados pode ser interessante após uma suficiente redução da região de interesse do AD delimitada durante o processo interactivo.

O sucesso e aceitação deste tipo de sistemas de apoio à decisão depende fortemente do tipo de interacção com o utilizador. O objectivo é oferecer ao AD uma comunicação fácil e flexível, por forma a captar melhor as suas preferências e tirar partido das suas capacidades de processamento de informação e de decisão. A base de diálogo controla todos os aspectos de interacção homem-máquina, destinando-se a oferecer ao AD um meio operacional fácil de aprender e de uso intuitivo. O espaço dos pesos tem um papel chave no "interface" gráfico oferecido por TOMMIX. É usado como um valioso meio para guardar a informação obtida sob várias formas e em diferentes métodos. Por exemplo, as limitações adicionais introduzidas nos valores das funções objectivo ou as contracções do cone dos critérios são transpostas para o espaço dos pesos, permitindo assim que a informação transferida entre métodos seja apresentada graficamente. O AD deve ser auxiliado na interpretação da informação no espaço dos pesos por um analista com conhecimentos de PLMO.

O módulo de gestão de dados é responsável pela entrada de dados, edição de problemas e saída de resultados (para impressora ou disco).

Todo o programa foi desenvolvido em Pascal no microcomputador Apple Macintosh II, recorrendo extensivamente às rotinas da "toolbox" em ROM. O programa foi construído de forma modular podendo facilmente incluir novos métodos, procedimentos auxiliares ou formas de apresentação de resultados.

3 - INTERACÇÃO HOMEM-MAQUINA

A interacção "homem-máquina" é uma componente fundamental de um sistema de apoio à decisão, tendo como objectivos contribuir para a criação do sistema de preferências do AD e a ampliação das suas capacidades de processamento de informação e de decisão.

No desenvolvimento computacional de sistemas de apoio à decisão, deve ser feito um equilíbrio entre as potencialidades que são oferecidas ao utilizador e a facilidade de aprendizagem e uso. Para isso, deve ter-se em atenção os limites das

capacidades humanas (sobretudo a memória e paciência), dando relevo às suas principais potencialidades (nomeadamente a inspecção visual).

De forma a tirar partido das capacidades de processamento de informação do AD, a base de diálogo faz uso extensivo de gráficos e oferece um ambiente flexível e amigável servindo-se de janelas múltiplas, menus, utilização do "rato" e caixas de diálogo (tendo em atenção a estrutura semântica das sequências de diálogo possíveis), mantendo sempre o utilizador no controlo do processo de pesquisa de soluções. A informação numérica encontra-se acessível sempre que requerida pelo utilizador.

O utilizador controla o comportamento do programa através de operações simples, como premir o "rato" em determinadas zonas do écran ou premir uma tecla. O "interface" com o utilizador, desenvolvido na implementação computacional de TOMMIX, tem as seguintes características:

- Uma barra de menus ao cimo do écran agrupa as acções permitidas ao utilizador. Os menus são "desenroláveis", não ocupando espaço no écran e não requerendo memorização de comandos. Ao premir o "rato" num menu são automaticamente apresentados os ítems correspondentes às acções disponíveis que o utilizador pode escolher.
- Janelas múltiplas são utilizadas para apresentação ao utilizador de informação gráfica e numérica. Quando o "rato" é premido numa janela parcialmente escondida, esta é transportada para o primeiro plano do écran tornando-se assim acessível toda a informação que ela contém.
- Caixas de diálogo são usadas para a obtenção de informação suplementar do AD, necessária para a execução de um dado comando. São úteis também para transmitir informação ao utilizador (por exemplo, apresentação de mensagens de erro ou obtenção de permissão do utilizador para determinadas operações potencialmente "perigosas").
- Controlos pictóricos dão ao utilizador a possibilidade de, numa forma intuitiva, especificar as suas preferências por meio de botões ou indicadores que são movimentados com o "rato".

4 - BREVE DESCRIÇÃO DOS MÉTODOS INCLUIDOS EM TOMMIX

A categorização dos métodos interactivos de PLMO adoptada neste texto baseia-se no tipo de estratégia usada para a redução do âmbito da pesquisa, na informação requerida do AD e no tipo de função utilizada para construir um equivalente escalar do problema multiobjectivo por forma a obter uma solução eficiente (função escalarizante). Existem métodos em que as fases de interacção com o AD têm lugar ao fim de cada iteração do algoritmo (por ex^o no STEM), enquanto noutros há diversas fases de interacção em cada iteração.

4.1 - Método STEM

O "Step Method" (STEM), desenvolvido por Benayoun et al. [12] é um método interactivo de redução da região admissível. Em cada interacção o AD é chamado a especificar as quantidades de relaxamento para os valores das funções objectivo que considera satisfatórios de modo a melhorar as restantes. Em cada fase de cálculo é minimizada uma distância pesada de Tchebycheff à solução ideal.

Em cada iteração, o problema a otimizar reflecte as escolhas do AD feitas em iterações precedentes, através da redução da região admissível. É apresentada ao AD a solução de compromisso calculada em cada iteração minimizando a distância de Tchebycheff à solução ideal. Se os valores das funções objectivo são considerados satisfatórios, o processo termina; caso contrário, o AD deve especificar quais pretende relaxar de forma a melhorar os outros objectivos. A região admissível é

então reduzida através das limitações nos valores das funções objectivo (calculadas com base nas quantidades de relaxamento introduzidas pelo AD).

Em cada iteração é apresentado um gráfico de barras, mostrando o valor de cada função objectivo.

4.1 - Método de Zionts-Wallenius

O método de Zionts-Wallenius [13] reduz progressivamente o espaço dos pesos, de acordo com as respostas do AD em cada interacção face a comparações entre pares de soluções e julgamentos sobre as tendências de variação unitária que conduzem a outras soluções não dominadas. Em cada fase de cálculo é otimizada uma soma ponderada das funções objectivo.

Inicialmente é escolhido um conjunto de pesos e otimiza-se a respectiva soma ponderada dos objectivos. Obtem-se assim, e em cada interacção, um ponto extremo não dominado do poliedro admissível.

O AD manifesta as suas preferências através da comparação entre pares de soluções extremas não dominadas, e avaliação de tendências de variação unitária ao longo de arestas que têm origem na solução corrente e conduzem a outras soluções não dominadas.

Partindo das respostas dadas pelo AD, o método introduz então restrições no espaço dos pesos, reduzindo progressivamente o domínio admissível para a selecção de um novo conjunto de pesos. O processo termina quando o espaço dos pesos for reduzido a uma região suficientemente pequena de forma a que seja identificada uma solução final, convergindo para o óptimo de uma função utilidade implícita do AD. Parte-se do princípio que as respostas do AD nas fases de diálogo são coerentes com essa função utilidade implícita, embora no caso de serem detectadas incoerências seja dada a possibilidade de eliminar as restrições mais antigas no espaço dos pesos.

Para problemas com três funções objectivo é apresentado, em cada interacção, o espaço dos pesos e gráficos de barras mostrando o desempenho relativo de cada solução. Por forma a facilitar a comparação entre os dois gráficos são utilizados os mesmos padrões (ou cores) na representação de uma dada solução no espaço dos pesos e nos gráficos de barras mostrando os valores das funções objectivo correspondentes. Podem também ser visualizadas (ou omitidas para aumentar a clareza do gráfico) as restrições impostas no espaço dos pesos.

4.3 - Método "Interval Criterion Weights"

O método "Interval Criterion Weights", desenvolvido por Steuer [14], é um método interactivo que reduz progressivamente o cone dos critérios (cone convexo gerado pelos gradientes das funções objectivo). Esta redução é feita de acordo com a escolha feita pelo AD de uma solução de uma amostra de soluções não dominadas. Em cada fase de cálculo são optimizadas várias somas ponderadas das funções objectivo.

Partindo de um conjunto particular de combinações convexas dos gradientes das funções objectivo, calcula-se uma série de soluções, da qual é apresentada uma amostra ao AD. A solução escolhida pelo AD indica a combinação convexa de gradientes das funções objectivo que corresponde mais de perto às suas preferências. O cone dos critérios é contraído em torno desse gradiente, resultando o cone de critérios reduzido da iteração seguinte. O cone dos critérios é assim gradualmente contraído e deslocado até se focar numa pequena porção da superfície da região admissível, contendo o ponto extremo eficiente com maior valor para a função utilidade do AD.

Para problemas com três funções objectivo é apresentado o espaço dos pesos (correspondendo ao cone de critérios contraído), assim como uma projecção

bidimensional dos pontos extremos não dominados e arestas já calculados no espaço dos objectivos.

4.4 - Método "Pareto Race"

O método "Pareto Race", proposto por Korhonen e Wallenius [15], cuja implementação original é o programa VIG (Visual Interactive Goal Programming), é um método de pesquisa direccional. A informação requerida do AD consiste na especificação das funções objectivo a melhorar alterando a direcção do movimento. As soluções eficientes são obtidas por meio da optimização de uma função escalarizante e programação paramétrica em relação aos termos independentes.

A partir de níveis de aspiração para os valores das funções objectivo, especificados inicialmente pelo AD, constrói-se uma direcção de referência. A direcção de referência é então projectada sobre o conjunto das soluções eficientes, gerando uma trajectória (subconjunto das soluções eficientes) que é apresentada ao AD. Esta técnica permite ao AD percorrer a fronteira eficiente, controlando a direcção do movimento (privilegiando diferentes funções objectivo) e a velocidade (obtendo soluções mais ou menos próximas umas das outras).

São apresentados ao AD gráficos de barras representando os valores das funções objectivo que variam dinamicamente à medida que a pesquisa da face não dominada vai decorrendo. A velocidade e a direcção do movimento são controladas através de um painel de controlo.

4.4 - Método TRIMAP

O método TRIMAP, desenvolvido por Clímaco e Antunes [16], é um método de pesquisa livre baseado na aprendizagem progressiva e selectiva do conjunto das soluções não dominadas, que combina a redução da região admissível com a redução do espaço dos pesos. O AD pode especificar níveis mínimos para os valores das funções objectivo e impor restrições no espaço dos pesos. Em cada fase de cálculo é optimizada uma soma ponderada das funções objectivo.

A finalidade do método não é convergir para uma solução de melhor compromisso, mas sim ajudar o AD a eliminar os subconjuntos de soluções não dominadas que não lhe interessam. O processo interactivo termina quando o AD considera conhecer o bastante sobre o conjunto das soluções não dominadas. TRIMAP combina três procedimentos fundamentais: decomposição do espaço dos pesos, introdução de restrições no espaço dos objectivos e introdução de restrições no espaço dos pesos. Além disso, as limitações introduzidas nos valores das funções objectivo podem ser traduzidas para o espaço dos pesos. O método é dirigido para problemas com três funções objectivo, o que sendo uma limitação, permite o uso de meios gráficos particularmente adequados ao diálogo com o AD. Privilegia-se assim a capacidade de processamento de informação do AD, simplificando-se a fase de diálogo.

O objectivo principal é o preenchimento progressivo e selectivo do espaço dos pesos, evitando um estudo exaustivo de regiões que correspondam a valores de funções objectivo muito próximos (frequente em casos de estudo reais).

A redução do âmbito da pesquisa é feita principalmente impondo limitações nos valores das funções objectivo, que são traduzidas para o espaço dos pesos. A introdução destas limitações adicionais pode ainda ser usada para obter soluções não dominadas que não sejam pontos extremos do poliedro admissível. É também possível impor restrições directamente no espaço dos pesos.

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

- directamente, em que o AD escolhe um conjunto de pesos de uma zona do triângulo não preenchida, que lhe parece importante para continuar a pesquisa;

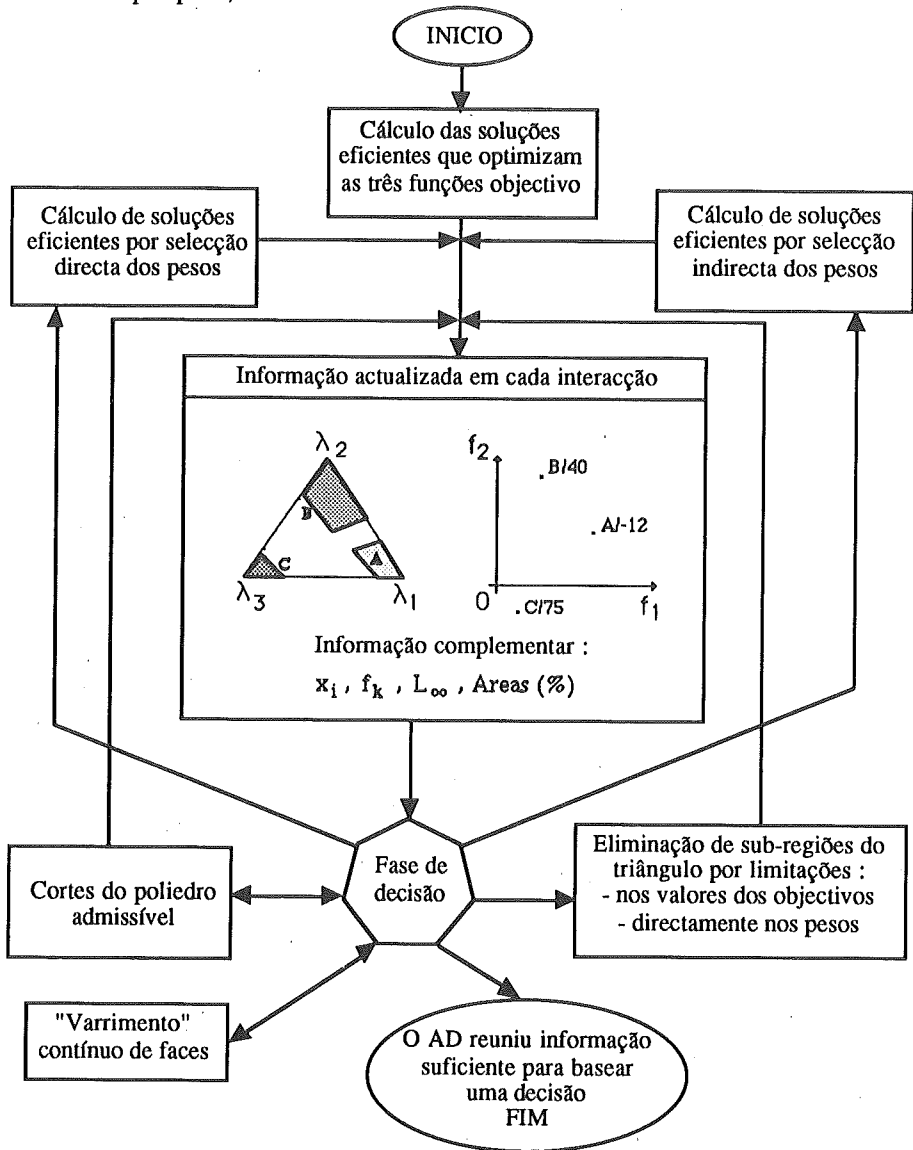


Fig. 2 - Diagrama de blocos do método TRIMAP

- indirectamente, construindo uma função ponderada cujo gradiente é normal a um plano de custo constante que passa por três soluções já calculadas e escolhidas pelo AD. Se os pesos assim obtidos não forem todos positivos, é feita uma pequena perturbação no gradiente da função objectivo ponderada por forma a assegurar essa condição.

O AD pode realizar "cortes" no poliedro admissível fixando o valor de uma ou duas funções objectivo. O programa calcula então as soluções não dominadas que

satisfazem estas restrições adicionais e que não são geralmente pontos extremos do poliedro admissível. TRIMAP permite ainda percorrer faces não dominadas.

Em cada interacção são apresentados ao AD dois gráficos. O primeiro é o espaço dos pesos mostrando as regiões correspondentes a cada um dos pontos extremos não dominados já conhecidos. Este gráfico mostra ainda as restrições à variação dos pesos, resultantes de limitações adicionais impostas nos valores das funções objectivo ou directamente introduzidas no espaço dos pesos. Para cada solução é apresentada informação complementar: a área da região correspondente no espaço dos pesos (em percentagem) e a distância de Tchebycheff à "solução ideal".

5 - UMA VISÃO GERAL DO FUNCIONAMENTO DE TOMMIX

TOMMIX permite total flexibilidade nas transições entre métodos (embora algumas das transições possam não acrescentar mais informação ao AD, todas são permitidas).

Na transição entre dois métodos, podem ou não ser consideradas as modificações previamente introduzidas (por ex., redução da região admissível ou contracção do cone dos critérios) de acordo com a vontade do AD. As modificações serão em princípio confirmadas sempre que o AD considere que as decisões tomadas anteriormente o conduziram a uma região de pesquisa de acordo com as suas preferências, tendo interesse em prosseguir a pesquisa num universo mais reduzido. As modificações podem ser desprezadas quando o AD considerar que não prosseguiu na " direcção correcta", não significando, no entanto, que não tenha sido um passo importante para um melhor conhecimento do problema. Neste caso, as preferências expressas não terão efeitos posteriores.

De acordo com o tipo de situação, podem distinguir-se diferentes " graus de aprendizagem" :

- a mudança de um método para outro, num dado momento do processo interactivo, significa, do ponto de vista do novo método (e devido à sua estrutura interna), reiniciar o processo (por ex., transição do ZW para o STEM);
- imposição de restrições na região de pesquisa, nomeadamente como resultado do estabelecimento de limitações inferiores nos valores das funções objectivo (ex., TRIMAP), ou contracção do cone dos gradientes das funções objectivo (ex., ICW);
- na transição de um método para outro, iniciar o novo método partindo da solução obtida na última interacção, existindo ou não redução da região de pesquisa (ex., TRIMAP-PR, TRIMAP-ZW).

Em certos casos, um método pode ser usado como um procedimento com objectivos diferentes daqueles que foram definidos no método original. Alguns exemplos são:

- executar uma (ou várias) iterações usando o método STEM pretendendo estabelecer limitações adicionais nos valores das funções objectivo;
- usar o método ZW localmente, i.e. partir de um ponto extremo eficiente e compará-lo com pontos extremos adjacentes e/ou fazer julgamentos sobre as variações unitárias nos valores das funções objectivo quando nos deslocamos desse ponto sobre uma aresta eficiente. Dá-se assim a possibilidade de introduzir restrições no espaço dos pesos, o que significa implicitamente o estabelecimento de limitações na região admissível.

Note-se que em TOMMIX não se pretende assegurar a convergência para o óptimo de uma função utilidade implícita do AD. Mesmo nos métodos onde isto é suposto (tal como foram originalmente propostos), a intenção é ajudar o AD a reunir o conhecimento suficiente sobre o problema para basear uma decisão.

É feita seguidamente uma breve referência à utilização mais comum (mas não exclusiva) de cada método, tendo em atenção as respectivas características fundamentais.

STEM é útil para a obtenção de soluções não dominadas que não são necessariamente pontos extremos e para uma rápida redução da região admissível nas primeiras interações com o AD. As limitações nos valores das funções objectivo impostas pelas quantidades de relaxamento pretendidas pelo AD podem ser consideradas nas transições para os métodos "Pareto Race", Zionts-Wallenius e TRIMAP. Na transição para estes dois últimos métodos essas limitações são traduzidas no espaço dos pesos.

O método Zionts-Wallenius é especialmente útil quando o AD está seguro das suas preferências, e para proceder a uma pesquisa local numa região anteriormente delimitada pelo AD. Durante a utilização do método, são apresentadas graficamente as restrições que vão sendo introduzidas no espaço dos pesos. É assim visualizada a região deste espaço onde a pesquisa deve prosseguir para ser coerente com respostas dadas previamente relativas à comparação de pontos extremos não dominados e/ou julgamentos de variações unitárias das funções objectivo. Quando ocorre uma transição para o método TRIMAP é apresentada graficamente a região reduzida do espaço dos pesos.

O método ICW é adequado para a redução da região admissível nas primeiras interações. Isto deve-se ao facto de o cone dos critérios se ir deslocando e contraindo "focando-se" numa parte da região admissível. Em cada interacção do método ICW é apresentado o espaço dos pesos correspondente ao cone dos critérios contraído. Quando é feita uma transição para os métodos Zionts-Wallenius ou TRIMAP é visualizado graficamente o espaço dos pesos correspondente ao cone dos critérios contraído ou ao original, respectivamente para os casos em que o AD pretenda ou não considerar o cone dos critérios contraído para o prosseguimento da sua pesquisa. Em caso negativo, o espaço dos pesos correspondente ao cone dos critérios contraído é representado dentro do espaço dos pesos do problema original.

O método "Pareto Race" tem especial interesse na exploração local de faces não dominadas (geralmente numa fase final do processo de pesquisa), sendo também útil para o estabelecimento de limitações nos valores das funções objectivo. Quando há transição do "Pareto Race" para os outros métodos são calculados todos os pontos extremos que definem a face onde a pesquisa direccionada foi interrompida. No caso de transição para o TRIMAP, todas essas soluções são adicionadas ao conjunto de soluções já calculadas em interações anteriores deste método. Para entrar nos métodos Zionts-Wallenius ou ICW, é pedido ao AD para escolher um desses pontos. Pode transitar-se para o método "Pareto Race" partindo do STEM, Zionts-Wallenius, ICW ou TRIMAP através de qualquer solução não dominada calculada nesses métodos que, nos casos do STEM ou TRIMAP, pode ser um ponto não extremo.

TRIMAP é um método especialmente adequado para uma pesquisa livre permitindo efectuar uma pesquisa global destinada a aumentar o conhecimento sobre o problema. O AD pode adoptar dois tipos de estratégias:

- proceder a uma pesquisa destinada a obter informação global sobre o problema e impor limitações adicionais nas funções objectivo (ou directamente no espaço dos pesos) uma vez identificados os valores mínimos que está disposto a aceitar, de acordo com o seu sistema de preferências;
- realizar uma pesquisa local, explorando determinadas zonas específicas.

6 - EXEMPLO ILUSTRATIVO

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

Restrições :

$$\begin{aligned}
 2.00 \ X1 + 1.00 \ X2 + 4.00 \ X3 + 3.00 \ X4 &\leq 60.00 \\
 3.00 \ X1 + 4.00 \ X2 + 1.00 \ X3 + 2.00 \ X4 &\leq 60.00
 \end{aligned}$$

Funções objectivo :

** f 1

$$3.00 \ X1 + 1.00 \ X2 + 2.00 \ X3 + 1.00 \ X4$$

** f 2

$$1.00 \ X1 - 1.00 \ X2 + 2.00 \ X3 + 4.00 \ X4$$

** f 3

$$-1.00 \ X1 + 5.00 \ X2 + 1.00 \ X3 + 2.00 \ X4$$

Suponhamos que o AD começa por usar o método ICW, com uma amostra de tamanho 3 (i.e. em cada interacção são apresentadas três soluções ao AD, entre as quais selecciona a sua preferida). Em cada iteração são resolvidos $2p+1$ problemas lineares (em que p é o número de funções objectivo). Os pontos dentro do espaço dos pesos representam as combinações convexas dos pesos (ponderando as funções objectivo para obter o respectivo equivalente escalar) automaticamente geradas pelo método.

Suponhamos que a solução 3 da fig. 3 é a solução preferida pelo AD (da amostra constituída pelas soluções 1, 2 e 3) e que é seleccionada para contrair o cone dos critérios.

Abandonando o método ICW, pode transitar-se para o método TRIMAP de duas formas:

- considerando o cone contraído (fig. 4a);
- considerando o cone original, e visualizando o espaço dos pesos correspondente ao cone contraído sobreposto no espaço dos pesos do problema original (fig. 4b).

Suponhamos que a transição para o método TRIMAP é feita de acordo com a segunda opção.

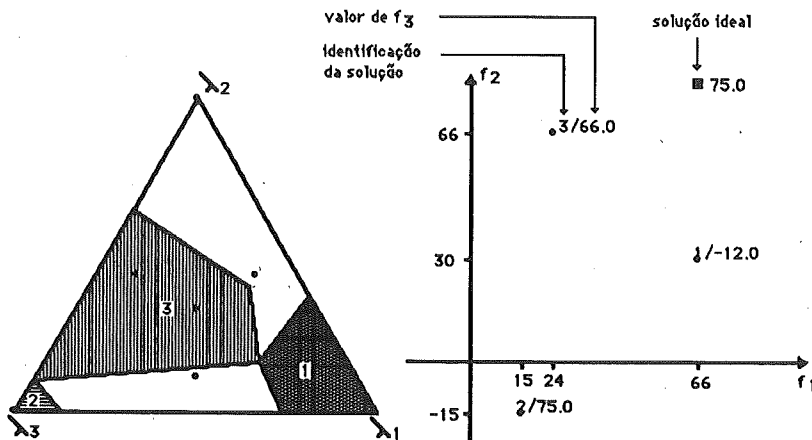


Fig. 3 - Método "Interval Criterion Weights"

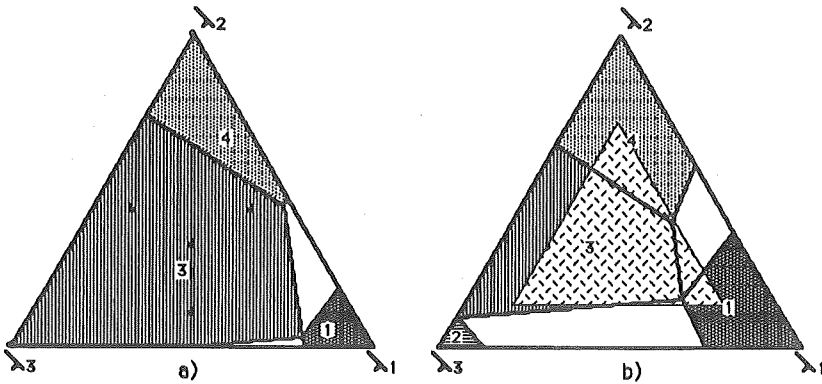


Fig. 4 - Transição do ICW para o TRIMAP

TRIMAP é especialmente útil para o estabelecimento de limitações nos valores das funções objectivo. Esta informação é traduzida para o espaço dos pesos, apresentando-se as regiões correspondentes às soluções não dominadas que satisfazem as restrições impostas pelo AD (ver fig. 5, em que foi introduzida a restrição $f_2(x) \geq 40$, satisfeita na região do triângulo a tracejado que inclui a região ocupada pelas soluções 3 e 4).

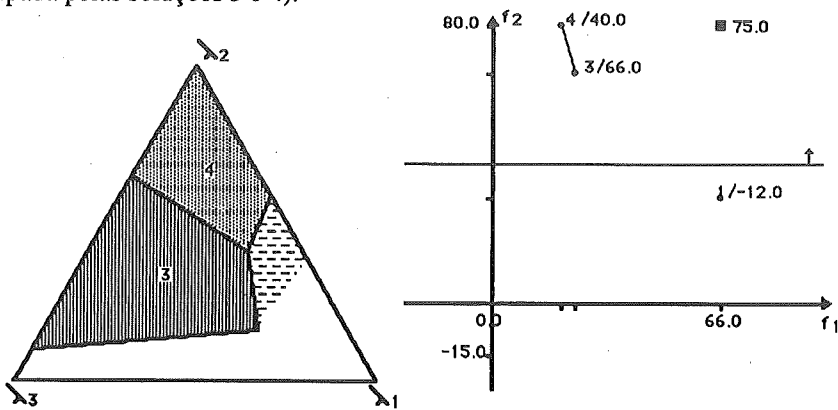


Fig. 5 - Uso do TRIMAP para impor restrições adicionais nos valores das funções objectivo

É feita em seguida uma transição para o STEM, considerando a restrição $f_2(x) \geq 40$ imposta durante o uso do TRIMAP.

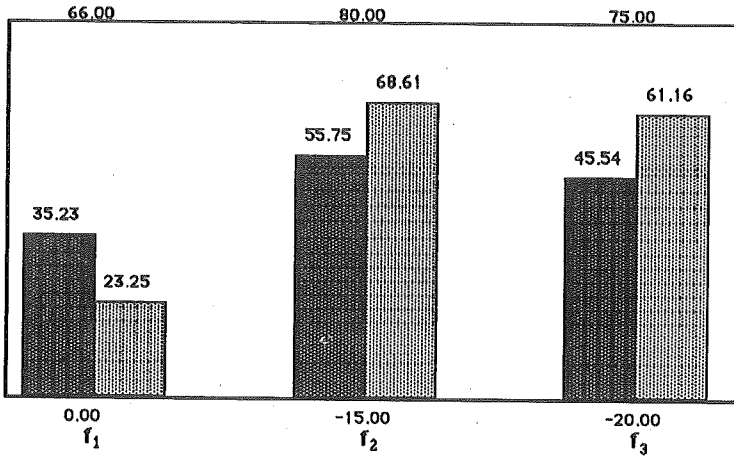


Fig. 6 - Uso do STEM

O resultado da primeira iteração do STEM (considerando a restrição adicional), é apresentado na fig. 6 (barras com padrão cinzento escuro).

O AD decide relaxar $f_1(x)$ de 15 e fazer mais uma iteração do método STEM (barras com padrão cinzento claro na fig. 6). Abandonando o método STEM, é dada ao AD a possibilidade de considerar ou não as restrições introduzidas, ou mesmo modificá-las. Suponhamos que para continuar a pesquisa numa região mais reduzida, usando o método TRÍMAP, o AD decide considerar as restrições $f_1 \geq 20.23$, $f_2 \geq 55.75$, $f_3 \geq 45.54$ (introduzidas durante o uso do STEM). Na fig. 7a, as subregiões do espaço dos pesos correspondentes a cada uma destas limitações adicionais são preenchidas com diferentes padrões que se sobrepõem, permitindo assim visualizar claramente as respectivas zonas de intersecção.

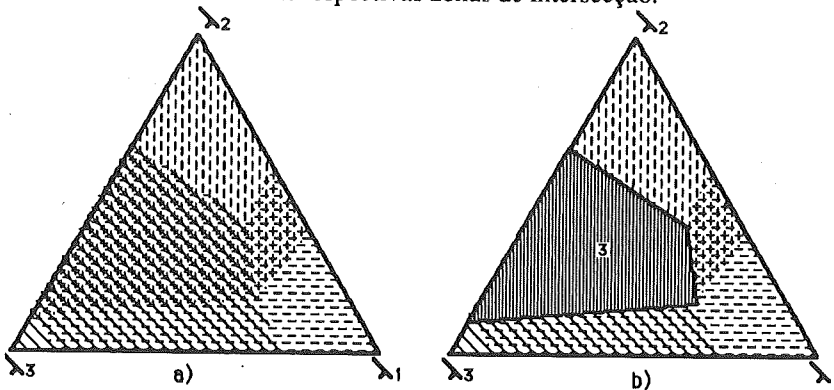


Fig. 7 - Regiões do espaço dos pesos que satisfazem as restrições adicionais "importadas" do STEM

Após ser realizada uma pesquisa exaustiva na região onde todas as restrições se intersectam, conclui-se que a solução 3 da fig. 7b é o único ponto extremo não dominado que satisfaz simultaneamente todas as limitações nos valores das funções objectivo. De modo a obter um conhecimento local dos pontos extremos não dominados próximos da solução 3 ($f_1=24.0$, $f_2=66.0$, $f_3=66.0$), é feita uma transição para o método Zionts-Wallenius, partindo desta solução (fig. 8).

Suponhamos que o AD prefere sempre a solução 3 quando é chamado pelo método Zionts-Wallenius a comparar esta solução com as suas adjacentes. Se o AD não aceitar também nenhuma das tendências de variação unitária correspondentes às arestas não dominadas que ligam esta solução às soluções adjacentes, estaria então encontrada a solução final.

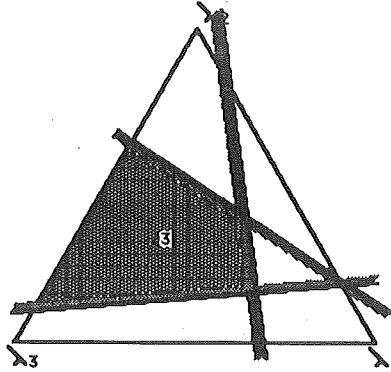


Fig. 8 - O espaço dos pesos com a representação das restrições resultantes das respostas a comparações de soluções, no método Zionts-Wallenius

Nesta fase do processo de decisão, Pareto Race é um método especialmente interessante para efectuar uma exploração local de faces não dominadas. Partindo da solução 3 e seleccionando uma direcção de modo a melhorar a função objectivo 1, usando os controlos disponíveis (fig. 9), o AD é conduzido para a face não dominada que contém os pontos extremos 1 e 3.

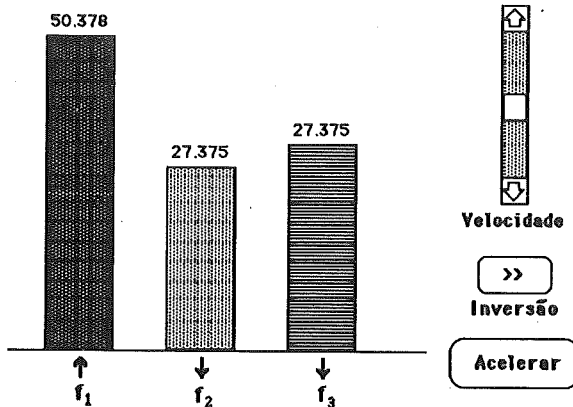


Fig. 9 - Utilização do método Pareto Race

7 - APLICAÇÕES DA BASE DE MÉTODOS DE PLMO

No que se refere a aplicações TOMMIX foi utilizado num problema de planeamento da expansão de um sistema produtor de energia e no estudo de estratégias de evolução de redes de telecomunicações [6,17,18].

No planeamento da expansão de um sistema produtor de energia (baseado na evolução da carga) as três funções objectivo objectivo quantificam o custo total, o impacto ambiental (ambas a minimizar) e a fiabilidade (a maximizar). As variáveis de decisão referem-se ao tipo de tecnologias de geração (fuel, nuclear e carvão) consideradas para a expansão do sistema produtor. As restrições podem ser

classificadas em três categorias: restrições de carga (toda a potência geradora em cada período de tempo deve ser pelo menos igual à carga nesse intervalo), restrições operacionais das unidades geradoras (impondo que a potência de uma unidade não pode exceder a respectiva potência nominal, previamente afectada por um factor de disponibilidade) e restrições orçamentais relativas ao maior valor admissível para a função objectivo custo. Detalhes sobre este caso de estudo podem ser encontrados em [17].

No estudo da evolução de redes de telecomunicações, os objectivos considerados são o valor da modernização da rede (a maximizar), uma função representando a dependência externa associada a uma dada política de evolução (a minimizar) e uma função representando o grau de modernização dos serviços oferecidos (a maximizar). As variáveis de decisão referem-se ao tipo de serviço oferecido a uma dada linha de assinante (serviço telefónico tradicional, serviço de dados de banda estreita e serviço digital integrado de banda larga) e ao tipo de transmissão usado para suportar esse serviço. As restrições expressam aspectos técnicos, orçamentais, de mercado e sociais envolvidos no problema. Detalhes sobre o modelo matemático a a aplicação de TOMMIX podem ser encontrados em [18].

8 - CONCLUSÕES

Neste artigo pretendeu fazer-se uma síntese do trabalho que temos vindo a desenvolver em métodos interactivos de PLMO. Tendo por base a implementação computacional de alguns dos mais representativos métodos interactivos de PLMO, a respectiva aplicação a casos de estudo em planeamento energético e a inovação a nível metodológico, foi desenvolvida uma base integrada e flexível de métodos, que constitui a componente fundamental de um sistema de apoio à decisão.

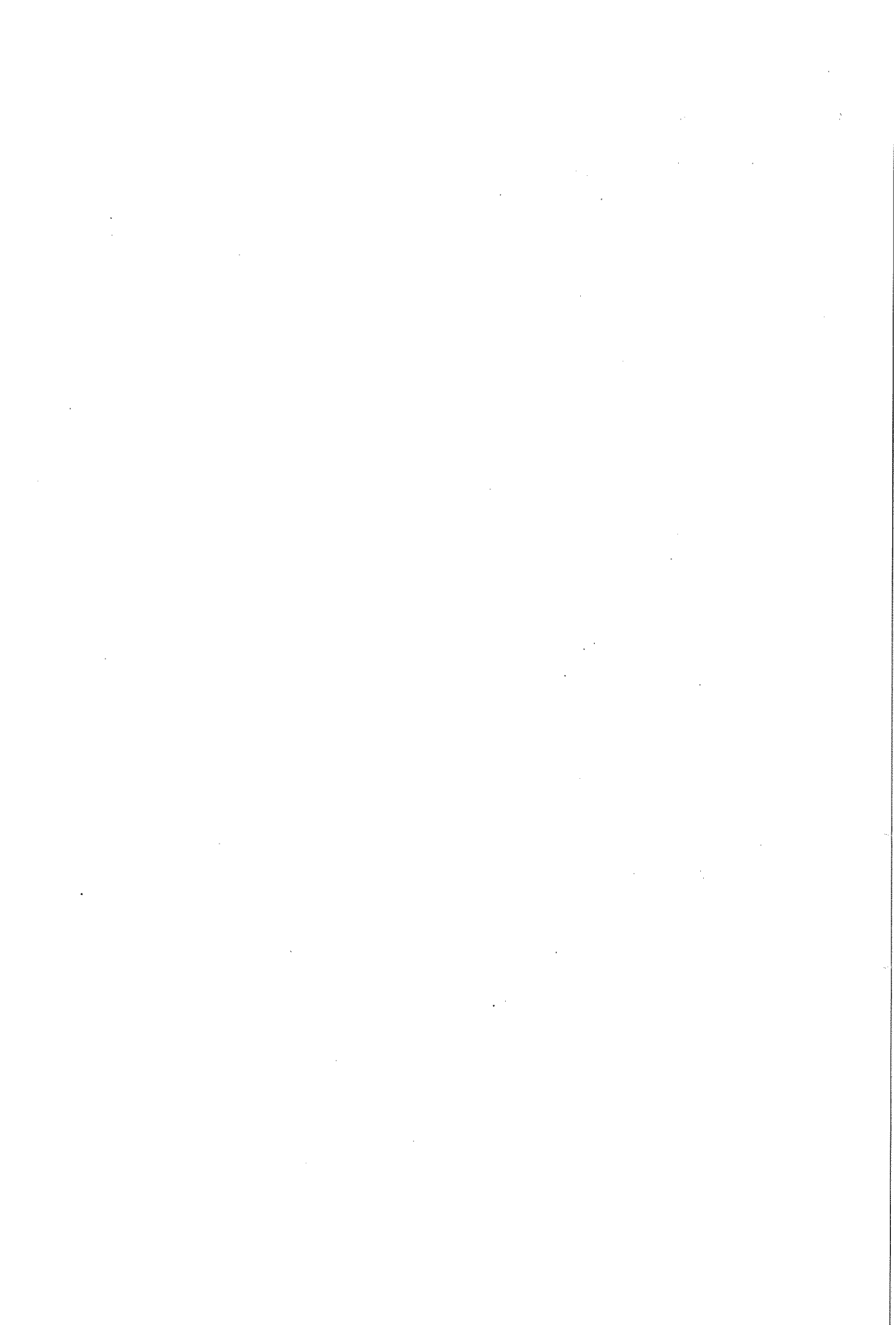
A base de métodos de PLMO integra actualmente os métodos interactivos STEM, Zions-Wallenius, Interval Criterion Weights, Pareto Race e TRIMAP (representativos de diferentes estratégias de redução do âmbito da pesquisa). TOMMIX permite tirar partido da combinação dos métodos e da transferência de informação recolhida em cada um deles que possa ser usada nas fases seguintes do processo de decisão. À base de métodos encontram-se associados um conjunto de procedimentos auxiliares, uma base de diálogo que controla todos os aspectos de interacção homem-máquina e um módulo de gestão de dados.

As reflexões feitas em torno de TOMMIX colocaram em evidência a importância da interacção homem-máquina (nomeadamente da utilização de gráficos e de um ambiente computacional amigável com o utilizador) e as características de flexibilidade desejáveis num utensílio deste tipo. A base integrada de métodos não pretende "produzir decisões", mas sim constituir um meio eficaz de "apoio à decisão", ajudando o AD na aprendizagem do problema, proporcionando um processo construtivo e criativo do seu sistema de preferências. Esta aproximação distancia-se das estratégias que reduzem o processo interactivo a uma convergência para o óptimo de uma função utilidade implícita do AD.

REFERENCIAS

- [1] Steuer, R. *Multiple Criteria Optimization: Theory, Computation and Application*. Wiley, (1986).
- [2] Vanderpooten, D.; Ph. Vincke. Description and analysis of some representative interactive multicriteria procedures. *Mathematical and Computer Modelling* 12, 1221-1238 (1989).
- [3] Zions, S.; Multiple criteria mathematical programming: an updated overview and several approaches. In *Mathematical models for decision*

- support, G. Mitra (ed.), Springer-Verlag NATO ASI Series F 48, 135-167 (1988).
- [4] Weistroffer, H. R.; S. Narula; K. Koong. An overview of recent MCDM applications. In Proc. of the Int. Conf. on *Multiple Criteria Decision Making*, Bangkok, M. Tabucanon and V. Chankong (eds.), 1109-1123 (1989).
- [5] Clímaco, J; C. H. Antunes. A comparison of microcomputer implemented interactive MOLP methods based on a case study. *Readings in Multiple Criteria Decision Aid*, C. Bana e Costa (ed.), Springer-Verlag 445-470 (1990).
- [6] Antunes, C. H.; M. J. Alves; A. L. Silva; J. Clímaco. A computer implementation of an interactive MOLP model base with applications to energy and telecommunications planning. In Proc. of the Int. Conf. on *Multiple Criteria Decision Making*, Bangkok, M. Tabucanon and V. Chankong (eds.), 401-420 (1989).
- [7] Clímaco, J.; C. H. Antunes. Flexible method bases and man-machine interfaces as key features in interactive MOLP approaches. Apresentado em *IASA Workshop on Multiple Criteria Decision Support*, Helsinki (1989).
- [8] Steuer, R. E. ; A. Whisman. Toward the consolidation of interactive multiple objective programming procedures. *Lecture Notes in Economics and Mathematical Systems* 273, G. Fandel, M. Grauer, A. Kurzhanski and A. P. Wierzbicki (eds.), Springer Verlag, 232-241 (1986).
- [9] Antunes, C. H.; M. J. Alves; A. L. Silva; J. Clímaco. An integrated MOLP method base package - a guided tour of TOMMIX. A ser publicado em *Computers and Operations Research*.
- [10] French, S. Interactive multi-objective programming : its aims, applications and demands. *Journal Operational Research Society*. 35, 827-834 (1984).
- [11] Buchanan, J. T.; H. G. Daellenbach. A comparative evaluation of interactive solution methods for multiple objective decision models. *European Journal Operational Research* 29, 353-359 (1987).
- [12] Benayoun, R.; J. de Montgolfier; J. Tergny; O. Larichev. Linear programming with multiple objective functions: step method (STEM). *Mathematical Programming* 1, 366-375 (1971).
- [13] Zionts, S.; J. Wallenius. An interactive multiple objective linear programming method for a class of underlying nonlinear utility functions. *Management Science* 29, 519-529 (1983).
- [14] Steuer, R. E. An interactive multiple objective linear programming procedure. *TIMS Studies in the Management Sciences* 6, 225-239 (1977).
- [15] Korhonen, P.; J. Wallenius. A Pareto race. *Naval Research Logistics* 35, 615-623 (1988).
- [16] Clímaco, J.; C. H. Antunes. Implementation of a user friendly software package - a guided tour of TRIMAP. *Mathematical and Computer Modelling* 12, 1299-1309 (1989).
- [17] Clímaco, J.; C. H. Antunes; A. Martins; J. Sá Marta; A. T. Almeida. A decision support system for power generation expansion planning with a case study. *Proceedings of the 29th IEEE Conference on Decision and Control* 264-270, Honolulu, USA, 1990 .
- [18] Antunes, C. H.; J. Clímaco; J. Craveirinha. A multiple criteria model for new telecommunication service planning. *Submetido para publicação*.



ÁRVORES DE SUPORTE COM RESTRIÇÕES DE GRAU

UMA APLICAÇÃO AO PROBLEMA DO CAIXEIRO VIAJANTE MÚLTIPLO

Maria Cândida Mourão
Maria João Ralha
Departamento de Matemática
Instituto Superior de Economia e Gestão (UTL)

Resumo

O Problema do Caixeiro Viajante Múltiplo (*m-TSP*) consiste na determinação das rotas óptimas a atribuir a dois ou mais caixeiros, que partindo todos de uma mesma cidade, a ela regressam no final da viagem. Todas as cidades, à excepção da inicial, são visitadas exactamente uma vez por um único caixeiro.

O presente trabalho tem por objectivo testar a eficiência, relativamente à qualidade das soluções, de um método heurístico para o Problema do Caixeiro Viajante Múltiplo. Tendo por base a determinação de uma árvore de suporte com restrições de grau num vértice, foram implementadas duas variantes do método, que se distinguem pelo algoritmo aproximativo para o *TSP* utilizado para encontrar uma solução admissível a partir desta árvore.

Os valores das soluções obtidas pelas duas variantes, num conjunto de problemas-teste, são comparados entre si e com o da heurística de *Frieze*. Dos resultados apresentados pode observar-se que o método proposto por *Frieze*, embora seja o que tem melhor análise de pior caso, para problemas completos simétricos e que verifiquem a desigualdade triangular, não é o que melhor se comporta na prática.

Palavras Chave

Problema do Caixeiro Viajante Múltiplo; Método Heurístico; Árvore de Suporte, com Restrições de Grau, de Custo Mínimo; Dualidade Lagrangeana.

1. INTRODUÇÃO

Neste trabalho são apresentadas e testadas computacionalmente duas variantes de um método heurístico para o Problema do Caixeiro Viajante Múltiplo.

O Problema do Caixeiro Viajante Múltiplo ("The Multiple Traveling Salesman Problem": *m-TSP*) consiste em determinar o conjunto de percursos a serem efectuados por *m* caixeiros, num dado conjunto de *n* cidades (para as quais se conhece a distância entre quaisquer duas cidades). Todos os caixeiros iniciam a viagem numa cidade, previamente fixada - cidade inicial ou depósito - onde têm que regressar no final da viagem. Cada cidade, com excepção da inicial, é visitada uma única vez por um só caixeiro e pretende-se que a distância total percorrida seja mínima.

Este problema é uma das generalizações do "bem conhecido" Problema do Caixeiro Viajante (*TSP*) no qual se considera apenas um caixeiro. São ambos problemas de "difícil" resolução, ou seja, para os quais não existem, e se presume ser impossível desenvolver, algoritmos que, no pior caso, determinem a sua

solução ótima em tempo de execução polinomial. Tal justifica a importância do desenvolvimento de "bons" métodos heurísticos, ou seja métodos que forneçam soluções admissíveis de valores não muito afastados do valor ótimo do problema.

Sendo o *m-TSP* uma extensão do Problema do Caixeiro Viajante Simples, parece natural que grande parte das heurísticas surjam como adaptação das existentes para o *TSP*. Pode ainda pensar-se numa transformação do *m-TSP* num *TSP* ([1], [7], [8], [10], [11], [12], [13]) e utilizar qualquer dos métodos existentes para o *TSP*. Contudo esta alternativa origina um aumento pouco conveniente da dimensão do problema.

Frederickson, Hecht & Kim [4], adaptam à determinação de uma solução admissível para o *m-TSP* dois métodos de construção para o *TSP* - o método da inserção mais próxima e o do vizinho mais próximo, que partindo de um nó inicial se diferenciam pela forma como é seleccionado o vértice seguinte a integrar no circuito. Apresentam, ainda uma terceira heurística ("*m-Splitour*") que consiste na subdivisão de um circuito para o *TSP* em *m* subcircuits. Este método é dos três o que apresenta melhor análise de pior caso, ou seja o que, no pior caso, obtém uma solução admissível com valor mais próximo do valor ótimo do *m-TSP*, caso seja utilizada a heurística de *Christofides* [2] para a determinação do circuito inicial para o *TSP*.

O método heurístico de *Frieze* [3] é uma generalização do de *Christofides* [2] para o *TSP*. Aplicável a problemas cujos grafos sejam completos este método é o que proporciona, em casos simétricos e que verifiquem a desigualdade triangular, melhor análise de pior caso. Nestas condições, o valor da solução por ele encontrado, não é mais do que 50% superior ao valor ótimo, resultado que *Frieze* [3] demonstra manter-se válido para o *m-TSP*.

Neste artigo é apresentada uma heurística que começa por determinar uma árvore de suporte com restrições de grau num vértice, de custo mínimo (*m-SST*) a partir da qual se definem subgrafos com os quais se obtém *m* soluções admissíveis para *m TSP's*, aplicando um qualquer método heurístico. Assim, diferentes métodos heurísticos para o *TSP* originam diferentes soluções para um mesmo *m-TSP*.

À definição e formulação do *m-TSP* como um problema de programação linear inteira segue-se a descrição do algoritmo global.

Com o objectivo de testar os métodos foi utilizado um programa para a determinação de minorantes do valor ótimo do problema. As variantes do método implementado são comparadas entre si e com o de *Frieze* em termos dos "gap's" que se obtiveram para um conjunto de problemas-teste cujos dados foram gerados aleatoriamente.

Com base na análise dos resultados são tiradas conclusões e enunciados alguns pontos de interesse para futuras investigações.

Em anexo são apresentados, de forma sucinta, os algoritmos utilizados para a determinação da *m-SST* e da solução admissível para o *TSP*.

2. FORMULAÇÃO MATEMÁTICA DO PROBLEMA DO CAIXEIRO VIAJANTE MÚLTIPLO

Neste ponto formula-se o *m-TSP* como um problema de programação inteira. Optou-se por apresentar a formulação proposta por *Gavish & Srikanth* [6], pois, como se verá, com base nela torna-se fácil a justificação do método heurístico desenvolvido. Por outro lado, o valor do minorante, necessário ao cálculo dos "gap's", foi também determinado a partir da mesma formulação, com uma implementação do método sugerido pelos mesmos autores ([6]), que se justifica no final deste ponto.

O Problema, tal como foi descrito, pode ser representado através de um grafo $G = (V, A)$ onde:

$V = \{1, 2, \dots, n\}$ é o conjunto dos vértices do grafo representando as n cidades do problema;

e

A o conjunto das arestas ou arcos do grafo, representando as ligações entre as cidades.

A cada aresta, ou arco, encontra-se associado um custo (interessa por vezes considerar a distância, o valor, etc.) que se representará por:

c_{ij} - custo de um caixeiro viajar da cidade $i \in V$ para a cidade $j \in V$;

$C = [c_{ij}]_{\substack{i=1, \dots, n \\ j=1, \dots, n}}$ matriz dos custos.

Seja :

1 - cidade de partida e de chegada, por vezes também designada por cidade inicial e final ou, simplesmente, por depósito;

$S = V \setminus \{1\}$ - conjunto dos vértices do grafo com excepção do inicial;

$m = n^\circ$ de caixeiros viajantes.

Formulação

Definam-se as variáveis binárias

$$x_{ij} = \begin{cases} 1 & \text{se um caixeiro viaja directamente da cidade } i \in S \\ & \text{para a cidade } j \in S \text{ ou de } j \in S \text{ para } i \in S \\ 0 & \text{caso contrário (c.c.)} \end{cases}$$

$$x_{1j} = \begin{cases} 1 & \text{se um caixeiro viaja directamente da cidade } 1 (i \in S) \\ & \text{para a cidade } j \in S(1) \\ 0 & \text{c.c.} \end{cases}$$

$$(x_{i1})$$

Considerando o caso simétrico e m fixo, o m -TSP pode ser formulado ([6]) como um problema de programação inteira, e consiste em:

- Determinar o valor das variáveis binárias x_{ij} que satisfazem:

$$(P) \quad z^* = \text{Min} \quad \sum_{i=1}^{n-1} \sum_{j=i+1}^n c_{ij} x_{ij} + \sum_{i \in S} c_{i1} x_{i1} \quad (1)$$

$$\text{s.a:} \quad \sum_{j \in S} x_{1j} = m \quad (2)$$

$$\sum_{i \in S} x_{i1} = m \quad (3)$$

$$x_{j1} + \sum_{i < j} x_{ij} + \sum_{i > j} x_{ji} = 2 \quad \forall j \in S \quad (4)$$

$$\sum_{\substack{i, j \in S_k \\ i < j}} x_{ij} \leq |S_k| - 1 \quad \forall S_k \subseteq S, S_k \neq \emptyset \quad (5)$$

$$x_{ij} = 0, 1 \quad 1 \leq i < j \leq n \quad (6)$$

$$x_{i1} = 0, 1 \quad \forall i \in S. \quad (7)$$

As restrições

- (2) e (3) garantem que, em qualquer solução, da cidade (vértice) inicial, 1 , partem exactamente m caixeiros (arestas) a ela regressando no final da viagem (existem m arestas incidentes em 1 representando as arestas de retorno dos caixeiros);
- (4) garantem que, em qualquer solução, em cada vértice de S incidem apenas duas arestas, ou seja, o grau de cada vértice é dois. Assim, cada cidade, com excepção da inicial, é visitada uma e uma só vez;
- (5) impedem a formação de subcircuitos nos subconjuntos de S , i.é., em qualquer solução, para qualquer $S_k \subseteq S$ existe sempre uma aresta que une um vértice de S_k com um vértice de $S \setminus S_k$, pois o número de arestas que ligam dois vértices de S_k nunca é superior a $|S_k| - 1$.

A função objectivo (1) garante a minimização da distância total, ou seja da soma das distâncias percorridas por cada um dos m caixeiros.

No método utilizado para o cálculo de minorantes, sugerido por *Gavish & Srikanth* [6], considera-se a relaxação Lagrangeana das $n-1$ restrições (4). Adiciona-se previamente a (P) a restrição:

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^n x_{ij} = n - 1 \quad (8)$$

que, sendo redundante nesta formalização, deixa de o ser quando relaxadas as restrições (4). Com a inclusão de (8), qualquer solução admissível do problema relaxado contém $n - 1 + m$ arestas (m resultantes das restrições (3) e $n - 1$ desta) igual ao número de arestas de qualquer solução admissível do m -TSP. Assim, na relaxação apenas não se impõe o facto de cada cidade, com excepção da inicial, ser visitada exactamente uma vez por um só caixeiro (ou seja, o grau de cada vértice ser dois).

Se, como é usual, se considerar π um vector de multiplicadores de Lagrange e fizer:

$$\tilde{c}_{ij} = c_{ij} - \pi_i - \pi_j \quad \text{com} \quad \pi_1 = 0$$

resultam dois subproblemas de fácil resolução, podendo o valor do minorante ser calculado com base no valor do solução fornecida por cada um dos subproblemas.

É fácil notar que a parte respeitante às variáveis x_{ij} do problema relaxado ou seja:

$$\text{Min} \quad \sum_{i=1}^{n-1} \sum_{j=i+1}^n \tilde{c}_{ij} x_{ij}$$

$$\text{s.a.:} \quad (2), (5), (6) \text{ e } (8)$$

traduz o problema da determinação de uma árvore de suporte com restrições de grau no vértice inicial, 1 , de menor custo possível (m -SST), pois a restrição (2) impõe que grau (1) = m .

Por outro lado, e no que concerne às variáveis x_{i1} tem-se:

$$\text{Min} \quad \sum_{i=2}^n \tilde{c}_{i1} x_{i1}$$

$$\text{s.a.:} \quad (3) \text{ e } (7)$$

representando a determinação das m arestas de menor custo incidentes no vértice 1 .

O valor do minorante que se obtém resolvendo o problema relaxado, para um dado vector π , ou seja o valor de:

$$L(\pi) = \text{Min} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \tilde{c}_{ij} x_{ij} + \text{Min} \sum_{i=2}^n \tilde{c}_{i1} x_{i1} + \sum_{j=2}^n \pi_j$$

pode ser melhorado pela resolução do problema Dual Lagrangeano associado. Para tal pode recorrer-se aos métodos usuais de optimização subgradiente.

É óbvio que a solução óptima do *m-TSP* poderia ser encontrada a partir das soluções óptimas de *m TSP's* se se soubesse, á partida, quais as cidades que cada um dos *m* caixeiros visitaria. O método para a determinação de minorantes, acima descrito, sugeriu a ideia de determinar soluções aproximadas do problema, tomando como base de construção dos *m* percursos a *m-SST* correspondente ao melhor minorante obtido, como se detalha no ponto seguinte.

3. MÉTODO HEURÍSTICO

Como anteriormente se referiu, a solução admissível para o *m-TSP* foi encontrada a partir de uma árvore em que o grau do vértice 1 seja *m*, de menor custo possível e de uma solução admissível para cada um de *m TSP's*.

No presente ponto é descrito o método heurístico global utilizado sendo em anexo apresentados os algoritmos utilizados nos diversos passos desta heurística.

Descrição Geral do Método Heurístico

Considerem-se *m* subgrafos de $G = (X, A)$:

$$G_\ell = \langle \{X_\ell\} \rangle \quad \ell = 1, \dots, m$$

onde $X = \bigcup_{\ell=1}^m X_\ell, X_\ell \subset X, \forall \ell_1, \ell_2 X_{\ell_1} \cap X_{\ell_2} = \{1\}$.

Como é óbvio uma solução admissível para o *m-TSP* pode ser encontrada com base em soluções admissíveis para os *TSP* nos *m* subgrafos G_ℓ .

Os subgrafos foram construídos com base numa árvore de suporte, onde o grau do vértice 1 seja igual ao número de caixeiros, *m*, de menor custo possível, que se designará por *m-SST*. Designando por subárvores as árvores resultantes de retirar o vértice 1 da *m-SST*, o conjunto de vértices de cada um dos *m* subgrafos coincide com os vértices de cada uma das subárvores da *m-SST* e o vértice 1. Ou seja:

$$X_\ell = \{1\} \cup V_\ell \quad \text{onde } V_\ell \text{ é o conjunto de vértices de uma das } m \text{ subárvores da } m\text{-SST.}$$

Ex.: Num problema com $n = 11$ cidades e $m = 3$ caixeiros, seja:

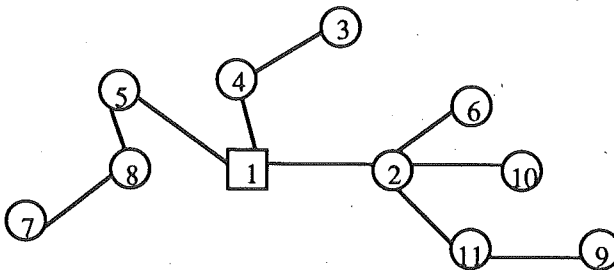


Fig. 3.1

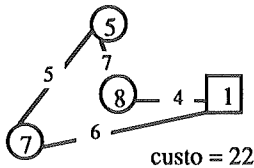
uma 3-SST. Determina-se então, uma solução admissível para cada um dos três TSP nos subgrafos:

$$G_1 = \langle \{1, 2, 6, 9, 10, 11\} \rangle$$

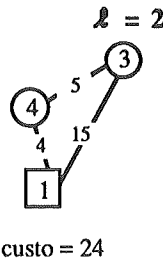
$$G_2 = \langle \{1, 3, 4\} \rangle$$

$$G_3 = \langle \{1, 5, 7, 8\} \rangle$$

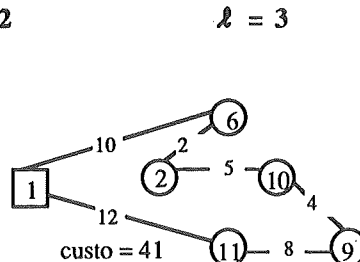
obtendo-se $\ell = 1$



(a)



(b)



(c)

Fig. 3.2

Neste caso, a solução admissível para o 3-TSP pode ser dada por:

O 1º caixeiro realiza o percurso da Fig. 3.2 (a); O 2º realiza o percurso da Fig. 3.2 (b); E o 3º realiza o percurso da Fig. 3.2 (c). O seu valor é $87 \geq z^*$. ♦

Logo, a determinação de uma solução admissível para o m -TSP e do seu valor, representando um majorante para o valor óptimo do problema, pode ser feita a partir do algoritmo que se descreve de seguida.

Algoritmo:

1. Determinar, T, uma árvore de suporte, onde o vértice 1 tenha grau m , de menor custo possível ([5]);
 - $\ell \leftarrow 1$;
2. Enquanto $\ell \leq m$
 - 2.1 Identificar o subgrafo $G_\ell = \langle \{X_\ell\} \rangle$, onde X_ℓ é o conjunto formado pelos vértices de uma subárvore ainda não considerada de T e o vértice 1; Seja A_ℓ o seu conjunto de arestas;
 - 2.2 Encontrar uma solução admissível para o TSP no subgrafo G_ℓ , considerando a matriz de custos C_ℓ , submatriz de C referente às arestas de A_ℓ (neste trabalho, a determinação de soluções admissíveis para o TSP foi feita utilizando o Método de Fusão ([9]) e o de Inserção de Menor Custo ([14]));
 - 2.3 Fazer $\ell \leftarrow \ell + 1$;

A solução admissível para o m -TSP é dada pelas m soluções admissíveis encontradas em 3, representando cada uma o percurso de um dos caixeiros. O valor da solução é a soma dos valores destas m soluções.

4. ANÁLISE DOS RESULTADOS COMPUTACIONAIS

Os métodos apresentados foram testados num Microcomputador PC-AT/286, sistema operativo MS-DOS 3.30. Os programas foram codificados em Pascal e compilados em Turbo Pascal 5.0. O programa para a determinação do emparelhamento perfeito de custo mínimo, necessário ao método heurístico de Frieze [3], estando codificado em Fortran foi compilado em Profor.

As coordenadas dos pontos foram geradas aleatoriamente seguindo uma distribuição uniforme em [0, 200], sendo de seguida calculadas as distâncias Euclidianas entre os diversos pontos, resultando assim problemas simétricos e que verificam a desigualdade triangular.

Os desvios percentuais dos majorantes foram determinados de acordo com a fórmula:

$$\text{gap} = (\bar{z} - z) / z$$

onde: \bar{z} é o valor do majorante;

z é o valor do minorante.

Os valores que se apresentam nas tabelas I e II resumem, respectivamente, os "gap's" e os tempos de execução, obtidos por cada um dos métodos, quando aplicados a um conjunto de problemas-teste. Assim, cada linha contém os valores mínimos, médios e máximos de 30 problemas de iguais dimensões (n, m).

"GAP's" (%)

| Dim. n, m | Frieze | | | Fusão | | | Inserção de Menor Custo | | |
|--------------|--------|-------------|-----|-------|-------------|-----|----------------------------|-------------|-----|
| | Min | Méd | Máx | Min | Méd | Máx | Min | Méd | Máx |
| 30, 2 | 5 | <u>12.4</u> | 22 | 5 | <u>14.7</u> | 27 | 7 | 14.3 | 27 |
| 40, 2 | 4 | <u>12.7</u> | 20 | 5 | <u>17.4</u> | 28 | 5 | 16.3 | 30 |
| 30, 5 | 10 | <u>19.2</u> | 37 | 5 | 12.7 | 26 | 5 | <u>12.6</u> | 23 |
| 40, 5 | 9 | <u>17.8</u> | 35 | 9 | 14.5 | 24 | 7 | 12.9 | 23 |
| 50, 5 | 12 | 17.3 | 28 | 14 | <u>18.1</u> | 25 | 13 | 16.9 | 24 |
| 60, 5 | 13 | 16.9 | 23 | 11 | <u>17.0</u> | 24 | 9 | 15.8 | 23 |
| 50, 10 | 16 | <u>26.8</u> | 43 | 6 | <u>12.5</u> | 19 | 7 | <u>12.5</u> | 21 |
| 60, 10 | 18 | <u>25.9</u> | 37 | 6 | 12.7 | 18 | 7 | 12.7 | 17 |
| 70, 10 | 16 | <u>22.1</u> | 30 | 7 | 14.8 | 21 | 6 | 13.3 | 23 |
| 80, 10 | 15 | <u>21.6</u> | 31 | 9 | 16.1 | 23 | 6 | 15.0 | 22 |
| 90, 10 | 6 | <u>31.3</u> | 67 | 0 | 26.4 | 57 | 0 | 25.4 | 54 |
| 100,10 | 14 | <u>24.1</u> | 32 | 11 | 15.3 | 20 | 7 | 13.7 | 21 |
| 50, 15 | 19 | <u>34.2</u> | 54 | 4 | 7.7 | 13 | 0 | 7.1 | 13 |
| 60,15 | 16 | <u>31.1</u> | 47 | 5 | 9.3 | 16 | 5 | 8.8 | 13 |
| 70, 15 | 12 | <u>28.8</u> | 50 | 5 | 10.2 | 17 | 4 | 9.3 | 17 |
| 80, 15 | 17 | <u>29.1</u> | 42 | 6 | 10.7 | 16 | 5 | 10.0 | 16 |
| 90, 15 | 20 | <u>28.5</u> | 37 | 8 | 11.2 | 16 | 5 | 10.2 | 15 |
| 100,15 | 17 | <u>24.7</u> | 36 | 8 | 12.3 | 17 | 7 | 11.2 | 16 |

Tab. I

Nota: Em cada linha, o número sublinhado representa o pior "gap" médio e o **sombreado** o melhor.

Da tabela I pode observar-se que as soluções obtidas pelo método de Frieze pioram com o aumento das dimensões do problema, quer em número de cidades,

quer em número de caixeiros. Qualquer uma das variantes do método heurístico fornece melhores soluções quando se aumenta o número de caixeiros, mantendo o de cidades. Tal não é de estranhar dado o algoritmo global acentar na determinação de soluções admissíveis para o TSP em tantos subgrafos como o número de caixeiros, reduzindo-se assim, em princípio, a dimensão de cada TSP.

Da comparação dos métodos, pode observar-se que o de *Frieze*, embora seja o que tem melhor análise de pior caso, é o que pior se comporta na prática. O método baseado no de inserção de menor custo é o que fornece melhores "gap's" médios, nunca atingindo o pior valor médio. Em termos dos "gap's" mínimos e máximos, embora não sejam comparáveis entre si as duas variantes da heurística são, em geral, melhores que o método de *Frieze*.

Analisando os valores dos "gap's" encontrados pelas variantes da heurística, e tendo em conta que foram calculados relativamente ao valor do minorante, pode dizer-se que não são muito elevados.

Tempos de CPU (s)

| Dim. n, m | Frieze | | | Fusão | | | Inserção de Menor Custo | | |
|--------------|--------|-------|-------|-------|-------|-------|----------------------------|------|-------|
| | Min | Méd | Máx | Min | Méd | Máx | Min | Méd | Máx |
| 30, 2 | 0.82 | 1.25 | 1.71 | 0.99 | 1.35 | 1.82 | 0.65 | 0.94 | 1.16 |
| 40, 2 | 1.31 | 1.93 | 3.35 | 1.70 | 2.28 | 3.24 | 1.25 | 1.63 | 2.25 |
| 30, 5 | 0.77 | 1.31 | 1.86 | 0.93 | 1.12 | 1.48 | 0.60 | 0.80 | 0.93 |
| 40, 5 | 1.31 | 2.20 | 4.17 | 1.53 | 1.94 | 2.58 | 0.99 | 1.39 | 1.71 |
| 50, 5 | 2.64 | 3.78 | 5.00 | 2.26 | 3.30 | 5.55 | 1.65 | 2.44 | 4.07 |
| 60, 5 | 3.78 | 4.65 | 6.05 | 3.52 | 4.39 | 5.65 | 2.64 | 3.17 | 3.79 |
| 50, 10 | 2.53 | 4.57 | 7.44 | 1.92 | 2.47 | 3.28 | 1.48 | 1.80 | 2.37 |
| 60, 10 | 4.47 | 6.44 | 9.46 | 2.63 | 3.88 | 5.61 | 2.09 | 2.77 | 3.46 |
| 70, 10 | 6.48 | 8.94 | 12.81 | 3.96 | 5.02 | 6.81 | 2.96 | 3.65 | 4.50 |
| 80, 10 | 6.59 | 10.57 | 16.47 | 4.66 | 6.72 | 8.53 | 3.74 | 4.97 | 5.73 |
| 90, 10 | 9.46 | 14.55 | 20.22 | 5.50 | 8.37 | 11.36 | 4.57 | 6.16 | 7.64 |
| 100, 10 | 12.58 | 18.15 | 30.37 | 7.25 | 11.20 | 17.30 | 5.87 | 7.91 | 10.61 |
| 50, 15 | 3.52 | 5.85 | 10.38 | 1.66 | 2.38 | 3.68 | 1.27 | 1.69 | 2.15 |
| 60, 15 | 5.17 | 8.55 | 14.29 | 2.64 | 3.48 | 4.73 | 2.04 | 2.52 | 3.14 |
| 70, 15 | 6.16 | 10.53 | 17.02 | 3.37 | 4.80 | 6.81 | 2.46 | 3.49 | 4.51 |
| 80, 15 | 7.75 | 13.59 | 23.89 | 4.88 | 6.79 | 10.00 | 3.63 | 4.65 | 6.14 |
| 90, 15 | 10.66 | 16.56 | 22.90 | 5.28 | 8.78 | 12.25 | 4.28 | 6.00 | 7.37 |
| 100, 15 | 15.35 | 21.21 | 28.29 | 7.52 | 9.95 | 11.98 | 5.71 | 7.05 | 8.29 |

Tab. II

Nota: Em cada linha, o número sublinhado representa o pior tempo de execução médio e o **sombreado** o melhor.

Relativamente aos tempos de execução (Tab. II) gastos o método de *Frieze* é, em geral, substancialmente mais pesado que os restantes, sendo significativamente elevado para os problemas de maiores dimensões. O método baseado no de inserção de menor custo é, como seria de esperar, o mais rápido. Tanto este método como o de fusão não se tornam demasiado pesados com o aumento das dimensões.

5. CONCLUSÕES

Com base na análise dos resultados feita no ponto anterior pode inferir-se que o método heurístico de *Frieze* [3], embora seja o que tem melhor análise de pior caso, em problemas simétricos e em que é válida a desigualdade triangular, caso em que se englobam os problemas testados, não conduz em geral a bons resultados práticos. Trata-se portanto de um método com interesse essencialmente teórico.

O método heurístico apresentado parece proporcionar melhores soluções admissíveis, embora teoricamente seja pior que o de *Frieze*.

Relativamente às variantes testadas, embora não sejam comparáveis entre si com base nos valores dos "gap's", parece poder afirmar-se que a de inserção de menor custo é melhor, pois gastando menos tempo de CPU, fornece sempre melhores "gap's" médios. Pode ainda observar-se que a diferença entre os valores das soluções admissíveis obtidas não é significativa sempre que o valor do majorante obtido por este método é superior ao do encontrado com base no método de fusão.

Como se referiu, o método em causa tem por base a determinação de uma *m-SST*. O raciocínio que presidiu ao desenvolvimento deste método sugeriu a ideia de partir de uma *2m-SST*, tentando assim obter uma solução admissível para o *m-TSP* com base em subgrafos diferentes dos considerados neste artigo.

Tal como já foi feito relativamente a alguns métodos heurísticos as autoras estão ainda a tentar generalizar para o *m-TSP* um método construtivo para o *TSP* - o método de Fusão - esperando, em breve, poder apresentar resultados sobre o seu desempenho computacional.

Agradecimentos

Foi muito valiosa a leitura do presente artigo pela Prof. Doutora Maria Teresa Almeida, assim como muito construtivas as suas críticas e sugestões. As autoras agradecem-lhe a participação chamando a si a exclusiva responsabilidade de eventuais erros.

Referências Bibliográficas

- [1] BELLMORE, M.; HONG, S. - 1974 - "*Transformation of Multisalesmen Problem to the Standard Traveling Salesman Problem*"; Journal of the Association for Computing Machinery, Vol. 21, pg. 500 - 504;
- [2] CHRISTOFIDES, N. - 1979 - "*The Travelling Salesman Problem*"; in Combinatorial Optimization, edited by Christofides, N.; Mingozzi, A.; Toth, P.; Sandi, C.; John Wiley & Sons;
- [3] FRIEZE, A.M. - 1983 - "*An Extension of Christofides Heuristic to the k-Person Travelling Salesman Problem*"; Discrete Applied Mathematics, Vol.6, pg. 79 - 83;
- [4] FREDERICKSON, G.N.; HETCH, M.S.; KIM, C.E. - 1978 - "*Approximation Algorithms for some Routing Problems*"; Siam Journal on Computing, Vol. 7, pg. 178 - 193;
- [5] GABOW, H.N. - 1978 - "*A Good Algorithm for Smallest Spanning Trees with a Degree Constraint*"; Networks, Vol. 8, pg. 201 - 208;
- [6] GAVISH, B.; SRIKANTH, K. - 1986 - "*An Optimal Solution Method for Large-Scale Multiple Traveling Salesman Problems*"; Operations Research, Vol. 34 (5), pg. 698 - 717;
- [7] HONG, S.; PÅDBERG, M.W. - 1977 - "*A Note on the Symmetric Multiple Traveling Salesman Problem with Fixed Charges*"; Operations Research, Vol. 25, pg. 871 - 874;

- [8] JONKER, R.; VOLGENANT, T. - 1988 - "*An Improved Transformation of the Symmetric Multiple Traveling Salesman Problem*"; *Operations Research*, Vol. 36 (1), pg. 163 - 167;
- [9] JOHNSON, D.S.; PAPADIMITIOU, C.H. - 1985 - "*Computational Complexity*"; in *The Traveling Salesman Problem - a guided tour of combinatorial optimization*, edited by Lawler, E.L.; Lenstra, J.K.; Rinnoy Kan, A.H.G.; Shmoys, D.B.; John Wiley & Sons;
- [10] LENSTRA, J.K.; RINNOOY KAN, A.H.G. - 1976 - "*On General Routing Problems*"; *Networks*, Vol. 6, pg. 273 - 280;
- [11] LENSTRA, J.K.; RINNOOY KAN, A.H.G. - 1979 - "*A Characterization of Linear Admissible Transformations for the m-Travelling Salesmen Problem: A Result of Berenguer*"; *European Journal of Operational Research*, Vol. 3, pg. 250 - 252;
- [12] ORLOFF, C.S. - 1974 - "*Routing a Fleet of M Vehicles to/from a Central Facility*"; *Networks*, Vol. 4, pg. 147 - 162;
- [13] RAO, M.R. - 1980 - "*A Note on the Multiple Traveling Salesman Problem*"; *Operations Research*, Vol. 28, pg. 628 - 632;
- [14] ROSENKRANTZ, D.; SASATEARNS, R.; LEWIS, P. - 1974 - "*Approximate Algorithms for the Traveling Salesman Problem*"; *Proceeding of the 15 Annual IEEE Symposium of Switching and Automata Theory*, pg.33 - 42.

HEURÍSTICA PARA O SEQUENCIAMENTO DE LOTES ECONÓMICOS DE n PRODUTOS EM m MÁQUINAS IDÊNTICAS

José Fernando Gonçalves
Manuel Pina Marques
Gabinete de Gestão e Engenharia Industrial
DEMEGI – FEUP

Resumo

Nesta comunicação apresenta-se uma heurística que permite sequenciar a produção de n produtos em m máquinas idênticas. Em primeiro lugar trata-se este problema para o caso da procura ser determinística, fazendo-se depois a extensão do trabalho para o caso da procura estocástica.

A heurística ajusta dinamicamente o tamanho dos lotes e faz a afectação dos vários produtos às m máquinas, de forma a minimizar a soma dos custos de preparação e posse.

Finalmente são apresentados os resultados obtidos por simulação da aplicação da heurística a dois problemas diferentes.

1. INTRODUÇÃO

No problema em estudo cada um dos n produtos é processado numa só operação. Esta pode ser realizada em qualquer uma das m máquinas idênticas existentes. Pretende-se definir em que máquina ou máquinas cada um dos produtos deve ser processado e, para cada uma delas, o sequenciamento a utilizar, de forma a minimizar a soma dos custos de preparação e posse.

Outros autores já trataram várias versões deste problema, as quais contemplam na sua maioria casos mais restritivos, já que consideram uma só máquina e/ou a procura como sendo determinística. Alguns desses trabalhos podem ser encontrados nas referências indicadas no final deste artigo.

A heurística que vai ser apresentada pode ser subdividida em duas partes distintas. A primeira, que iremos designar por **Parte 1**, determina, para cada produto, se se deve ter uma ou mais máquinas exclusivamente dedicadas à sua produção (eliminando assim os custos de preparação e de posse), ou se esta se deve fazer por lotes. Neste último caso, determina-se o tamanho do lote ideal para a sua produção. Na segunda parte, **Parte 2**, procura-se encontrar um sequenciamento exequível para os produtos que são produzidos por lotes. Para esse efeito ajusta-se dinamicamente não só a afectação dos produtos às máquinas, mas também os tamanhos dos lotes e os inícios de fabrico em cada máquina. A heurística baseia-se, no que diz respeito ao ajuste do tamanho dos lotes e inícios de fabrico, na heurística de Leachman e Gascon (1988).

Embora o sistema de produção deva ser objecto de um acompanhamento contínuo, este é frequentemente impraticável. Assim, será utilizada a revisão periódica, com um período que dependerá das características do processo produtivo, tempos de preparação das máquinas, tamanho dos lotes, taxas de produção, etc.

2. NOTAÇÃO

Apresenta-se de seguida a simbologia que será utilizada ao longo deste trabalho.

- i – Índice do produto, $i = 1, \dots, n$.
Os produtos são reordenados sempre que a heurística é aplicada.
- m – Índice da máquina, $m = 1, \dots, M$.
- i_m – Índice do i -ésimo produto afectado à máquina m .
O produto i_m é o produto que a máquina m está a produzir, ou está a ser preparada para produzir.
- t – Índice do período de tempo.
 $t = 1$ representa o próximo período; $t = t^*$ representa o último período no horizonte de planeamento.
- R – Período de revisão.
- A_i – Custo de preparação associado ao produto i .
- d_i^t – Previsão, disponível no início do período t , da procura do produto i no período t .
- H_i – Custo de posse do produto i por unidade e por período.
- P_i – Taxa de produção do produto i .
- C_i – Tempo de preparação para o produto i .
- cl_i – Tempo necessário para completar a preparação da máquina para a produção do produto i .
- I_i – Quantidade em stock do produto i no início do período de revisão corrente.
- σ_i^1 – Desvio padrão do erro da previsão da procura do produto i no primeiro período com procura estocástica.

3. A HEURÍSTICA

A heurística é apresentada em primeiro lugar para o caso da procura ser determinística ($d_i^t = D_i$ e $\sigma_i^1 = 0$), sendo feito posteriormente o relaxamento deste pressuposto.

3.1. Heurística - Parte 1

O Procedimento 1 reserva máquinas para a produção dedicada dos produtos cujos tempos de produção excedam a capacidade de uma só máquina. Assim, reserva-se M_i^0 máquinas para a produção dedicada de cada produto i , sendo M_i^0 o maior inteiro menor que D_i/P_i . Ao mesmo tempo modificamos a sua procura para $D_i' = D_i - M_i^0 P_i$ (isto é, D_i' representará a procura do produto i não satisfeita pelas M_i^0 máquinas dedicadas à sua produção).

Seja $M^r = M - \sum_{i \in I} M_i^0$ o número de máquinas que ainda têm a totalidade da sua

capacidade livre. Se $M^r < 0$ então a capacidade de produção não é suficiente para satisfazer a procura.

Depois de termos completado o *Procedimento 1*, o nosso problema consiste em determinar a forma de satisfazer a procura restante de cada produto i , em que $D_i' < P_i$ a partir das M_i máquinas ainda não utilizadas. O *Procedimento 2* resolverá este problema.

Vamos verificar se a procura restante de algum dos produtos deve ocupar só por si a totalidade da capacidade de uma máquina (evitando desse modo os custos de posse e de preparação associados a esse produto). Para os produtos que forem produzidos em conjunto com outros na mesma máquina, calculamos os respectivos lotes económicos de produção Q_i^* . Caso seja possível produzir mais do que um produto sozinho, deverão escolher aqueles que conduzam a uma maior redução dos custos.

O *Procedimento 2* consiste na resolução de um problema de programação inteira não linear com a seguinte forma:

Seja $y_i = \begin{cases} 1 & \text{se o produto é produzido sózinho numa máquina} \\ 0 & \text{caso contrário} \end{cases}$

$$\text{Minimizar: } Z = \sum_{i=1}^n (1 - y_i) \left[\frac{A_i D_i'}{Q_i} + \frac{1}{2} H_i \left(1 - \frac{D_i'}{P_i} \right) Q_i \right]$$

$$\text{Sujeito a: } \sum_{i=1}^n (1 - y_i) \left[\frac{C_i D_i'}{Q_i} + \frac{D_i'}{P_i} \right] \leq M^r - \sum_{i=1}^n y_i$$

Este problema pode ser resolvido através da técnica do *Branch and Bound*. A cada nó da árvore de pesquisa corresponderá uma combinação de valores das variáveis inteiras y_i . Para cada uma das combinações temos de determinar os respectivos lotes económicos. Para esse efeito calculamos em primeiro lugar a solução óptima do problema sem restrição (EMQ_i), dada por

$$EMQ_i = \left[\frac{2A_i D_i'}{H_i \left(1 - \frac{D_i'}{P_i} \right)} \right]^{1/2} \tag{1}$$

Seguidamente verificamos se os valores encontrados para os vários EMQ_i satisfazem a restrição de capacidade. Se isso acontecer, então esses valores são óptimos (para a combinação de valores y_i fixados); caso contrário, a solução óptima é obtida quando a restrição de capacidade é activa. Para este último caso podemos obter os valores óptimos utilizando a função *Lagrangiana* correspondente

$$Q_i^* = EMQ_i^2 + \lambda^* = \left[\frac{2C_i D_i'}{H_i \left[1 - \frac{D_i'}{P_i} \right]} \right]^{1/2} \tag{2a}$$

em que λ^* representa o valor de λ que satisfaz a equação

$$\sum_{i=1}^n (1-y_i) \frac{C_i D_i'}{\left[EMQ_i^2 + \lambda \frac{2C_i D_i'}{H_i \left(1 - \frac{D_i'}{P_i}\right)} \right]} = M^r - \sum_{i=1}^n y_i - \sum_{i=1}^n (1-y_i) \frac{D_i'}{P_i} \quad (2b)$$

O valor de λ^* pode ser obtido utilizando um método iterativo como, por exemplo, a pesquisa binária.

Seja $I^0 = \{i \mid y_i = 0\}$ o conjunto dos produtos que não têm nenhuma máquina exclusivamente dedicada à sua produção, e $I^1 = \{i \mid y_i = 1\}$ o conjunto de todos os outros produtos. Seja ainda Q_i^* o lote óptimo para cada produto $i \in I^0$ e defina-se M_i da seguinte forma

$$M_i = \begin{cases} M_i^0 + 1 & \text{se } i \in I^1 \\ M_i^0 & \text{se } i \in I^0 \end{cases} \quad (3)$$

Então, cada produto $i \in I^1$ terá M_i máquinas exclusivamente dedicadas à sua produção, produzindo em cada período a quantidade necessária para satisfazer a procura. Os produtos $i \in I^0$ serão produzidos não só nas M_i máquinas exclusivamente dedicadas à sua produção, mas também em lotes nas restantes $M^r - \sum_{i \in I^1} y_i$ máquinas. O nível de enchimento ideal para a produção do produto i é dado por

$$S_i^* = \left(1 - \frac{D_i'}{P_i}\right) Q_i^*$$

A Parte 2 da heurística explica como se deve fazer a programação da produção por lotes de todos os produtos $i \in I^0$ de forma a que os níveis de enchimento ideais S_i^* sejam tanto quanto possível atingidos.

3.2. Heurística - Parte 2

A segunda parte da heurística calcula valores escalados de S_i^* que dão origem a sequências de produção exequíveis. Determina igualmente em qual das $M^r - \sum_{i=1}^n (1-y_i)$

máquinas partilhadas se deve produzir cada produto $i \in I^0$, bem como quando a sua produção deve ter início e fim.

Admita-se que a procura dos produtos $i = 1_m, 2_m, \dots, n_m$ é afectada à máquina m . Conforme já referido, designamos por 1_m o produto que a máquina m está a produzir, ou está a ser preparada para produzir.

A produção seguirá aquilo a que designaremos por *níveis de enchimento operacionais* $S_{1_m}^*$, que são valores escalados dos níveis de enchimento ideais $S_{1_m}^*$, já fixados na Parte 1. O factor de escala para os produtos afectados à máquina m é designado por α_m , e o seu valor será deduzido mais tarde.

O nível de enchimento operacional do produto 1_m é dado por

$$S_{1_m} = ss_{1_m} + Q_{1_m}^* \left(1 - \frac{D'_{1_m}}{P_{1_m}}\right) \quad (4)$$

O parâmetro ss_{1_m} representa um factor associado ao stock de segurança, e será discutido mais tarde. Para o caso determinístico $ss_{1_m} = 0$.

No caso de existir um valor de m para o qual $I_{1_m} \leq S_{1_m}$ então a produção do produto 1_m deve terminar, sendo renumerados os índices dos produtos afectados à máquina m de forma a que o índice 1_m represente o produto afectado à máquina m com o menor valor esperado do tempo até à ruptura de stock.

O tempo restante de utilização da máquina m pelo produto 1_m , RUT_{1_m} , engloba o tempo que falta para completar a preparação da máquina e o tempo necessário para aumentar a existência em stock I_{1_m} até ao seu nível de enchimento S_{1_m} . O Tempo Restante de Utilização é então calculado por (ver figura 1)

$$RUT_{1_m} = \max \left\{ 0, \frac{S_{1_m} - I_{1_m} + D'_{1_m} c_{1_m}}{P_{1_m} - D'_{1_m}} + c_{1_m} \right\} \quad (5)$$

Definimos valor esperado do tempo até à ruptura de stock do produto i , RO_i , da seguinte forma (ver figuras 2a e 2b)

$$RO_i = \begin{cases} RUT_{1_m} + \frac{S_{1_m} - ss_{1_m}}{D'_{1_m}} - C_{1_m} & \text{se } i=1_m \text{ e } I_{1_m} < S_{1_m} \\ \frac{I_i - ss_i}{D'_{1_m}} - C_i & \text{para todos os outros casos} \end{cases} \quad (6)$$

Os índices dos n artigos afectados à máquina m serão numerados $2_m, 3_m, \dots, (n+1)_m$ por ordem crescente dos seus tempos esperados até à ruptura de stock $RO_{2_m} \leq RO_{3_m} \leq \dots \leq RO_{(n+1)_m}$.

As rupturas de stock dos produtos $2_m, 3_m, \dots, (n+1)_m$, que vão ser produzidos na máquina m , estarão adequadamente espaçadas ao longo do tempo de forma a permitir a produção até aos respectivos níveis de enchimento, se verificarem

$$RO_{2_m} \geq RUT_{1_m} \quad (7)$$

$$RO_{i_m} \geq RO_{i-1_m} + C_{i-1_m} + \frac{Q_{i-1_m}^*}{P_{i-1_m}} \alpha_m \quad i = 3, 4, \dots, (n+1).$$

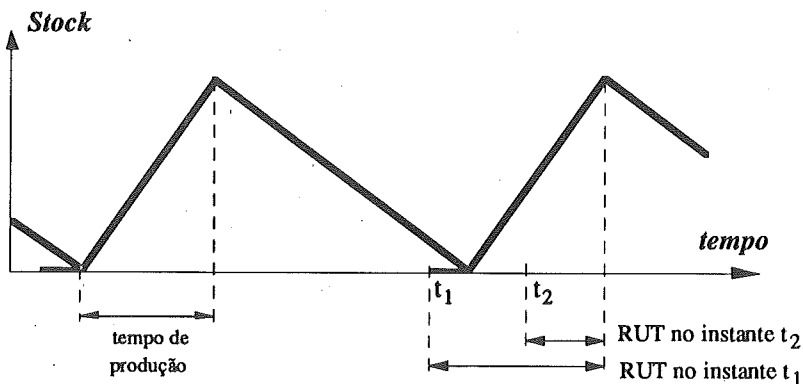


Figura 1 - Tempo Restante de Utilização (RUT)

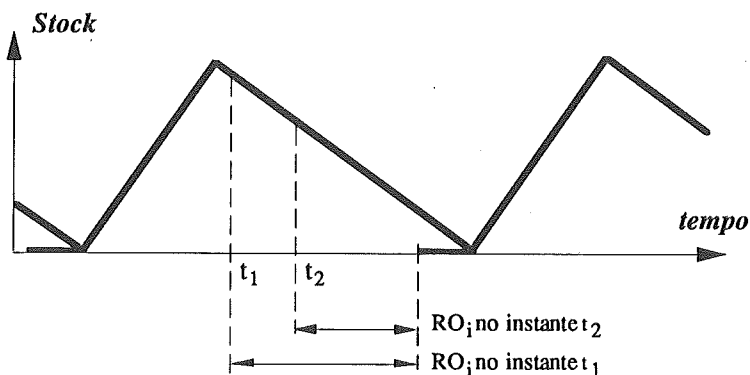


Figura 2a - Tempo Previsto até à Ruptura do produto i quando este não se encontra em produção.

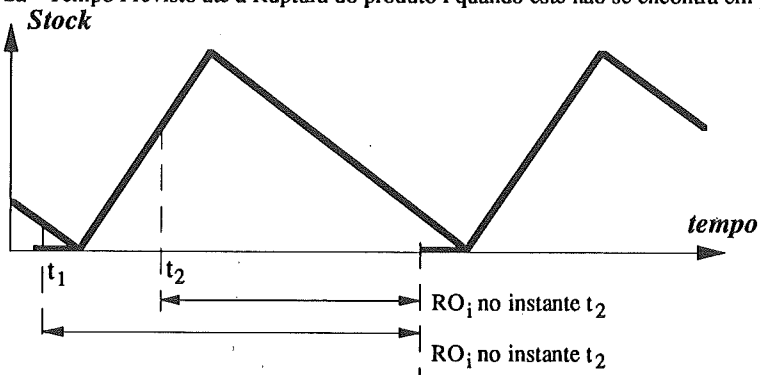


Figura 2b - Tempo Previsto até à Ruptura do produto i quando este se encontra em produção.

Um produto i_m tem *folga positiva* se a última desigualdade se verificar; caso contrário diz-se que o produto tem *folga negativa*.

A folga positiva de um produto i_m pode ser utilizada para compensar a folga negativa de qualquer outro produto j_m (com $j_m > i_m$). Para tal, basta antecipar o início da produção dos lotes dos produtos $i_m, (i + 1)_m, \dots, (j - 1)_m$. É pois importante conhecer para cada produto a *folga acumulada* relativamente aos tempos esperados até à ruptura de stock. A folga acumulada TS_{i_m} para um produto i_m é calculada, para um determinado valor α_m por

$$TS_{i_m}(\alpha_m) = \begin{cases} RO_{2_m} - RUT_{1_m} & \text{para } i = 2 \\ RO_{i_m} - \sum_{i=2_m}^{i=(i-1)_m} C_i - \alpha_m \sum_{i=2_m}^{i=(i-1)_m} \frac{Q_i}{P_i} - RUT_{1_m} & \text{para } i \geq 3 \end{cases} \quad (8)$$

Caso a folga acumulada seja negativa para algum produto i_m , isto é, $TS_{i_m} < 0$, quando $\alpha_m = 1$, isso significa que os níveis de enchimento desejados não serão exequíveis. A fim de tornar realizável o programa de fabrico vamos, em vez de alterar os níveis de enchimento desejados de um ou de alguns produtos i_m , fixar o factor de escala α_m no maior valor ≤ 1 de forma a que $TS_{i_m} \geq 0$ para todos os produtos i_m . O *factor de escala eficiente* α_m^* é dado por

$$\alpha_m^* = \min \left\{ 1, \left\{ \max \alpha_m \mid TS_{i_m}(\alpha_m) \geq 0 \text{ para todos } i_m \right\} \right\} \quad (9)$$

Caso seja inevitável haver uma ruptura, α_m^* pode tomar valores demasiado pequenos ou mesmo negativos. A fim de se evitarem estes casos, o utilizador define um parâmetro α_0 que será um limite inferior para α_m . A expressão (9) pode ser então reescrita da seguinte forma

$$\alpha_m^* = \max \left\{ \alpha_0, \min \left\{ 1, \left\{ \max \alpha_m \mid TS_{i_m}(\alpha_m) \geq 0 \text{ para todos } i_m \right\} \right\} \right\} \quad (10)$$

O Procedimento de Afectação

O procedimento de afectação (*Procedimento 3*) afecta todos os produtos $i \in I^0$ às máquinas não dedicadas exclusivamente à produção de um produto, e determina um conjunto de valores de α_m que tornam a folga acumulada de cada um dos produtos i_m positiva ou, caso tal não seja possível, o menos negativa possível. De notar que, para os produtos que estão a ser inicialmente fabricados em cada uma das máquinas, será fixada uma máquina para a sua futura produção, mas não para a produção corrente; isto é, não haverá interrupção da produção do lote correntemente em fabrico.

Os produtos são afectados às máquinas um de cada vez e por ordem crescente dos respectivos tempos esperados até à ruptura de stock, RO_i . O primeiro produto afectado à máquina m terá associado o índice 2_m , o segundo 3_m , etc. (O índice 1_m está associado ao produto para o qual a máquina m está a produzir, ou para cujo fabrico está a ser preparada).

Suponhamos que os produtos $i = 1, 2, \dots, i - 1$ já foram afectados a uma determinada máquina m , e que queremos agora seleccionar uma máquina para a produção do produto i . Para esse efeito, vamos, para cada máquina m , calcular através de (8) a folga acumulada desse produto i , $TS_i(\alpha_m^*)$, utilizando o valor actual de α_m^* .

Seja M^+ o conjunto definido pelas máquinas que, caso fossem escolhidas para afectar o produto i , davam origem a folgas acumuladas positivas

$$M^+ = \{m \mid TS_i(\alpha_m^*) \geq 0\}$$

A regra para a afectação depende de M^+ , como se passa a explicar.

Se $M^+ \neq \emptyset$ - o produto i será afectado à máquina $m \in M^+$ que minimiza a diferença entre a folga acumulada do produto que lhe foi afectado em último lugar, $TS_{n_m}(\alpha_m^*)$, e a folga acumulada do produto i , caso este lhe seja afectado, $TS_i(\alpha_m^*)$. Por outras palavras, o produto i será afectado à máquina para a qual será necessária a menor antecipação do início da produção dos produtos que o precedem.

Se $M^+ = \emptyset$ - calcula-se α_m^* de acordo com (10) para todas as máquinas m , assumindo que o produto i lhe foi afectado. O produto i será afectado à máquina que permita obter a maior folga acumulada, $TS_i(\alpha_m^*)$. No caso de haver duas ou mais máquinas que conduzam à mesma folga, deverá ser escolhida a que tiver associado o maior valor de α_m^* . O factor de escala da máquina m à qual o produto i for efectivamente afectado deve ser actualizado, sendo os restantes α_m^* s repostos nos valores que tinham antes deste passo.

Após termos completado o procedimento de afectação dos produtos às máquinas, devemos verificar se $I_{1_m} \geq S_{1_m}(\alpha_m)$ para todas as máquinas m , por ordem decrescente de α_m . Seja m^1 a primeira máquina para a qual esta desigualdade se verifica. Então, devemos interromper a produção do produto 1_{m^1} e reiniciar o procedimento de afectação. Caso não se encontre nenhuma máquina m^1 , a afectação estará concluída.

Definiu-se em que máquina cada produto deve ser produzido, bem como quais os produtos que devem ver interrompida a sua produção. No ponto seguinte apresenta-se uma política de programação que vai determinar, para cada máquina, se esta deve estar parada um período, continuar com a produção do produto para o qual está configurada, ou ser preparada para o fabrico de outro produto e iniciar a sua produção até o próximo período de revisão.

Política de Programação

Caso a produção do produto 1_m tenha terminado, pode-se manter parada a máquina m se as folgas acumuladas $TS_{i_m}(\alpha_m^*)$ forem maiores do que um período de revisão; isto é, se

$$TS_{i_m}(\alpha_m^*) \geq R \quad \text{para } i \geq 2 \quad (11)$$

calculando-se TS_{i_m} pela expressão (8).

Pelo contrário, se $TS_{i_m}(\alpha_m^*) < R$ para algum produto i_m , dever-se-à começar a preparar a máquina m para produzir o produto 2_m . A produção do item 2_m deve elevar o stock I_{2_m} até ao nível de enchimento S_{2_m} antes de se preparar a máquina para a produção do produto 3_m , e assim sucessivamente até o próximo período de revisão.

Se a máquina ainda está a ser preparada para produzir o produto 1_m , ou se este está a ser produzido, parar a máquina um período de tempo implicaria que demoraria mais

do que RUT_{1_m} , dado em (5), para o stock de 1_m atingir o nível S_{1_m} . Neste caso a máquina pode estar parada um período se

$$\frac{I_{1_m} - SS_{1_m}}{D_{1_m}} \geq R \tag{12a}$$

e se

$$TS_{i_m}(\alpha_m^*) \geq R + \frac{D_{1_m}' R}{P_{i_m} - D_{1_m}'} \quad \text{para } i \geq 2 \tag{12b}$$

sendo TS_{i_m} dado por (8). No caso da máquina não poder estar parada um período, deve-se continuar a produzir o produto 1_m até o seu stock I_{1_m} atingir o nível de enchimento S_{1_m} , após o que se deve preparar a máquina para a produção do produto 2_m , e assim sucessivamente, seguindo os níveis de enchimento operacionais até ao próximo período de revisão.

3.3. Heurística - Procura estocástica

Até agora a heurística baseou-se nos valores médios da procura D_i . Na prática, no entanto, estes valores são desconhecidos, sendo fixados em termos de previsão da procura.

No caso da *procura estocástica ser dependente do tempo* podemos caracterizar a procura pela média da procura prevista para os próximos t_u períodos (13), e pelo desvio padrão associado aos erros de previsão para o período seguinte, σ_i^1 .

$$D_i = \frac{1}{t_1} \sum_{t=1}^{t=t_u} d_i^t \quad i = 1, 2, \dots, n \tag{13}$$

Silver e Peterson (1985) mostraram que, para a maior parte dos sistemas de stocks, o desvio padrão dos erros de previsão para L períodos pode ser aproximado por

$$\sigma^L = L^c \sigma^1 \tag{14}$$

em que σ^1 representa o desvio padrão associado à previsão da procura para o período imediatamente seguinte, e c é uma constante estimada empiricamente. Assume-se que os erros de previsão de cada produto i são variáveis aleatórias independentes, identicamente distribuídas para todos os períodos, seguindo uma distribuição normal $N(0, \sigma_i^1)$. Nessas condições teremos $c = 1/2$.

Dada a incerteza associada às previsões que lhes estão na origem, os valores de D_i dados em (13) são aproximados. A heurística desenvolvida deve ser modificada a fim de ter tal facto em consideração.

O *Procedimento 1* tem de ser alterado de forma a que a carga a afectar às máquinas dedicadas exclusivamente à produção de um artigo englobem uma determinada incerteza associada à procura, para além da sua média prevista. A procura X afectada a cada máquina dedicada deve então satisfazer a igualdade

$$X + Z_D \frac{X}{D_i} \sigma_i^1 = P_i$$

em que Z_D é um parâmetro escolhido em função do nível de serviço desejado. Isto é equivalente a utilizar o *Procedimento 1* descrito anteriormente, utilizando os valores D_i e P_i^1 , tal que

$$P_i' = \frac{P_i}{1 + \frac{Z_D \sigma_i^1}{D_i}} \quad (15)$$

A capacidade extra afectada a cada máquina ao aplicar o *Procedimento 1* com uma taxa de produção P_i' será utilizada na criação de um stock que absorva a incerteza da procura.

No *Procedimento 2* iremos utilizar para cada produto i uma procura média D_i' definida por

$$D_i' = D_i - M_i^0 P_i'$$

em que P_i' é calculado por (15). Neste caso, em vez de diminuirmos a capacidade disponível, como fizemos no *Procedimento 1*, aumentamos o tamanho dos lotes óptimos (aumentando assim os níveis de enchimento desejados) para todos os produtos $i \in I^0$. O tamanho dos lotes será aumentado de uma quantidade proporcional à incerteza das previsões da procura no intervalo de tempo compreendido entre produções. Os lotes serão alterados para

$$Q_i^* = Q_i + Z_Q \left[\frac{Q_i'}{D_i'} \left(1 - \frac{D_i'}{P_i'} \right) \right]^{1/2} \frac{D_i'}{D_i} \sigma_i^1 \quad (16)$$

em que Z_Q representa um parâmetro fixado em função do nível de serviço desejado e Q_i' é o valor do lote óptimo obtido pelo *Procedimento 2*. À medida que aumentamos o tamanho dos lotes o número de preparações das máquinas vai diminuindo, aumentando assim a capacidade disponível. Consequentemente existirá uma maior flexibilidade para reagir contra a incerteza da procura.

O stock de segurança do produto i , ss_i , é dado por

$$ss_i = \begin{cases} Z_S \sigma_i^1 & \text{se } i \in I^1 \\ Z_C C_i \sigma_i^1 & \text{se } i \in I^0 \end{cases} \quad (17)$$

em que Z_C e Z_S são parâmetros fixados de acordo com o nível de serviço desejado.

Os produtos $i \in I^1$ terão M_i máquinas exclusivamente dedicadas à sua produção ($M_i - 1$ máquinas reservadas pelo *Procedimento 1* e uma máquina pelo *Procedimento 2*). Estas M_i máquinas dedicadas devem produzir até que o stock do produto i atinja o nível do stock de segurança ss_i . Nessa altura deve ser interrompida a sua produção.

A procura dos produtos $i \in I^0$ será satisfeita quer pela produção das M_i máquinas dedicadas reservadas pelo *Procedimento 1* quer pela produção em lotes das $M^T - \sum_{i \in I^1} y_i$

máquinas não dedicadas. Sendo d_i^t a procura durante o período de revisão t , as M_i máquinas dedicadas devem produzir em cada período t uma determinada quantidade P_i^t dada por

$$P_i^t = \min \left\{ \max \left\{ S_i - (I_i^{t-1} - d_i^t), 0 \right\}, M_i P_i \right\} \quad (18)$$

em que I_i^{t-1} representa o stock do produto i no final do período de revisão $t - 1$. No caso da produção das máquinas dedicadas não ser suficiente para satisfazer a procura no

período, a procura restante, $\max\{d_i^t - p_i^t, 0\}$ terá de ser satisfeita a partir de stock. Sempre que o stock de um produto $i \in I^0$ for inferior ao stock de segurança ss_i , as máquinas dedicadas e as máquinas não dedicadas serão utilizadas para elevar o stock até ao nível S_i dado por (4). A produção em lotes dos produtos $i \in I^0$ é programada na Parte 2 da heurística.

Alteração da Parte 2 da heurística, a fim de reflectir a incerteza

A Parte 2 da heurística baseia-se na estimativa dos tempos esperados até à ruptura de stock dos produtos, que são determinados a partir da previsão da procura. A fim de termos em conta a incerteza associada à estimativa dos tempos esperados até à ruptura de stock, vamos modificar o cálculo dos ciclos operacionais. Vamos fazê-lo da forma proposta por *Leachman e Gascon (1988)*.

Podemos definir por (14) um intervalo de confiança para o valor esperado do tempo até à ruptura de stock do produto i

$$RO_i + Z_R RO_i^{1/2} \frac{\sigma_i}{D_i}$$

em que Z_R é fixado em função do nível de confiança desejado.

Os tempos esperados até à ruptura de stock são utilizados na Parte 2 da heurística na afectação dos produtos às máquinas e na determinação dos ciclos operacionais. São igualmente utilizados quando se pretende decidir sobre a eventual paragem da máquina durante um período.

No primeiro caso, isto é, no procedimento de afectação, pode acontecer que os ciclos desejados sejam exequíveis mesmo que $TS_i \leq 0$ para algum produto i com um valor esperado do tempo até à ruptura, RO_i , bastante grande. Valores da procura superiores às previsões podem temporariamente tornar impraticáveis os ciclos desejados. No entanto, valores inferiores voltarão a tornar exequíveis os ciclos desejados antes que RO_i atinja o valor zero. Dado que o valor RO_i é bastante grande, existem muitas oportunidades de reduzir os ciclos operacionais antes de RO_i atingir o valor zero. Como consequência, podemos ser insensíveis às folgas negativas devidas a produtos com grandes valores esperados de tempos até à ruptura de stock. Pelo contrário, devemos dar particular atenção às folgas negativas devidas a produtos com pequenos tempos até à ruptura de stock. Para traduzir este raciocínio na prática, propomos que se use para o cálculo dos valores de TS_i (no procedimento de afectação) o valor de RO_i^+ em vez de RO_i

$$RO_i^+ = \begin{cases} RO_i & \text{se } RO_i \leq 1 \\ RO_i + Z_R RO_i^{1/2} \frac{\sigma_i}{D_i} & \text{para os outros casos} \end{cases}$$

No segundo caso, isto é, relativamente à paragem da máquina por um período, devemos também ter em conta o aumento da incerteza associado a grandes tempos esperados até à ruptura de stock. Neste caso assumimos uma atitude mais cautelosa, pois não se deseja parar agora por um período a produção de uma determinada quantidade que teremos de produzir uns períodos mais tarde. Isto porque se corre o risco de haver entretanto falta de capacidade, devido a se terem feito previsões baixas. Assim, propomos que os valores de TS_i utilizados nas expressões (11) e (12b) sejam calculados utilizando RO_i^- em vez de RO_i , sendo

$$RO_i^- = \begin{cases} RO_i & \text{se } RO_i \leq 1 \\ RO_i - Z_R RO_i^{1/2} \frac{\sigma_i^1}{D_i} & \text{para os outros casos} \end{cases}$$

4. RESULTADOS DE SIMULAÇÃO

Apresentam-se seguidamente os resultados obtidos por simulação da aplicação da heurística a dois conjuntos de dados distintos.

Ambas as simulações foram feitas tendo por base um horizonte de planeamento de 200 dias. As durações de preparação das máquinas foram fixadas em 1/2 dia, sendo o custo associado a cada preparação de 768 USD ($C_i = 0.5$ e $A_i = 768$). Considerou-se ainda que o custo de posse era de 0.00367 USD por unidade e por dia ($H_i = 0.00367$). A procura dos produtos foi gerada aleatoriamente, período a período, partindo do pressuposto de que seguia uma distribuição normal de média D_i e desvio padrão σ_i^1 . Partiu-se do princípio de que a procura era conhecida no início de cada período de tempo. Testou-se a heurística para cada conjunto de dados com três valores distintos de coeficiente de variação, $\sigma_i^1/D_i = 0$, $\sigma_i^1/D_i = 0.25$ e $\sigma_i^1/D_i = 0.5$. Foram utilizadas três séries diferentes de números aleatórios nos testes para $\sigma_i^1/D_i = 0.25$ e $\sigma_i^1/D_i = 0.5$.

O **Problema I** refere-se a 5 produtos a produzir numa única máquina. Na tabela 1 são dados os pormenores relativamente à procura, taxa de produção e stock inicial de cada um dos produtos. Como se trata de uma só máquina não é necessário aplicar o *Procedimento 1*. Pela mesma razão, no *Procedimento 2* só foi necessário determinar qual o tamanho dos lotes que minimizavam a função de custos, satisfazendo simultaneamente a restrição de capacidade. Foram utilizados os valores óptimos dados pela expressão (1), visto eles satisfazerem a restrição de capacidade. Um resumo dos resultados obtidos pela simulação são apresentados na tabela 2.

O **Problema II** refere-se a um conjunto de 13 artigos a produzir em 4 máquinas idênticas. Os dados relativos a este problema são apresentados na tabela 3. Para este caso, o *Procedimento 1* reserva uma máquina para a produção dedicada do produto 4. O *Procedimento 2* foi facilmente resolvido já que só se poderia dedicar, no máximo, outra máquina para a produção dedicada de qualquer um dos produtos (já que qualquer

combinação de 12 produtos $\sum_{i=1}^{i=12} D_i^1/P_i > 1$). Não foi reservada mais nenhuma máquina

para a produção dedicada de qualquer produto, e novamente os valores obtidos por (1) eram óptimos para a fixação dos tamanhos dos lotes. A tabela 4 apresenta um resumo dos resultados da simulação.

| i | D_i | P_i | I_i |
|---|-------|-------|-------|
| 1 | 1232 | 4000 | 6160 |
| 2 | 420 | 12000 | 12600 |
| 3 | 714 | 6000 | 3570 |
| 4 | 1190 | 10000 | 11900 |
| 5 | 1428 | 12000 | 14200 |

Tabela 1 - Procura, taxa de produção e stock inicial para o Problema I

| Coeficiente de variação | Parâmetros | | | Procura cativa % | Custo de posse | Custo de preparação | CUSTO TOTAL |
|-------------------------|----------------|----------------|----------------|------------------|----------------|---------------------|-------------|
| | Z _C | Z _O | Z _R | | | | |
| $\sigma_i/D_i = 0$ | 0.0 | 0.0 | 0.0 | 0 | 36216 | 36096 | 72311 |
| $\sigma_i/D_i = 0.25$ | 3.0 | 0.1 | 2.0 | 0 | 38613 | 37632 | 76245 |
| | | | | 0 | 38009 | 36096 | 74105 |
| $\sigma_i/D_i = 0.5$ | 3.0 | 0.1 | 2.0 | 0 | 37520 | 37632 | 75152 |
| | | | | 0 | 40523 | 37632 | 78154 |
| | | | | 0 | 39733 | 38400 | 78133 |
| | | | | 0 | 39122 | 38400 | 77552 |

Tabela 2 - Resultados da simulação para o Problema I

| i | D _i | P _i | I _i |
|----|----------------|----------------|----------------|
| 1 | 1390 | 12000 | 15000 |
| 2 | 1550 | 10000 | 20000 |
| 3 | 1300 | 8000 | 15000 |
| 4 | 16500 | 15000 | 55000 |
| 5 | 1550 | 10000 | 15000 |
| 6 | 590 | 4000 | 10000 |
| 7 | 880 | 6000 | 8000 |
| 8 | 1250 | 7000 | 20000 |
| 9 | 1530 | 10000 | 5000 |
| 10 | 850 | 5000 | 15000 |
| 11 | 1420 | 9000 | 15000 |
| 12 | 1790 | 11000 | 21000 |
| 13 | 1550 | 9800 | 16000 |

Tabela 3 - Procura, taxa de produção e stock inicial para o Problema II

| Coeficiente de variação | Parâmetros | | | | Procura cativa % | Custo de posse | Custo de preparação | CUSTO TOTAL |
|-------------------------|----------------|----------------|----------------|----------------|------------------|----------------|---------------------|-------------|
| | Z _C | Z _O | Z _R | Z _D | | | | |
| $\sigma_i/D_i = 0$ | 0.0 | 0.0 | 0.0 | 0.0 | 0 | 108618 | 102912 | 211529 |
| $\sigma_i/D_i = 0.25$ | 3.0 | 0.1 | 2.0 | 0.1 | 0 | 116552 | 109824 | 226345 |
| | | | | | 0 | 116429 | 107520 | 223948 |
| $\sigma_i/D_i = 0.5$ | 3.0 | 0.1 | 2.0 | 0.1 | 0 | 116217 | 105984 | 222200 |
| | | | | | 0 | 127234 | 110592 | 237825 |
| | | | | | 0 | 126086 | 118272 | 244357 |
| | | | | | 0 | 123726 | 121344 | 245070 |

Tabela 4 - Resultados da simulação para o Problema II

A fim de se poder avaliar o desempenho da Parte 2 da heurística, compara-se na tabela 5 o custo médio por unidade de tempo nas várias simulações com o custo para o caso determinístico obtido no *Procedimento 2* somado ao custo por unidade de tempo do stock de segurança.

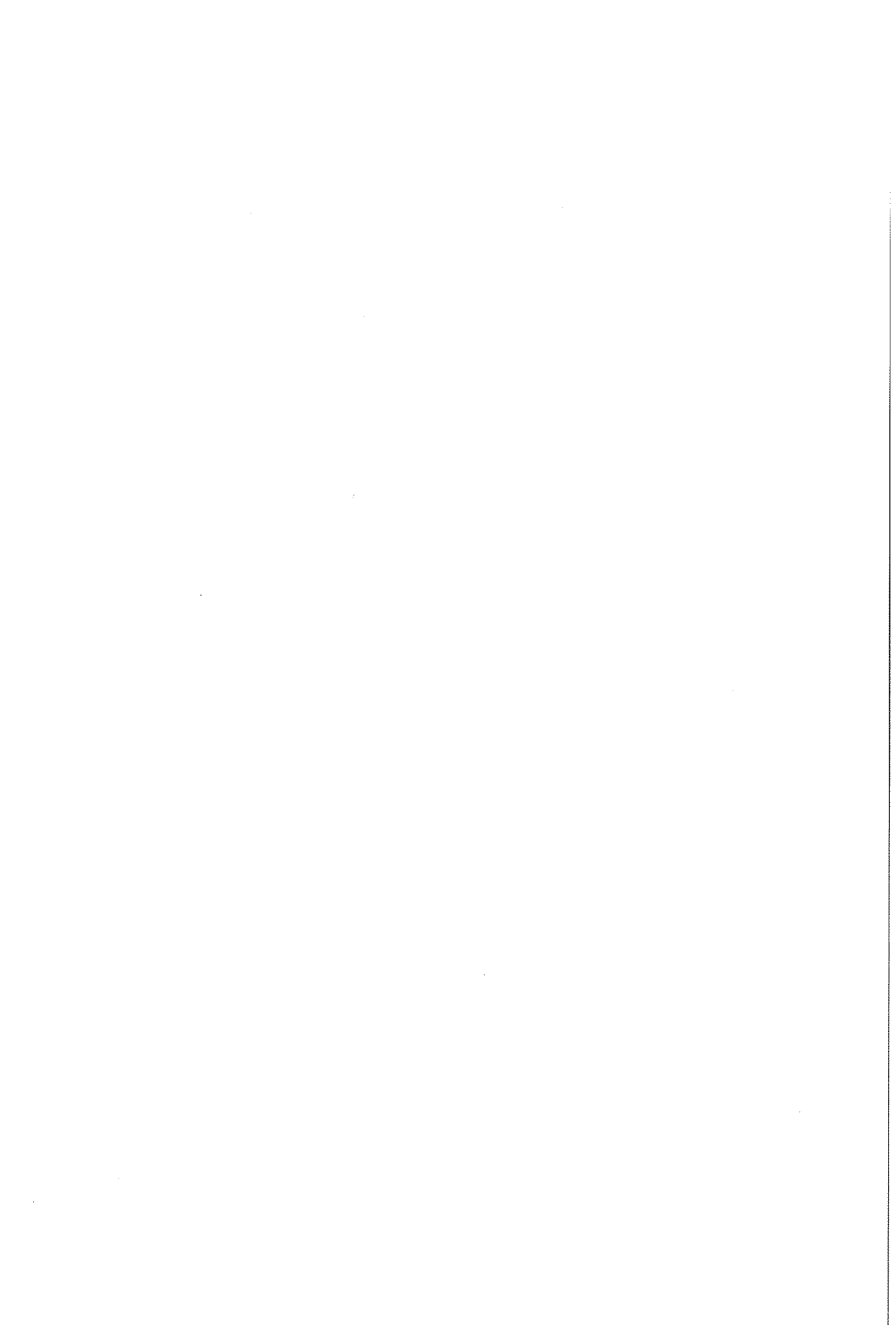
| Problema | Coefficiente de variação | Custo médio simulado por dia | Custo do caso determinístico mais custo do ss por dia | % diferença do custo |
|----------|--------------------------|------------------------------|---|----------------------|
| I | $\sigma_i^1/D_i = 0$ | 361.6 | 337.7 | 7.1 |
| | $\sigma_i^1/D_i = 0.25$ | 375.8 | 344.6 | 9.0 |
| | $\sigma_i^1/D_i = 0.5$ | 389.7 | 351.5 | 10.9 |
| II | $\sigma_i^1/D_i = 0$ | 1057.7 | 1023.4 | 3.4 |
| | $\sigma_i^1/D_i = 0.25$ | 1120.8 | 1076.0 | 4.2 |
| | $\sigma_i^1/D_i = 0.5$ | 1211.8 | 1128.6 | 7.4 |

Tabela 5 - Desempenho da Parte 2 da heurística

REFERÊNCIAS

- [1] DIXON, PAUL S. e EDWARD A.SILVER, "A Heuristic for the Multi-Item, Single-Level, Limited Capacity, Lot-Sizing Problem", *J.Oper.Management*, 2 (1982), 23-39.
- [2] DOGRAMACI, ALI, JOHN C.PANAYIOTOPOULOS e NABIL R. ADAM, "The Dynamic Lot-Sizing Problem for Multiple Items Under Limited Capacity", *AIEE Trans.*, 13 (1981), 294-303.
- [3] DOLL, C.LOREN e D.CLAY WHYBARK, "An Iterative Procedure for the Single Machine Multi-Product Lot Scheduling Problem", *Management Sci.*, 20 (1973), 50-55.
- [4] ELMAGHRABY, SALAH E., "The economic Lot Scheduling Problem (ELSP): Review and Extensions", *Management Sci.*, 24 (1978), 587-598.
- [5] GASCON, ANDRÉ e ROBERT C.LEACHMAN, "A Dynamic Programming Solution to the Dynamic, Multi-Item, Single-Machine Scheduling Problem", *Oper. Res.*, 36 (1988), 50-56.
- [6] GLASSEY, C.R., "Minimum Change-Over Scheduling of Several Products on One Machine", *Oper. Res.*, 16 (1968), 343-352.
- [7] GONÇALVES, JOSÉ F., "Continuous Time Analysis of Production Planning and Scheduling Models", *Tese de Doutorado (Ph.D.)*, Universidade da California, Berkeley, 1987.
- [8] GRAVES, STEPHEN C., "The Multi-Product Production Cycling Problem", *AIEE Trans.*, 12 (1980), 233-240.
- [9] HAESSLER, ROBERT W., "An Improved Extended Basic Period Procedure for Solving the Economic Lot Scheduling Problem", *AIEE Trans.*, 11 (1979), 336-340.
- [10] LEACHMAN, ROBERT C. e ANDRÉ GASCON, "A Heuristic Scheduling Policy for Multi-Item, Single Machine Production Systems with Time-Varying, Stochastic Demands", *Management Sci.*, 34 (1988), 377-390.
- [11] MCKINNEY, JACQUELINE, "Optimal Multi-Product Scheduling on One Machine Over Finite Horizon", *Tese de Doutorado (Ph.D.)*, Universidade de Johns Hopkins, 1980.

- [12] NEWSON, E.F.PETER, "Multi-Item Lot Size Scheduling by Heuristic. Part I. With Fixed Resources. Part II. With Variable Resources", *Management Sci.*, 21 (1975), 1186-1203.
- [13] PARK, K.S. e D.K.YUN, "A Stepwise Partial Enumeration Algorithm for the Economic Lot Scheduling Problem", *IIE Trans.*, 16 (1984), 363- 370.
- [14] SILVER, EDWARD A. e REIN PETERSON, *Decision Systems for Inventory Management and Production Planning*, Second Ed., John Wiley & Sons, New York, 1985.
- [15] VAN NUNEN, J.A.E.E. e J.WESSELS, "Multi-Item Lot Size Determination and Scheduling Under Capacity Constraints", *European J. Oper. Res.*, 2 (1978), 36-41.
- [16] VERGIN, ROGER C. e TERRY N.LEE, "Scheduling Rules for the Multiple Product Single Machine System with Stochastic Demand", *Infor*, 16 (1978), 64-73.



MODELIZAÇÃO ESTOCÁSTICA DE TRÁFEGO EM REDES MULTISERVIÇO

por

Lino M. Tralhão*

J. Craveirinha**

J. P. Paixão***

Resumo

Nos últimos anos, tem sido desenvolvido um novo tipo de redes de telecomunicação que permitem a transmissão integrada de diferentes serviços (telefone, videotelefone, teletexto, telefax, transmissão de dados, correio electrónico e outros): as Redes Digitais Integradas de Serviços – RDIS ("ISDN"). A introdução deste novo tipo de redes multiserviço, levanta novos problemas de modelização estocástica dos fluxos de tráfego, fundamentalmente devido ao facto de se tratarem de sobreposições complexas de fluxos de tráfego de naturezas diferentes. Trata-se de um tema de grande actualidade, quer do ponto de vista da teoria estocástica dos sistemas de serviço, quer do ponto de vista da aplicação em modelos de análise da qualidade do serviço (probabilidade de bloqueio e tempos de atraso) e planeamento de redes.

Pretende-se com este trabalho fazer uma resenha sobre os vários tipos de modelos estocásticos mais utilizados e referir resultados e contribuições importantes neste domínio.

Em primeiro lugar, far-se-á uma descrição muito breve das características funcionais mais relevantes das RDIS (de banda estreita e de banda larga). Em seguida, e após uma revisão de conceitos básicos sobre processos pontuais, iremos apresentar as definições e propriedades fundamentais dos processos estocásticos: alternante de renovações, "random hazard Poisson process", de Poisson comutado, e de Poisson modulado por uma cadeia de Markov. Para além da formalização de propriedades relevantes dos modelos matemáticos, serão igualmente referidas contribuições importantes sobre aplicações em problemas de modelização estocástica de redes de comunicação.

1 - INTRODUÇÃO. REDES MULTISERVIÇO

O estado presente de desenvolvimento das tecnologias da comunicação permite a concepção de um novo tipo de rede de telecomunicações (Rede Integrada de Serviços – RDIS, "Integrated Services Digital Network" – ISDN), a qual assume duas feições fundamentais: por um lado a representação digital de todas as informações transmitidas e processadas, independentemente dos seus tipos (voz, texto, dados e imagens); por outro a integração de sistemas e serviços, ou seja, a concentração de forma padronizada de uma vasta gama de teleserviços (o objectivo é todos) nessa (uma só) rede, permitindo que os vários serviços possam ser utilizados através de um procedimento padronizado para o estabelecimento e terminação de chamadas, via um "interface" também padronizado entre o equipamento terminal e a rede, e assim o suporte simultâneo, no utilizador, de tipos diferentes de informação (terminais multiserviço). As características da RDIS têm vindo a ser padronizadas, ao longo da presente década, por grupos de peritos de todo o mundo, sob a égide do CCITT ("Comité Consultatif International Télégraphique et Téléphonique"). Todos os principais fabricantes de equipamento e

* Departamento de Matemática

** Departamento de Engenharia Electrotécnica

Universidade de Coimbra / INESC - Núcleo de Coimbra

*** Departamento de Estatística, Investigação Operacional e Computação – Universidade de Lisboa

empresas operadoras estão a trabalhar na sua implementação, dados os substanciais benefícios de que é portadora (Bocker 1987):

- Os utilizadores obterão serviços adicionais e avançados, a maioria dos quais destinados a prover o crescimento de tráfego não vocal. O acesso unificado ao ISDN permitirá aos utilizadores o uso de sistemas existentes de modo menos custoso do que através de várias redes dedicadas aos vários tipos de comunicação.
- As empresas operadoras beneficiarão do conceito de rede universal, que lhes permitirá introduzir novos serviços sem grande esforço financeiro.
- Os fabricantes de equipamento darão boas vindas, acima de tudo ao acesso padronizado à rede, o qual, ao permitir o acesso a uma larga gama de serviços, por parte do utilizador, através de um só terminal (multiserviço), lhes alargará o mercado.

A rede permitirá o providenciamento de serviços, não só de tipo interactivo, mas também serviços distribucionais, nos quais existem emissores permanentes de informação que pode ser recebida por assinantes autorizados (ex. informação meteorológica permanente, serviço de horas, etc., e, no futuro, com a introdução das redes de banda larga: rádio e televisão). Nos interactivos teremos, não só os conversacionais (ex. telefone, videotelefone, teletexto, telefax, transmissão de dados), normalmente serviços em tempo real, mas também os serviços de mensagens (ex. correio fax, correio de texto), que requerem a existência de unidades de memorização ("caixas de correio"), e ainda serviços de busca (ex. software, bancos de dados), em que a informação está memorizada e disponível numa fonte a que o assinante pode aceder.

Inicialmente as redes ISDN destinavam-se a usar como suporte físico as redes telefónicas digitais já existentes. A velocidade a que uma dada informação digital é transmitida fica limitada a 2Mbits/s; temos então a chamada ISDN de banda estreita. Esta rede usa o modo de transmissão síncrono com comutação por circuitos, com tramas de 125µs, de 30 canais de transmissão, mais dois de sinalização, por trama. Temos assim um canal básico de transmissão de 64kbit/s (canal B), sendo o acesso básico à rede feito através de 2 canais do tipo B e um auxiliar de 16kbits/s (canal D 16), para sinalização e transmissão de dados de baixa velocidade. As comunicações que exigem maior velocidade, utilizarão simultaneamente mais do que dois canais de 64kbit/s. No entanto, a emergência de tecnologias de comunicação de alta velocidade (p.ex. fibras ópticas), irá permitir a transmissão a velocidades muito maiores (140Mbit/s). Temos então a chamada ISDN de banda larga (BISDN). Para esta rede já se encontra padronizado o modo assíncrono de transmissão, mais concretamente o modo ATD ("Asynchronous Time Division Mode"). Neste modo, cada comunicação digitalizada é "partida" no tempo em pequenas unidades --células. A forma como estas células são transmitidas e comutadas através da rede, é uma forma intermédia entre o sistema atrás referido e o de comutação por pacotes.

A introdução deste novo tipo de redes multiserviço, levanta novos problemas de modelização estocástica dos fluxos de tráfego, de grande interesse, quer no campo teórico, por suscitar novos desenvolvimentos da teoria dos processos estocásticos e dos sistemas estocásticos de serviços, quer no campo das aplicações no contexto dos modelos de análise e planeamento de redes de comunicação. Trata-se agora de sobreposições de fluxos de tráfego, caracterizados por ocupações de canais (no caso do modo síncrono) com múltiplicidades diferentes, diferentes durações médias de chamadas, e correlacionados, dando origem a fluxos muitas vezes não modelizáveis por processos de Poisson. Um dos aspectos de mais realce destes fluxos é, por vezes, o seu carácter explosivo ("bursty"), implicando coeficientes de variação elevados, não só quando se trata de fluxos de transbordo, mas também,

em redes de pacotes de voz e dados, devido às intermitências no envio de pacotes nas fontes, provocadas pelos períodos de silêncio. A importância da modelização estocástica destes fluxos, reside no facto de poder servir de base a modelos de determinação do grau de serviço da rede (especificado em termos de congestão de chamadas, atrasos no estabelecimento de chamadas e/ou transmissão de pacotes), componente essencial de qualquer modelo de planeamento.

Neste trabalho pretende-se fazer uma resenha actualizada do "estado da arte" nesta matéria, incluindo definições, conceitos e propriedades básicas de modelos estocásticos que têm sido utilizados neste contexto, em particular o 2-MMPP, e referir uma série de artigos relevantes nesta área. Na secção 2, começaremos por rever conceitos básicos e propriedades gerais dos Processos Pontuais, em cujo âmbito se enquadram todos os processos estocásticos abordados neste trabalho. No ponto 3, é abordado, com algum detalhe, o Processo Alternante de Renovações, por ser a base da teoria dos tempos de permanência, que serve de suporte à análise de alguns tipos de processos duplamente estocásticos. Na secção 4 são apresentadas as propriedades fundamentais do "Random Hazard Poisson Process", caso particular do Processo de Poisson Duplamente Estocástico. No ponto 5 é abordado o Processo de Poisson Comutado (ou processo 2-MMPP) e referidas algumas aplicações importantes em redes de comunicação. Na Secção 6 é apresentado e definido o m-MMPP (Processo de Poisson Modulado por uma Cadeia de Markov com m estados), e referidos alguns trabalhos de aplicação. Finalmente, no ponto 7 referem-se outros trabalhos relevantes sobre modelização de redes de comunicação. Finalmente, em apêndice, revêem-se conceitos e resultados básicos sobre processos pontuais, utilizados ao longo do texto.

2. - CONCEITOS GERAIS SOBRE PROCESSOS PONTUAIS (PP)

2.1 - Definição

Seja \mathcal{T} um conjunto infinito, e consideremos o conjunto não vazio Ω , contável. Seja agora o conjunto $\mathcal{P} = \{(f, g)\}$ de pares de funções dos tipos $f : \mathbb{N} \rightarrow \mathcal{T}$, $g : \mathbb{N} \rightarrow \Omega$, no qual se encontra definida uma medida de probabilidade σ . O espaço de probabilidade $\{\mathcal{P}, \zeta, \Omega\}$, estabelece um processo pontual, sendo cada par (f, g) uma realização, ou caminho de amostragem, do processo; Ω chama-se conjunto dos tipos de pontos.

Em qualquer processo pontual podemos também considerar o conjunto de variáveis aleatórias (VAs), de contagem, $\mathfrak{B} = \{N(A), A \subset \mathcal{T}\}$:

$$N(A) : \mathcal{P} \rightarrow \mathbb{N} \tag{2.1}$$

$$(f, g) \rightarrow \#\{n \in \mathbb{N} : f(n) \in A\}$$

Em cada realização ($w := (f, g)$) do processo, os "pontos" são os elementos do contradomínio de f . Os elementos de Ω são os tipos de pontos. Se Ω tiver um só elemento, deixa de ser pertinente na definição de PP; neste caso o PP diz-se simples. Caso $\#\Omega > 1$, o PP diz-se marcado. Se $\Omega \subset \mathbb{N}$, $\#\Omega > 1$ e as funções f injectivas, temos um processo múltiplo.

Para os casos em estudo neste trabalho, vamos considerar que $t = \mathbb{R}^+$, representando o "tempo". Uma realização (w) do PP corresponde então a uma sequência de eventos (pontos) no tempo; $f(i)$, $i \in \mathbb{N}$, passará a representar-se por t_i e indica o instante (ponto) do evento de ordem i . As VAs $N(A)$ indicam então o

número de eventos nos conjuntos A, de instantes. Vamos também considerar que os conjuntos (A) acima referidos, são de Borel.

2.2 - O PP como processo de entrada num sistema estocástico de serviço

Como o processo de entrada num sistema estocástico de serviço corresponde ao processo de chegada, ao longo do tempo, de utilizadores do(s) servidor(es) do sistema, este processo é um processo pontual, sendo $t = \mathbb{R}^+$ (admite-se aqui a existência de um instante zero, sendo o estudo do processo feito somente para os eventos ocorridos depois desse instante); então t_i é o instante de chegada de um utilizador. Se os utilizadores chegarem em grupos ("batches"), temos um PP múltiplo ou por lotes; se houver tipos diferentes de utilizadores, temos um PP marcado.

2.3 - Algumas propriedades dos PP ($t = \mathbb{R}^+$)

Para caracterizar um PP deste tipo ($t = \mathbb{R}^+$) é usual considerarem-se as distribuições conjuntas:

$$\text{pr}\{N(A_i) = n_i \in \mathbb{N} \cup \{0\}, i = 1, \dots, k\}, k = 1, 2, \dots \quad (2.2)$$

em que "pr" indica "probabilidade" e sendo os A_i conjuntos de Borel, arbitrários, de \mathbb{R}^+ . Convencionamos também, para $N(a, b]$ a notação equivalente $N(a, b)$, e para $N(0, t)$, $N(t)$. Esta caracterização é insuficiente para o caso de PPs marcados, sendo aqui usual completá-la com as distribuições conjuntas:

$$\text{pr}\{M_i = \beta, \beta \in \Omega, i = 1, \dots, k\}, k = 1, 2, \dots \quad (2.3)$$

em que a VA M_i indica o tipo do evento de ordem i .

Outra forma de caracterização de um PP consiste na consideração das distribuições conjuntas das VAs:

$$\{T_i, i = 1, \dots\} \quad (2.4)$$

em que a VA T_i indica o instante do i -ésimo evento. Esta caracterização, juntamente com (2.3) provem directamente da definição de PP enunciada em 2.1.

De (2.4) podemos encontrar a distribuição conjunta dos comprimentos de intervalos consecutivos:

$$\{X_i, i = 0, 1, \dots\} \quad (2.5a)$$

sendo:

$$\begin{aligned} X_0 &= T_1 \\ X_i &= T_{i+1} - T_i \quad i = 1, \dots \end{aligned} \quad (2.5b)$$

obviamente temos $T_i = X_0 + \dots + X_i$ (2.5.c).

No caso de um PP simples, existe uma conexão entre as caracterizações (2.4) e (2.2), através da importante relação:

$$\text{pr}\{N(t) > n\} = \text{pr}\{T_{n+1} \leq t\} \quad (2.6)$$

2.4 - Estacionaridade

O PP diz-se completamente estacionário se a distribuição conjunta (2.2) for invariante para qualquer translação de x unidades de todos os conjuntos A_i . Se esta propriedade for verdadeira para $k=1$, não se sabendo o que acontece para outros valores de k , o PP diz-se simplesmente estacionário. Se a média de $N(A)$ ($E[N(A)]$) e a variância de $N(A)$ ($\text{var}[N(A)]$) forem invariantes para qualquer translação de x unidades do conjunto (de Borel) A, o processo diz-se fracamente estacionário.

O PP tem estacionaridade nos intervalos sse a distribuição conjunta de:

$$X_{s_1}, X_{s_2}, \dots, X_{s_k}$$

depende somente do vector $(s_2 - s_1, s_3 - s_1, \dots, s_k - s_1)$, $\forall (s_1, \dots, s_k)$, $s_i \in \mathbb{N} \cup \{0\}$, $\forall k \in \mathbb{N}$.

A estacionaridade completa não implica estacionaridade nos intervalos. A questão da estacionaridade prende-se com a questão da escolha do instante 0 do PP. Assim, de um PP estacionário resultará um PP com as mesmas propriedades estocásticas, se for feita uma qualquer translação do instante inicial. Ora, se num PP completamente estacionário se fizer uma translação aleatória do instante de origem, a nova origem tem mais probabilidade de "cair" num intervalo longo do que num curto. Assim sendo, e havendo dependência estatística entre os intervalos, o novo PP obtido não terá as mesmas propriedades no tocante à sequência de intervalos. Por outro lado, se começarmos com a propriedade da estacionaridade nos intervalos, estamos efectivamente a condicionar o processo à existência de um evento no instante $t = 0$ (i.e. $\lim_{h \rightarrow 0^+} N(-h, 0) > 0$). Neste caso, o processo de contagem resultante não é, em geral, estacionário. Sabe-se, no entanto, que se tivermos um PP completamente estacionário, o PP resultante do anterior, condicionando a origem a um evento arbitrário, tem estacionaridade nos intervalos (vide, p.ex., eq. A.3.6).

Resulta do acima exposto que podemos considerar duas abordagens à 'estacionaridade' de um mesmo PP (melhor dizendo, a um tipo de PP): por um lado podemos considerar a sua origem num instante arbitrário (escolha arbitrária da origem, no tempo) e obtemos um processo estacionário no tocante às contagens; seguidamente, observamo-lo supondo a origem num evento arbitrário e temos um PP estacionário nos intervalos. Sempre que falarmos em PP estacionário, sem indicação mais precisa, estamos-nos a referir a um PP nestas condições. De ambas as abordagens podemos obter informações úteis sobre o processo, e relacionadas entre si, pelo que, em geral, ambas são feitas conjuntamente. Sempre que for caso disso, usaremos o índice (N) nos símbolos referentes a um PP estacionário (ou à sua abordagem) com origem num instante arbitrário, e o índice (A) para o caso de termos a origem num evento arbitrário. Neste último caso a sequência $\{T_i\}$ (de 2.5) poderá ser designada por $\{S_i\}$.

Em Apêndice são apresentados conceitos e propriedades relevantes para a descrição e análise matemática dos processos pontuais. Um caso particular é o Processo de Renovações (PR), em que os intervalos entre eventos são independentes entre si e, à excepção do primeiro, identicamente distribuídos. Este processo é muito utilizado como modelo aproximado de outros PPs cujo estudo exacto acarrete grande esforço computacional. Por exemplo (modelos de base ERT (Wilkinson 1956) e BPP (Delbrouck 1981, 1983)), como modelo de fluxos de tráfego de 'chamadas' em redes telefónicas com transbordo.

Seguidamente é apresentado um processo pontual (o ARP), cujo modelo serve de base ao estudo dos processos RHPP e SPP, depois analisados. Neste estudo seguiu-se a metodologia proposta por Rossiter (1986), baseada na chamada Teoria dos Tempos de Permanência. Entretanto introduziram-se algumas alterações na notação e precisaram-se alguns conceitos no que toca às "fases" do ARP.

3 - O PROCESSO ALTERNANTE DE RENOVAÇÕES ("ALTERNING RENEWAL PROCESS" - ARP)

3.1 - Definições

Seja $\{(I_n, Z_n), n=0, 1, \dots\}$ um processo estocástico de 2 variáveis. Seja $Z_0=0$, $\text{pr}\{I_0=0\} = p_0$, $\text{pr}\{I_0=1\} = p_1$, $(p_1 + p_0 = 1)$, onde a sequência $\{I_n, n=0, 1, \dots\}$ satisfaz a $I_n = 1 - I_{n-1}$, $n \geq 1$. Consideremos que, condicionado ao valor de I_0 , todos os $X_n = Z_n - Z_{n-1}$, $n \geq 1$, são VAs não negativas e mutuamente

independentes. Sejam H' (ou G') a FD de X_1 se $I_0 = 0$ (se $I_0 = 1$), e H (ou G) a FD de X_n , $n > 1$ se $I_{n-1} = 0$ (se $I_{n-1} = 1$). As VAs de FD H e G têm, respectivamente médias $1/\gamma$ e $1/\omega$.

Nestas condições, o PP representado pela sequência $\{I_n, n = 0, 1, \dots\}$ é um Processo Alternante de Renovações (ARP). I_n representa o tipo (0 ou 1), e Z_n o instante do n -ésimo evento (ponto). Trata-se assim de um PP marcado, com 2 tipos de pontos. Por outro lado, à excepção do primeiro, os intervalos terminados em pontos do tipo 1 têm FD H e os terminados em pontos do tipo 0, FD G . Podemos então considerar também a existência de dois tipos de intervalos: 0 e 1. Convencionamos que o tipo de um intervalo é o contrário do do seu supremo.

Se $H' = H$ e $G' = G$, temos um ARP ordinário (OARP).

Se H' (ou G') for a FD do tempo de recorrência para a frente da VA de FD H (G), então o ARP diz-se em equilíbrio (EARP). Trata-se, como no caso do PR, de um ARP obtido pela translação arbitrária da origem de um outro ARP.

Nos outros caso temos um ARP modificado (MARP).

Um caso particular de OARP é aquele em que tanto H com G são exponenciais. Temos neste caso um Processo Alternante de Poisson (APP).

3.2- Algumas VAs pertinentes

Seja c_E a função característica do conjunto E .

Sejam as VAs:

$$N_{(i)}(t) = \sum_{n=1}^{\infty} \chi_{(x : x \leq t)}(Z_{2n-1}) \tag{3.1}$$

$$N_{(p)}(t) = \sum_{n=1}^{\infty} \chi_{(x : x \leq t)}(Z_{2n})$$

As VAs $N_{(i)}(t)$ e $N_{(p)}(t)$ (fig. 3.1) são, respectivamente, as VAs de contagem de pontos de ordem ímpar e par.

Os intervalos do ARP também se designam por fases. Uma fase diz-se ímpar ou par, consoante a ordem do seu ponto terminal (supremo).

Sejam as VAs:

$$I_{(i)}(t) = \chi_{(x : x \neq N_{(p)}(t))}(N_{(i)}(t)) \tag{3.2}$$

$$I_{(i)}(t) = 1 - I_{(p)}(t)$$

As VAs $I_{(p)}(t)$ e $I_{(i)}(t)$ (fig. 3.1) são as VAs indicadoras, respectivamente, de fases pares e ímpares do processo ($I_{(p)} = 1$ (0) se o processo estiver, no instante t , numa fase par (ímpar)).

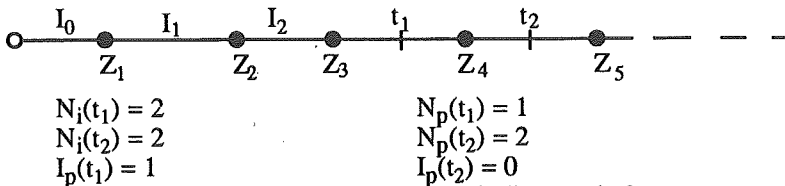


Fig. 3.1 – Contagem de pontos e indicação de fases

Então, a VA $I(t)$, indicadora do estado do processo no instante t , vem dada por:

$$I(t) = I_{(p)}(t) \chi_{\{0\}}(I_0) + I_{(i)}(t) \chi_{\{1\}}(I_0), t \geq 0 \tag{3.3}$$

Seja a VA $\alpha(t)$, o tempo acumulado de permanência no estado 1, durante o intervalo de tempo $(0,t]$:

$$\alpha(t) = \int_0^t I(u)du, \quad t > 0 \tag{3.4}$$

Sejam as VAs:

$$S_n := \sum_{i=1}^n X_{2i-1} \tag{3.5}$$

$n \geq 1$

$$T_n := \sum_{i=1}^n X_{2i}$$

S_n é a soma de ordem n das fases ímpares.

T_n é a soma de ordem n das fases pares.

Sejam:

$$\alpha_{(p)}(t) = T_{N_{(p)}(t)} I_{(i)}(t) + (t - S_{N_{(p)}(t)}) I_{(p)}(t) \tag{3.6}$$

$$\alpha_{(i)}(t) = t - \alpha_{(p)}(t)$$

$\alpha_{(p)}(t)$ ($\alpha_{(i)}(t)$) é a VA tempo de permanência no tipo par (ímpar), isto é a porção de $(0,t]$ em que o processo esteve em intervalos do tipo X_{2n-1} (X_{2n}).

De (3.3), (3.4) e (3.6) temos:

$$\alpha(t) = \alpha_{(p)}(t) \lambda_{\{0\}}(I_0) + \alpha_{(i)}(t) \lambda_{\{1\}}(I_0), \quad t > 0 \tag{3.7}$$

Seja $\phi(x, t)$ a FD de $\alpha(t)$. Sejam $\phi_{(1)}(x, t)$ e $\phi_{(0)}(x, t)$ as FD de $\alpha(t)$, condicionadas ao tipo de instante inicial:

$$\phi_{(j)}(x, t) = \text{pr}\{\alpha(t) \leq x \mid I_0 = j\} \tag{3.8}$$

Temos então:

$$\phi(x, t) = p_0 \phi_{(0)}(x, t) + p_1 \phi_{(1)}(x, t) \tag{3.9}$$

Seja:

$G_n^{(j)}$, $j = 0, 1$ a FD da VA soma dos tempos nas primeiras n fases do tipo 1, sendo $I_0 = j$.

$H_n^{(j)}$, $j = 0, 1$ a FD da VA soma dos tempos nas primeiras n fases do tipo 0, sendo $I_0 = j$. Então:

$$G_n^{(j)} = \text{FD} \left\{ \sum_{i=1}^{2n} X_i I_{i-1}, \quad I_0 = j, \quad n = 1, 2, \dots \right\} = \text{FD} \{ jS_n + (1 - j)T_n \} \tag{3.10}$$

$$H_n^{(j)} = \text{FD} \left\{ \sum_{i=1}^{2n} X_i I_i, \quad I_0 = j, \quad n = 1, 2, \dots \right\} = \text{FD} \{ jT_n + (1 - j)S_n \}$$

Dada a equidistribuição de todos os intervalos X_n , com $n > 1$, do mesmo tipo:

$$G_n^{(0)(ls)}(s) = (G^{(ls)}(s))^n$$

$$G_n^{(1)(ls)}(s) = G^{(ls)}(s) (G^{(ls)}(s))^{n-1} \tag{3.11a}$$

$$H_n^{(1)(ls)}(s) = (H^{(ls)}(s))^n$$

$$H_n^{(0)(s)} = H^{(s)}(s) (H^{(s)}(s))^{n-1}$$

Se se tratar de um processo em equilíbrio (EARP), temos:

$$H^{(s)}(s) = \frac{1-H^{(s)}(s)}{s} \gamma \quad G^{(s)}(s) = \frac{1-G^{(s)}(s)}{s} \omega \quad (3.11b)$$

No caso de um OARP:

$$H^{(s)}(s) = H^{(s)}(s) \quad G^{(s)}(s) = G^{(s)}(s) \quad (3.11c)$$

3.3 - Alguns resultados (Rossiter 86), (Takacs 57):

Para um EARP:

$$\phi_{(e)}(x, t) = \frac{\phi}{\phi+\omega} \phi_{(0)}(x, t) + \frac{\omega}{\phi+\omega} \phi_{(1)}(x, t) \quad (3.12)$$

em que o índice (e) indica um processo em equilíbrio, não se sabendo qual a fase inicial.

Para um MARP, com $I_0 = 0$:

$$\phi_{(0)}(x, t) = \begin{cases} 0 & \text{se } x < 0 \\ \sum_{n=0}^{\infty} G_n(x) \cdot [H_n - H_{n+1}] [(t-x)^-] & \text{se } x \in (0, t) \\ 1 & \text{se } x \geq t \end{cases} \quad (3.13a)$$

$$\phi_{(1)}(x, t) = \begin{cases} 0 & \text{se } x < 0 \\ 1 - \sum_{n=0}^{\infty} H_n((t-x)^-) [G_n - G_{n+1}](x) & \text{se } x \in (0, t) \\ 1 & \text{se } x \geq t \end{cases} \quad (3.13b)$$

com $H_1 = H'$, $G_1 = G'$.

Para um MARP, temos, para $x \geq 0$ e $|s| > 0$:

$$s\phi_{(0)}^{(1)}(x, s) = 1 - e^{-sx} \frac{H_1^{(s)}(s)}{H^{(s)}(s)} \left(1 - (1 - H^{(s)}(s)) \sum_{n=0}^{\infty} (x)(H^{(s)}(s))^n\right) \quad (3.14)$$

$$s\phi_{(1)}^{(1)}(x, s) = 1 - e^{-sx} \sum_{n=0}^{\infty} ([G_n - G_{n+1}](x))(H^{(s)}(s))^n$$

Para o caso de termos um APP, (3.14) fica:

$$\phi_{(0)}^{(1)}(x, s) = \frac{1}{s} \left\{ 1 - \frac{\omega}{s+\omega} e^{-\frac{s(s+\omega+\gamma)x}{s+\gamma}} \right\} \quad (3.15a)$$

$$\phi_{(1)}^{(1)}(x, s) = \frac{1}{s} \left\{ 1 - e^{-\frac{s(s+\omega+\gamma)x}{s+\omega}} \right\}$$

Se se tratar de um EAPP com fase inicial desconhecida:

$$\phi_{(e)}^{(1)}(x, s) = \frac{1}{s} \left(1 - \frac{\omega(s+\omega+\gamma)}{(\omega+\gamma)(s+\omega)} e^{-\frac{s(s+\omega+\gamma)x}{s+\omega}} \right) \quad (3.15b)$$

Para um MARP, temos, para $\theta, s > 0$:

$$s\phi_{(0)}^{(ls,l)}(\theta, s) = 1 - \frac{\theta[1-G^{(ls)}(s+\theta)]H^{(ls)}(s)}{(s+\theta)[1-G^{(ls)}(s+\theta)H^{(ls)}(s)]} \quad (3.16a')$$

$$s\phi_{(1)}^{(ls,l)}(\theta, s) = 1 - \frac{\theta[1-G^{(ls)}(s+\theta)-[G^{(ls)}-G'^{(ls)}](s+\theta).H^{(ls)}(s)]}{(s+\theta)[1-G^{(ls)}(s+\theta).H^{(ls)}(s)]} \quad (3.16a'')$$

Para um EARP, temos, para $\theta, s > 0$:

$$s\phi_{(0)}^{(ls,l)}(\theta, s) = 1 - \frac{\theta\omega(1-G^{(ls)}(s+\theta))(1-H^{(ls)}(s))}{s(s+\theta)(1-G^{(ls)}(s+\theta)H^{(ls)}(s))} \quad (3.16b)$$

e se a fase inicial for desconhecida:

$$s\phi_{(e)}^{(ls,l)}(\theta, s) = 1 - \frac{\theta\omega}{(\omega+\gamma)(s+\omega)} - \frac{\theta^2\omega\gamma(1-G^{(ls)}(s+\theta))(1-H^{(ls)}(s))}{(\omega+\gamma)s(s+\theta)^2(1-G^{(ls)}(s+\theta)H^{(ls)}(s))}$$

Para um APP, temos, para $\theta, s > 0$:

$$\phi_{(0)}^{(ls,l)}(\theta, s) = \frac{s+\omega+\gamma+\theta}{s(s+\omega+\gamma)+(s+\omega)\theta}$$

$$\phi_{(1)}^{(ls,l)}(\theta, s) = \frac{s+\omega+\gamma}{s(s+\omega+\gamma)+(s+\omega)} \quad (3.16c)$$

$$\phi_{(e)}^{(ls,l)}(\theta, s) = \frac{s+\omega+\gamma \frac{\theta\gamma}{\omega+\gamma}}{s(s+\omega+\gamma)+(s+\theta)\theta}$$

Para um MARP, começando num ponto "OFF", temos :

$$\mathfrak{L}\{E[\alpha(t)]\} = \frac{[1-G^{(ls)}(s)]H^{(ls)}(s)}{s^2[1-G^{(ls)}(s)H^{(ls)}(s)]} \quad (3.17a)$$

se começar num ponto "ON":

$$\mathfrak{L}\{E[\alpha(t)]\} = \frac{1-G'^{(ls)}(s)-[G^{(ls)}-G'^{(ls)}](s)H^{(ls)}(s)}{s^2[1-G^{(ls)}(s)H^{(ls)}(s)]} \quad (3.17b)$$

em que $\mathfrak{L}\{f\}$ é a transformada de Laplace da função f .

4 - O "RANDOM HAZARD POISSON PROCESS" (RHPP)

4.1 - Introdução

Este processo começou a ser estudado por Gaves (1963). Lawrence (1972) foi o primeiro a verificar a vantagem da abordagem deste processo através dos tempos de permanência. Com a teoria apresentada na secção 3 deste trabalho, é possível dar-se um registo muito completo das propriedades deste processo.

Tran-Gia (1983) mostrou a aplicabilidade do processo como modelo de tráfego em redes de computadores. Gaver (1963) e Kopinska (1984) propuseram-no como modelo de fiabilidade em sistemas sujeitos a níveis alternados de ambiente de operação.

4.2 - Definição

Para a sua definição, interessa considerar primeiramente a definição do Processo de Poisson Duplamente Estocástico (PPDE), de que o RHPP é um caso particular:

Um PPDE é um PP com um processo de intensidades $\{\Lambda(t), t > 0\}$ tal que, para todas as realizações de $\Lambda(t)$, o número de eventos em todo e qualquer intervalo E tem distribuição de Poisson de média $\int_E \Lambda(u) du$. Por outro lado, trata-se de um processo pontual completamente aleatório, isto é, as V.A.s $N(A)$ e $N(B)$ (ver secção 2.1) são independentes, para intervalos A e B disjuntos.

O RHPP é um PPDE em que o processo $\Lambda(t)$ é dado por:

$$\Lambda(t) = \begin{cases} \lambda_1 \in [0, \infty[& \text{se } I(t) = 1 \\ \lambda_2 \in [0, \infty[& \text{se } I(t) = 0 \end{cases} \quad (4.1)$$

Onde $I(t)$ é a VA indicadora do estado de um ARP (eq. 3.3). Portanto a intensidade do processo assume alternadamente os valores λ_1 e λ_2 , de acordo com as fases, respectivamente 1 e 0, de um ARP. Podemos então falar de RHPP marcado, ordinário ou em equilíbrio, consoante o tipo de ARP subjacente.

4.3 - Algumas grandezas (parâmetros) pertinentes

No caso de um RHPP equilibrado, com escolha arbitrária no tempo, do instante $t = 0$ (origem dos tempos), este "cai" numa fase de intensidade λ_2 com probabilidade $p_0 = \gamma/(\omega + \gamma)$ (4.2) (γ e ω referem-se ao ARP subjacente e encontram-se definidas em 3.1), tal como acontecia no EARP (que é o ARP subjacente a este RHPP). No entanto, se se eleger aleatoriamente, para instante $t = 0$, um evento, teremos agora $p_0 = \lambda_2 \gamma / (\lambda_1 \omega + \lambda_2 \gamma)$ (4.3) (no OARP $p_0 = 1/2$), como facilmente se verifica se tivermos em conta que no RHPP, "evento" não significa mudança de fase. Note-se que, neste último caso não teremos um RHPP ordinário, mas sim equilibrado, ao contrário do que acontece no ARP.

Outras grandezas pertinentes a considerar são:

-- Intensidade média:

$$\lambda = \frac{\lambda_1 \omega + \lambda_2 \gamma}{\omega + \gamma} \quad (4.4)$$

-- Frequência de comutação (número médio de fases por unidade de tempo):

$$f = \omega + \gamma \quad (4.5)$$

-- Razão de fases (k) (a razão entre a média dos tempos de permanência nas fases 0 e nas fases 1):

$$k = \gamma / \omega \quad (4.6)$$

4.4 - Alguns resultados

Para a determinação da importante função geradora de probabilidades do número de eventos no intervalo $]0, t]$, $P(z, t)$ (vide secção A.5), para o caso do RHPP, convém abordar primeiramente um seu caso particular – o Processo de Poisson Interrompido Generalizado ("Generalized Interrupted Poisson Process" GIPP) (Tran-Gia, 1983). Trata-se de um RHPP em que $\lambda_2 = 0$.

Seja então um GIPP, e sejam $N(t)$ e $\alpha(t)$, respectivamente, a VA número de eventos em $]0, t]$, e tempo de permanência no estado 1 do ARP subjacente. Então, para $x \in [0, t]$:

$$\text{pr}\{N(t) = k \mid \alpha(t) = x\} = \frac{(\lambda_1 x)^k}{k!} \cdot e^{-\lambda_1 x} \quad (4.7)$$

uma vez que num GIPP o número de eventos em]0,t] é igual ao de um processo de Poisson, de intensidade λ_1 , no intervalo]0, $\alpha(t)$], dado que, durante o tempo de permanência no estado 0 do ARP subjacente, não houve eventos ($\lambda_2 = 0$). Sendo assim, e pelo teorema da probabilidade total:

$$p_k(t) := \text{pr}\{N(t) = k\} = \int_{x=0}^{x=t} \frac{(\lambda_1 x)^k}{k!} e^{-\lambda_1 x} d\phi(x, t) \tag{4.8}$$

em que $\phi(x, t)$ é a FD de $\alpha(t)$. Então $P(z, t)$, função geradora de probabilidade de $\{p_k(t)\}$ vem dada por ($0 \leq z < 1, t > 0$):

$$P(z, t) = \int_{x=0}^{x=t} e^{-\lambda_1 x(1-z)} d\phi(x, t) \tag{4.9}$$

ou seja:

$$P(z, t) = \phi^{(1s)}(\lambda_1(1-z), t) \tag{4.10}$$

ou:

$$P^{(1)}(z, s) = \phi^{(1s,1)}(\lambda_1(1-z), s) \tag{4.11}$$

Substituindo agora, (3.16) em (4.11), encontramos expressões para $P^{(1)}(z, s)$, nos diversos tipos de GIPP (modificado, ordinário, etc.).

Um caso interessante de GIPP é aquele em que as fases 1 têm distribuição exponencial, com média $1/\gamma$; Kingman (1964) demonstrou ser este um processo de renovações. Num outro caso interessante, o ARP subjacente é um APP. Temos então o Processo de Poisson Interrompido (IPP), introduzido, a propósito da modelização de tráfegos de transbordo em redes telefônicas, por Kuczura (1973).

A abordagem do RHPP torna-se agora relativamente fácil:

Vamos supor, sem perda de generalidade, que $\lambda_1 > \lambda_2$. Então o RHPP é a sobreposição de um GIPP de intensidade $\lambda_1 - \lambda_2$ nas fases 1, e de um processo de Poisson independente daquele e de intensidade λ_2 . Vem então que:

$$N(t) = N_1(t) + N_2(t) \tag{4.12}$$

em que as VAs de contagem N_1 e N_2 dizem respeito, respectivamente ao GIPP e ao processo de Poisson. Como a função geradora de probabilidade da soma de 2 VAs independentes é igual ao produto das funções geradoras de probabilidade dos termos da soma, vem que:

$$P(z, t) = P_1(z, t) \cdot P_2(z, t) \tag{4.13}$$

Mas $P_2(z, t)$ é independente da fase inicial (diz respeito a um processo de Poisson), sendo bem conhecido que:

$$P_2(z, t) = e^{-\lambda_2(1-z)t} \tag{4.14}$$

então:

$$P(z, t) = P_1(z, t) \cdot e^{-\lambda_2(1-z)t} \tag{4.15}$$

Tendo agora (4.11) em consideração, vem, finalmente:

$$P^{(1s)}(z, s) = \phi^{(1s,1)}((\lambda_1 - \lambda_2)(1-z), s + \lambda_2(1-z)) \tag{4.16}$$

Tal como sugerido para o caso do GIPP, basta agora substituir as equações (3.16) em (4.16), para termos expressões de $P^{(1)}(z, s)$, para os diferentes tipos de RHPP.

5 - O PROCESSO DE POISSON COMUTADO ("SWITCHED POISSON PROCESS" - SPP)

5.1 - Introdução

O nome deste processo é devido a Tran-Gia (1983) e a Van Hoorn e Seelen (1983), tendo estes últimos apresentado uma solução numérica para o sistema estocástico de serviço SPP/G/1, usando-a para testar aproximações da fila de espera GI/G/1. Apareceu na literatura, pela primeira vez, como o processo de entrada de um sistema de serviço em Eisen e Tainiter (1963). Os mesmos resultados foram obtidos por Yechiali e Naor (1971). Kuczura (1973) introduz um caso particular de SPP -- o IPP já aqui referido. Heffes (1973, 1976) e Ross (1978) analisaram sistemas de serviço com o IPP como entrada. Heffes (1980) usa-o num modelo para agregados de fluxos de tráfego oferecidos a uma rede com comutação por pacotes.

O processo em causa é um modelo de tráfego de dados particularmente apropriado, uma vez que o sistema de serviço SPP/G/1 pode ser resolvido exactamente (Rossiter 1986). Se um fluxo de tráfego de dados pode ser modelizado por um SPP, o desempenho de um sistema de atraso ao qual é oferecido, pode ser prontamente determinado.

Apresentaremos aqui, primeiramente alguns resultados àcerca do SPP, e seguidamente abordaremos o problema da caracterização de fluxos de tráfego de dados, usando o SPP, em particular no caso de termos sobreposições de fluxos.

O estudo aqui feito, resulta de ideias retiradas de Whitt (1982), Sriram e Whitt (1986), Holtzman (1986), Heffes e Lucantoni (1986) e, fundamentalmente, Rossiter (1986).

5.2 - Definição

O SPP é um RHPP em que o MARP subjacente é um APP. Assim sendo, trata-se de um MMPP (vide secção 6 deste trabalho) em que a cadeia de Markov subjacente tem somente 2 estados -- 2-MMPP.

5.3 - Análise do Processo

5.3.1 - Estacionaridade

Sendo o MARP subjacente um APP, torna-se evidente que, dado estarem envolvidas unicamente distribuições exponenciais, um SPP em que a fase na origem tem intensidade λ_2 com probabilidade p_0 dada por 4.2, é completamente estacionário, não existindo qualquer distinção entre SPP ordinário e em equilíbrio, tomando estas designações com o significado já enunciado a propósito do RHPP. Por outro lado as definições dos parâmetros λ , λ_1 , λ_2 , ω , γ , f , k , feitas para o RHPP, são aqui igualmente válidas.

5.3.2 - Tipos, consoante o instante de origem

No que toca a condicionamentos sobre o instante de origem teremos, tal como no RHPP, os seguintes tipos de SPP:

- 1 - Origem numa fase 0 do APP. Neste caso será usado o índice 0 em símbolos que lhe digam respeito.
- 2 - Origem numa fase 1 do APP. Aqui o índice 1 será o usado.
- 3 - Origem num evento arbitrário. Tal com referido na secção 2.4, usaremos o índice A.
- 4 - Escolha arbitrária, no tempo, da origem (logo ausência de qualquer condicionamento sobre a origem). De acordo com a mencionada notação, usaremos o índice N.

5.3.3 - Alguns resultados

A análise feita no caso do RHPP, para a obtenção de uma expressão para $P^{(l)}(z, s)$, aplica-se, com é evidente, ao SPP: o SPP é a sobreposição de um Processo de Poisson com um IPP. Então, de (3.16c), do exposto em 4.4, e após algumas manipulações algébricas:

A - Para um $SPP_{(0)}$:

$$P_{(0)}^{(l)}(z, s) = \frac{s+\omega+\gamma+\lambda_1(1-z)}{d(z,s)} \tag{5.1}$$

B - Para um $SPP_{(1)}$:

$$P_{(1)}^{(l)}(z, s) = \frac{s+\omega+\gamma+\lambda_2(1-z)}{d(z,s)} \tag{5.2}$$

C - Para um $SPP_{(A)}$:

$$P_{(A)}^{(l)}(z, s) = \Pi^{(l)}(z, s) = \frac{s+\omega+\gamma+\lambda_1\lambda_2 \frac{(\omega+\gamma)(1-z)}{\lambda_1\omega+\lambda_2\gamma}}{d(z,s)} \tag{5.3}$$

D - Para um $SPP_{(N)}$:

$$P_{(N)}^{(l)}(z, s) = \frac{s+\omega+\gamma + \frac{(\lambda_1\gamma+\lambda_2\omega)(1-z)}{\omega+\gamma}}{d(z,s)} \tag{5.4}$$

onde:

$$d(z, s) = s(s+\omega+\gamma) + [(s+\omega)\lambda_1 + (s+\gamma)\lambda_2] (1-z) + \lambda_1\lambda_2(1-z)^2 \tag{5.5}$$

de notar que $d(z, s)$ pode ser factorizado na forma $(s + \eta_1) (s + \eta_2)$ (5.6), onde $\eta_1 = \eta_1(z)$ e $\eta_2 = \eta_2(z)$ são as raízes da equação:

$$\eta^2 - [(\lambda_1+\lambda_2)(1-z)+\omega+\gamma]\eta + [\lambda_1 \omega + \lambda_2 \gamma + \lambda_1 \lambda_2(1-z)](1-z) = 0 \tag{5.7}$$

cujo discriminante $\Delta(z)$, $0 \leq z < 1$, é não negativo, supondo, sem perda de generalidade, $\lambda_1 + \lambda_2$, tal como para o caso do RHPP.

De (5.6) podemos concluir que as T.L. em (5.1) a (5.4) são facilmente invertíveis, de modo a encontrarmos expressões para as funções geradoras de probabilidade das distribuições $\{p_k(t)\}$, nos vários tipos de SPP.

Analizemos mais em pormenor o caso $SPP_{(N)}$. Invertendo (5.4) vem (Tran-Gia (1983)):

$$P_{(N)}(z, t) = \frac{1}{\sqrt{\Delta}} [(R - \eta_1)e^{-\eta_1 t} + (\eta_2 - R)e^{-\eta_2 t}] \tag{5.8}$$

$$R = \omega + \gamma + \frac{\lambda_1\gamma+\lambda_2\omega}{\omega+\gamma} (1-z)$$

De (5.4) e (A.1.13) (com $k = 1$) podemos deduzir a T.L.S. da FD da V.A. intervalo entre eventos:

$$\begin{aligned} F_x^{(ls)}(s) &= F_{s_1}^{(ls)}(s) = 1 - sp_0^{(l)}(s) = 1 - sP_{(N)}^{(ls)}(0, s) = \\ &= \frac{(\lambda_1^2\omega+\lambda_2^2\gamma)s+(\lambda_1\omega+\lambda_2\gamma+\lambda_1\lambda_2)(\lambda_1\omega+\lambda_2\gamma)}{(\lambda_1\omega+\lambda_2\gamma)d(0,s)} \end{aligned} \tag{5.9}$$

Invertendo a T.L. de (5.9) obtemos uma expressão para a p.d.f. da V.A. intervalo entre eventos. Do facto de $d(0,s)$ ser um polinómio de grau 2 em s , com discriminante não negativo, podemos concluir que $f_X(t)$ é da forma:

$$f_X(t) = A_1 e^{-r_1 t} + A_2 e^{-r_2 t} \quad (5.10)$$

em que A_1, A_2, r_1, r_2 são expressões nos parâmetros $\lambda_1, \lambda_2, \omega, \gamma$. Concluímos então que os intervalos entre eventos de um SPP têm distribuição hipere exponencial.

De (A.5.4) e de (5.4) podemos obter uma expressão para o quadrado do coeficiente de variação:

$$c^2 = 1 + \frac{2\omega\gamma(\lambda_1 - \lambda_2)^2}{(\lambda_1\omega + \lambda_2\gamma + \lambda_1\lambda_2)(\omega + \gamma)^2} \quad (5.11)$$

e, daqui, $\text{var}[X]$ ($\text{var}[X] = c^2/\lambda^2$) (vide secções 2.4 e A.3.2). λ é a intensidade do processo, sendo dada por (4.4). Notar que se trata de um processo estacionário, pelo que (A.3.10) é aplicável.

Usando (5.4), (A.4.4) e (A.5.3) obtemos os coeficientes de correlação ($\rho_k, k = 1, 2, \dots$):

$$\rho_k = \frac{\lambda^2 a}{c^2} b^k, \quad k \geq 1 \quad (5.12)$$

com:

$$a = \frac{\omega\gamma(\lambda_1 - \lambda_2)^2}{(\lambda_1\omega + \lambda_2\gamma)^2(\lambda_1\omega + \lambda_2\gamma + \lambda_1\lambda_2)}$$

$$b = \frac{\lambda_1\lambda_2}{\lambda_1\omega\lambda_2\gamma + \lambda_1\lambda_2}$$

De (5.12) concluímos facilmente que o conhecimento de ρ_1 e ρ_2 permite o conhecimento de toda a sequência $\{\rho_k\}$, entre os intervalos X_i e X_{i+k} .

Usando (A.3.12), (A.3.13), (A.3.10) e (5.12), ficamos a conhecer a sequência $\{Z_k, k = 1, 2, \dots\}$ dos índices de dispersão dos intervalos:

$$Z_k = c^2 + \frac{2\lambda_1\lambda_2\omega\gamma(\lambda_1 - \lambda_2)^2}{(\omega + \gamma)^2(\lambda_1\omega + \lambda_2\gamma)^2} \quad (5.13a)$$

fazendo $k \rightarrow \infty$, e tendo em consideração que, de (5.12), $b < 1$:

$$z(\infty) = z_\infty = 1 + \frac{2\omega\gamma(\lambda_1 - \lambda_2)^2}{(\lambda_1\omega + \lambda_2\gamma)(\omega + \gamma)^2} \quad (5.13b)$$

em que $z(\infty)$ é o valor assintótico do índice de dispersão para as VAs de contagem (eq. A.3.14).

Do que é dito a propósito das eqs. (A.5.1), (A.5.5), e da eq. (5.8), podemos determinar a covariância dos números de eventos em intervalos adjacentes, cuja importância é mencionada por Heffes e Lucantoni (1986):

$$C(t) = \frac{\omega\gamma(\lambda_1 - \lambda_2)^2}{(\omega + \gamma)^4} (1 - e^{-(\omega + \gamma)t})^2 \quad (5.14a)$$

O seu valor assintótico vem dado, como é evidente, por:

$$C(\infty) = \omega\gamma(\lambda_1 - \lambda_2)^2 / (\omega + \gamma)^4 \quad (5.14b)$$

Um último resultado que importa aqui salientar, é o de que o SPP tende para o processo de Poisson (de intensidade $\lambda = (\lambda_1 + \lambda_2 k) / (1 + k)$, (5.16)), se fizermos a

frequência de comutação ($f = \omega + \gamma$) tender para ∞ , sem que a razão de fases ($k = \gamma/\omega$) do APP subjacente se altere.

5.4 - Caracterização de Fluxos de Tráfego, Usando o SPP

Convém inicialmente precisar os conceitos de processo canónico, parâmetros definidores de um processo canónico, e parâmetros caracterizadores de um fluxo de tráfego, tendo em vista a sua modelização através de um processo canónico (Rossiter 1986). Um processo canónico é um PP estacionário que pode ser usado, com razoável aproximação, para representar um fluxo de tráfego. O ponto fundamental a ter em consideração, para que um determinado PP seja eleito como canónico, é a facilidade relativa de análise de sistemas estocásticos, com um ou mais servidores, tendo esse processo (canónico) como entrada. Um conjunto U_D de parâmetros definidores de um processo canónico é um conjunto de parâmetros, através do qual o processo fica estabelecido; por exemplo, para o caso do SPP, $U_D = \{\lambda_1, \lambda_2, \omega, \gamma\}$, como facilmente se pode depreender das eqs. apresentadas na secção 5.3. O conjunto U_C de parâmetros caracterizadores de um fluxo, é um conjunto de certas quantidades características do fluxo, usadas para a sua modelização, eventualmente aproximada, através de um processo canónico, isto é, a partir desses parâmetros caracterizadores é possível obter o conjunto U_D , definidor do processo canónico usado na sua modelização. Uma condição essencial que U_C deve respeitar é a de que, se tivermos n fluxos independentes, caracterizados por $U_C(1), \dots, U_C(n)$, então o conjunto U_C , caracterizador do fluxo sobreposição dos anteriores (fluxo composto), deve poder ser obtido exactamente, a partir dos elementos de $U_C(1), \dots, U_C(n)$. Sendo U_C obtido exactamente, não depende da ordem (1 a n) pela qual os fluxos componentes foram ordenados. É costume referir estas propriedades dizendo que o método de modelização deve ser comutativo. Deve igualmente ser associativo.

Um processo, há muito usado como canónico, é o de renovações. O seu uso está subjacente ao método ERT de Wilkinson, para a modelização de fluxos de tráfego de transbordo em redes telefónicas; Whitt (1982) apresenta duas metodologias ('intervalo estacionário' e 'assimptótico') para o seu uso. Aqui o conjunto U_C é $U_C = \{\lambda, c^2\}$, respectivamente a intensidade média do fluxo e o quadrado do coeficiente de variação. Contudo, este método de modelização peca no tocante à modelização de sobreposições de fluxos de tráfego, dada a dificuldade da obtenção exacta do valor de c^2 para o fluxo composto. Por outro lado os intervalos inter-eventos de um processo de renovações são independentes, o que não sucede em muitos fluxos de tráfego.

Heffes (1980), Meier (1984), Sriram e Whitt (1986), Heffes e Lucantoni (1986) e Rossiter (1986) apresentam estudos de sistemas com SPP como processo canónico. Heffes (1980) apresenta um método de modelização de PPDEs, pelo SPP. O conjunto de parâmetros característicos proposto é $U_C = \{m, v, \mu_{3p}, \tau_c\}$, respectivamente, a média, a variância, o 3º momento da intensidade, e uma constante de tempo proporcional ao integral da função covariância do processo de intensidades. São também apresentadas as equações que relacionam estes parâmetros com os definidores ($U_D = \{\lambda_1, \lambda_2, \omega, \gamma\}$), bem como um método para a sua inversão:

$$m = \lambda \quad (5.15a)$$

$$v = \omega\gamma(\lambda_1 - \lambda_2)^2 / (\omega + \gamma)^2 \quad (5.15b)$$

$$\mu_{3p} = (\lambda_1^3\omega + \lambda_2^3\gamma) / (\omega + \gamma) \quad (5.15c)$$

$$\tau_c = 1/f = 1/(\omega + \gamma) \quad (5.15d)$$

Rossiter (1986) apresenta um outro conjunto de parâmetros caracterizadores. É ele $U_C = \{\lambda, c^2, z(\infty), C(\infty)\}$. As eqs. (4.4), (5.11), (5.13b) e (5.14b) constituem o sistema (que aqui fica referenciado por 5.16) que relaciona os elementos de U_C com os de U_D . Para a sua inversão, é usado o conjunto intermediário $U = \{\lambda, c^2, b, v\}$, em que os elementos estão relacionados com os de U_D em (4.4), (5.11), (5.12) e (5.15b). Seguidamente é apresentado um processo de inversão deste último sistema. De notar que este método só é válido no domínio $\{U_C = \{\lambda, c^2, z(\infty), C(\infty)\} \mid \lambda > 0, 1 < c^2 \leq z(\infty), C(\infty) > 0\}$.

A modelização de fluxos compostos é feita do seguinte modo:

Seja $N(t)$ a VA de contagem do fluxo composto, sendo este a sobreposição de n fluxos (componentes) independentes. Ao fluxo componente i corresponde a VA de contagem $N_i(t)$ e o conjunto de parâmetros caracterizadores $U_C(i) = \{\lambda_i, c_i^2, z_i(\infty), C_i(\infty)\}$.

O conjunto $U_C = \{\lambda, c^2, z(\infty), C(\infty)\}$ de parâmetros caracterizadores do fluxo composto é obtido do seguinte modo:

- 1) Pela aditividade das médias:

$$\lambda = \lambda_1 + \dots + \lambda_n \quad (5.17)$$

- 2) Invertendo o sistema (5.16) para o fluxo i , ficamos a conhecer o conjunto definidor do SPP (SPP_i) que o modeliza. Usando a eq. (5.4) ficamos a conhecer a função $P_i^{(1)}(z, s)$ referente a esse fluxo. Fazendo o mesmo para os outros $n - 1$ fluxos componentes ficamos com o conjunto de funções $\{P_i^{(1)}(z, s), i = 1, \dots, n\}$. Ora, é evidente que $N(t) = N_1(t) + \dots + N_n(t)$. Então:

$$P^{(1)}(z, s) = \prod_{i=1}^n P_i^{(1)}(z, s) \quad (5.18)$$

em que a função $P^{(1)}(z, s)$ se refere ao fluxo composto. Usando agora a eq. (A.5.4), ficamos a conhecer o coeficiente de variação (c) do fluxo composto.

- 3) Sendo os fluxos componentes independentes entre si, $\text{var}[N(t)] = \text{var}[N_1(t)] + \dots + \text{var}[N_n(t)]$. Sendo assim, e tendo em consideração (5.17) e a definição de $z(t)$ (Ap. A.3), podemos pôr:

$$z(\infty) = \sum_{i=1}^n \frac{\lambda_i}{\lambda} z_i(\infty) \quad (5.19)$$

- 4) Pelas razões apontadas em 3) e tendo (A.5.5) em consideração, podemos concluir que:

$$C(\infty) = \sum_{i=1}^n C_i(\infty) \quad (5.20)$$

Assim fica conhecido o conjunto U_C , correspondente ao fluxo composto. Entrando agora com os elementos de U_C no sistema (5.16) e invertendo-o, ficamos a conhecer os elementos de U_D do SPP modelizador do fluxo composto.

6 - O PROCESSO DE POISSON MODULADO POR UMA CADEIA DE MARKOV DE m ESTADOS ("m-STATE MARKOV MODULATED PROCESS" m-MMPP)

6.1 - Introdução

Este PP aparece, pela primeira vez, em Neuts (1971), onde são apresentados alguns resultados, e analisado o sistema de serviço MMPP/G/1, onde o tempo de serviço também depende do estado da cadeia de Markov subjacente ao PP de entrada (ver definição do processo). Em Neuts (1981) e Neuts (1984) é apresentada uma metodologia (a metodologia matriz-analítica) que resolve sistematicamente uma ampla classe de sistemas de serviço. Ramaswami (1980) aplica-a ao sistema N/G/1 de que o sistema MMPP/G/1 é um caso particular; os resultados obtidos são, no entanto, de grande custo computacional. Holtzman (1985), e Heffes e Lucantoni (1986) usam-no na modelização de fluxos de tráfego em redes de pacotes de dados e voz; nestes últimos, o sistema de espera MMPP/G/1 é resolvido de forma menos custosa computacionalmente, através da modelização (aproximada) do fluxo de tráfego pelo SPP. Para a modelização de fluxos de tráfego de pacotes, Fontana e Guerrero (1988) apresentam um modelo próximo do MMPP, a que é dado o nome de multi-modo, consistindo na sobreposição de fluxos de Poisson interdependentes; a diferença está em que os instantes de mudança de fase se encontram imediatamente a seguir a eventos (chegadas de pacotes); neste artigo é também sugerido o Processo de Poisson em Lotes ("Batch Poisson"), para modelizar as entradas nos sistemas de transmissão por pacotes em que não haja significativa correlação entre chegadas sucessivas.

6.2 - Definição

O m-MMPP é um processo de Poisson duplamente estocástico em que o processo de intensidades $\{L(t), t > 0\}$ é governado por uma cadeia de Markov, irreduzível, contínua no tempo e com m estados, ou seja:

$$L(t) = \begin{cases} \lambda_1 & \text{se } I(t) = 1 \\ \cdot \\ \cdot \\ \lambda_m & \text{se } I(t) = m \end{cases} \quad (6.1)$$

Em que $I(t)$ indica o estado de uma cadeia de Markov, irreduzível e contínua no tempo, no instante t . Um m-MMPP diz-se na fase i , no instante t , sse $L(t) = \lambda_i$.

Da definição, podemos concluir ser o SPP um caso particular de MMPP, o SPP é o 2-MMPP.

6.3 - Caracterização e Grandezas Pertinentes

Da definição, podemos concluir que um m-MMPP pode ser caracterizado pela matriz Q , geradora infinitesimal da cadeia de Markov subjacente:

$$Q = \begin{bmatrix} -\sigma_1 & \sigma_{12} & \cdot & \cdot & \sigma_{1m} \\ \sigma_{21} & -\sigma_2 & \cdot & \cdot & \sigma_{2m} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{m1} & \sigma_{m2} & \cdot & \cdot & -\sigma_m \end{bmatrix} \quad (6.2)$$

pela matriz diagonal $\Sigma = D(\lambda_1, \lambda_2, \dots, \lambda_m)$, e pelo vector de probabilidades γ , do estado do processo de Markov subjacente, no instante inicial ($t = 0$). Prova-se (Meier 1984) que:

Para o caso de escolha arbitrária, no tempo, da origem do processo:

$$\gamma = \pi \quad (6.3)$$

em que π é o vector estacionário de Q , isto é:

$$\pi : \pi Q = 0, \quad \pi e = 1, \quad e = (1, 1, \dots, 1) \quad (6.4)$$

Para o caso da escolha arbitrária da origem num evento:

$$\gamma = p \quad (6.5)$$

em que $p = \lambda \pi \sum$, sendo $\lambda = \pi \lambda$ a intensidade do processo; $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_m)$ é o vector de intensidades.

Consideremos a sequência $\{(J_n, X_n), n \geq 0\}$, em que a VA J_n indica a fase do evento de ordem n , e a VA X_n $\{n > 0\}$ indica o tempo entre o eventos de ordem $n-1$ e n , sendo $X_0 = 0$. Visto por esta óptica, o MMPP é um Processo de Renovações de Markov ("Markov Renewal Process" – MRP). A sua matriz de distribuições de probabilidades de transição (Pyke 1961) facilmente se conclui ser dada por:

$$F(x) = \int_0^x e^{(Q-\Lambda)t} \Lambda dt, \quad x > 0 \quad (6.6)$$

Os elementos $F_{ij}(x)$ são as probabilidades condicionais:

$$F_{ij}(x) = \text{pr}\{J_n = j, X_n \leq x \mid J_{n-1} = i\} \quad (6.7)$$

6.4 - Algumas Propriedades

Em Kingman (1964), Grandell (1976) e Meier (1984) provam-se as seguintes propriedades do m-MMPP:

6.4.1 - Momentos

Seja $N(t)$ a VA número de eventos em $(0, t]$, e $J(t)$ a fase do processo no instante t . Então:

- a) A matriz $M(t) : M_{ij} = E[N(t), J(t) = j \mid J(0) = i]$, do número médio de eventos em $(0, t]$ é dada por:

$$M(t) = \sum_{n=1}^{\infty} \frac{t^n}{n!} \sum_{m=0}^{n-1} Q^m L Q^{n-1-m} \quad (6.8)$$

- b) O vector $\mu(t) = M(t)e = (E[N(t) \mid J(0) = i])$ é dado por:

$$\mu(t) = \mu \lambda t e + (I - e\pi(e\pi - Q)^{-1} \Lambda e + [e\pi - e^{Qt}](e\pi - Q)^{-1} \Lambda e \quad (6.9)$$

6.4.2 - MMPP – Processo de Renovações

Um MMPP estacionário no tempo, no qual somente há eventos numa das fases é um processo de renovações. Esta propriedade é corolário de um teorema, mais geral, de Kingman (1964), referido em 4.4 a propósito do GIPP.

6.4.3 - Convergência para o Processo de Poisson

Se $p \rightarrow \pi$, então o MMPP converge fracamente para um processo de Poisson de intensidade λ (intensidade do MMPP).

Seja um MMPP, caracterizado por (kQ, Λ) , em que k é conhecido por parâmetro de 'agitação' ("jitterness"). Fazendo $k \rightarrow \infty$, o processo converge fracamente para um processo de Poisson de intensidade λ .

7 - OUTROS TRABALHOS

Nesta secção iremos fazer uma breve referência a outros trabalhos, focando problemas de modelização estocástica de fluxos de tráfego, com aplicação em redes de diferentes tipos.

Oda e Watanabe (1988) propõem métodos para a análise das probabilidades de perda das parcelas individuais de tráfego em redes com comutação por circuitos, e com esquemas de encaminhamento alternativo, sujeitas a fluxos de entrada, constituídos por chamadas de classes diferentes no tocante à velocidade ("baud rate"). Os momentos dos fluxos de transbordo emergentes dos grupos primários são calculados por dois métodos alternativos: um exacto e o outro aproximado. O fluxo (T), oferecido ao grupo secundário final é, assim, constituído pela sobreposição dos vários fluxos de transbordo emergentes dos grupos primários e do fluxo primário directo. São apresentados dois métodos para a análise do sistema fluxo T - grupo secundário final: no primeiro, menos preciso, é encontrado um sistema "equivalente" em que o fluxo é constituído por uma só classe de chamadas; no segundo os vários fluxos são modelizados através de IPPs e o fluxo de sobreposição (que obviamente constitui um m-MMPP) também (método IPP de Kuczura (1977)). Este tipo de problemática, para o caso de sistemas de transbordo mútuo, é abordado por Meier (1989). Aos grupos primários são oferecidos MMPPs, sobreposições de IPPs. Os fluxos marginais de transbordo de cada grupo primário (correspondentes, cada qual, a um IPP oferecido), que são MMPPs, são modelizados de acordo com o método de Heffes (1980) (eqs. 5.15) por SPPs, constatando-se que, afinal, estes SPPs são IPPs. Aos grupos secundários são então oferecidos MMPPs sobreposições destes IPPs de transbordo.

Um modelo para a análise de desempenho, no tocante às suas características de atraso e transbordo, de sistemas de comutação por pacotes, no contexto duma ISDN de banda larga, é apresentado por Kuribayashi e Takahashi (1988). Este estudo utiliza um modelo de fila de espera, designado por $M_1^{X_1}, M_2^{X_2}, \dots, M_n^{X_n}/G/S(NP)$. Cada pacote a ser transmitido é "partido" em "fatias" (células), todas do mesmo tamanho (modo ATD de transmissão). Há assim várias (n) classes de pacotes, consoante o número (X_i) de células em que são "partidos". O fluxo de entrada é, então, a sobreposição de n fluxos, cada qual correspondendo a uma classe. O fluxo correspondente à classe i é considerado como sendo de Poisson em lotes ("batch Poisson"), sendo os lotes de tamanho fixo. Sendo assim, a sua sobreposição é também um processo de Poisson em lotes, só que agora o tamanho dos lotes já não é constante, constituindo um processo estocástico discreto. A chegada simultânea de dois lotes de classes diferente é resolvida, havendo prioridade de entrada na fila de espera, do lote de maior tamanho. Os tempos de ocupação dos S servidores têm uma distribuição comum arbitrária, não havendo (símbolo NP) preemptividade.

ABREVIATURAS (e locais de definição)

- APP – Processo Alternante de Poisson (sec. 3.1)
- ARP – Processo Alternante de Renovações (sec. 3.1)
- BPP – Bernouilli Pascal Poisson
- EARP – Processo Alternante de Renovações em Equilíbrio (sec. 3.1)
- ERT – Equivalent Random Traffic (Tráfego de Azar Equivalente)
- FD – Função de Distribuição
- GIPP – Processo de Poisson Interrompido Generalizado (sec. 4.4)
- IPP – Processo de Poisson Interrompido (sec. 4.4)
- MARP – Processo Alternante de Renovações Modificado (sec. 3.1)

- MMPP – Processo de Poisson Modulado por uma Cadeia de Markov (sec. 6.2)
 MRP – Processo de Renovações de Markov ("Markov Renewal Process")
 OARP – Processo Alternante de Renovações Ordinário (sec. 3.1)
 pdf – Função Densidade de Probabilidade
 PP – Processo Pontual (sec. 2.1)
 PPDE – Processo de Poisson Duplamente Estocástico (sec. 4.2)
 PR – Processo de Renovações (sec. 2.4)
 RHPP – "Random Hazard Poisson process" (sec. 4.2)
 SPP – Processo de Poisson Comutado (sec. 5.2)
 TL – Transformada de Laplace
 TLS – Transformada de Laplace-Stieljes
 VA – Variável aleatória

BIBLIOGRAFIA E REFERÊNCIAS

- [1] Asmussen, S. (1987): Applied Probability and Queues. John Wiley and Sons
 [2] Bauer, H. (1981): Probability Theory and Elements of Measure Theory. Academic Press
 [3] Bocker, P. (1987): ISDN, The Integrated services Digital Network. Springer-Verlag
 [4] Cox, D. R. (1962): Renewal Theory. Methuen London
 [5] Cox, D. R.; Isham, V. (1980): Point Processes. Chapman and Hall
 [6] Delbrouck, L. E. N. (1981): A Unified Approximate Evaluation of Congestion Functions for Smooth and Peaky Traffics. IEEE Transactions on Communications, Vol. Com-29, nº 2, Feb. 1981
 [7] Delbrouck, L. E. N. (1983): The Uses of kosten's Systems in the Provisioning of alternate Trunk Groups Carrying Heteroneous Traffic. IEEE Transactions on Communications, Vol. Com-31, nº 6, Jun. 1983
 [8] Delbrouck, L. E. N. (1983): On the Steady-State Distribution in a Service Facility Carrying Mixtures of Traffic with Different Peakedness Factors and Capacity Requirements. IEEE Transactions on Communications, Vol. Com-31, nº 11, Nov. 1983
 [9] Eisen, M.; Trainer, M. (1983): Stochastic Variations in Queueing Processes. Opns. Res., 11, pp922-927
 [10] Gaver, D. P. (1983): Random Hazard in Reliability Problems. Technometrics, 5, pp211-226
 [11] Grandel, J. (1976): Doubly stochastic Poisson Processes. Springer-Verlag
 [12] Halmos, P. R. (1950): Measure Theory. D. Van Nostrand Company, inc
 [13] Heffes, H. (1980): A Class of data Traffic Processes - Covariance Function Characterization and Related Queueing Results. Bell Syst. Tech. J., 59, pp897-929
 [14] Heffes, H.; Lucantoni, D. M. (1986): A Markov Modulated Characterisation of Packetised Voice and data Traffic and Related Statistical Multiplexer Performance. IEEE Select. Areas Commun., SAC-4, pp856-868
 [15] Holtzman, J. M. (1986): Characteristics of Superpositions of Traffic Streams. Brussels Specialist Seminar on ISDN Traffic Issues, Brussels
 [16] Kingman, J. F. C. (1964): On Doubly Stochastic Poisson Processes. Royal Cambridge Phil. Soc., 60, pp 547-567
 [17] Kingman, J. F. C. (1964): The Stochastic Theory of Regenerative Events. Z. Wahrscheinlichkeitstheorie, 2, pp 180-224, 1964

- [18] Kingman, J. F. C. (1965): Linked Systems of Regenerative Events. Proc. London Math. Soc., 15, pp 125-150
- [19] Kopinska, I. (1984): The Reliability of an Element with Alternating Failure. Rate. Zast. Mat., 18, pp187-194
- [20] Kuczura, A. (1973): The Interrupted Poisson Process as an Overflow Process. Bell Syst. Tech. J., 52, pp437-448
- [21] Kuczura, A.; Bajaj, D. (1977): A Method of Moments for the Analysis of a Switched Communication Network's Performance. IEEE Trans. Comm., COM-25, nº2, pp185-193, Feb. 1977
- [22] Kuribayashi, S.; Takahasi, Y. (1988): Performance Analysis of ISDN Packet Switching Systems. Proc. ITC 12, Torino, pp5.1A.2.1-5.1A.2.7, June 1989
- [23] Lawrence, A. J. (1972): Some Models for Stationary Series of Univariate Events. Stochastic Processes, ed. P.A.W. Lewis, Wiley, New York, pp199-256
- [24] Meier, M. F.(1984): A Statistical Procedure for Fitting Markov Modulated Poisson Processes. Ph. D. Dissertation, Univ. Delaware
- [25] Meier, M. F.(1989): The Analysis of a Queue Arising in Overflow Models. IEEE Trans. Comm., vol 37, nº 4, pp367-372, April 1989
- [26] Minzer, S. E. (1989): Broadband ISDN and Asynchronous Transfer Mode (ATM). IEEE Comm. Magazine, September 1989, pp17-57
- [27] Neuts, M. F. (1971): A Queue Subject to Extraneous Phase Changes. Adv. Apl. Prob., 3, pp78-119
- [28] Neuts, M. F. (1981): Matrix Geometric Solutions in Stochastic Models: An Algorithmic Approach. The John Hopkins University Press, Baltimore
- [29] Neuts, M. F. (1984): Matrix-Analytic Methods in Queuing Theory. Eur. Jour. Oper. Res., 15 (1984), pp2-12
- [30] Oda, T.; Watanabe, Yu (1988): A Method for Analysing Circuit-Switched Networks with Multiple Bit Rate Classes. ITC-12, Torino, pp5.1A.6.1-5.1A.6.7, June 1988
- [31] Ramaswami, V. (1980): The N/G/1 Queue and its Detailed Analysis Adv. Apl. Prob.,12, pp222-261
- [32] Regtershot, G. J. K.; Smit, J. H. A. (1986): The Queue M/G/1 with Markov Modulated Arrivals and Services. Math. O. R., 11, pp465-483
- [33] Rossiter, M. H. (1986): Sojourn Time Theory and the Switched Poisson Process. Research Laboratories, Telecom Australia, Report 7835, 1986
- [34] Sims, B. T. (1976): Fundamentals of Topology. Macmillan Publ.
- [35] Sriram, K. and Whitt, W. (1986): Characterising Superposition Arrival Processes in Packet Multiplexers for Voice and Data. IEEE J. Sel. Areas. SAC-4, pp833-846
- [36] Takacs, L. (1957): On Certain Sojourn Time Problems in the Theory of Stochastic Processes. Acta. math. Sci. Hung., vol 8, pp169-191
- [37] Takacs, L. (1962): Introduction to the Theory of Queues. Oxford University Press
- [38] Tran-Gia, P. (1983): A Renewal Approximation for the Generalized Switched Poisson Process. J. R. Statist. Soc. B, 24, pp364-382
- [39] Van Hoorn, M. H.; Seelen, L. P. (1983): The SPP/G/1 Queue: A single server Queue with a Swiched Poisson Process as Input Process. O. R. Spectrum, 5, pp205-218
- [40] Whitt, W. (1982): Approximating a Point Process by a Renewal process I: Two Basic Methods. Opns. Res., 30, pp125-147
- [41] Wilkinson, R. I. (1956): Theory of Toll Traffic Engineering in the U.S.A.. Bell System Tech. J., vol. 35, nº 3, pp421-514

- [42] Yechiali, U.; Noar, P. (1971): Queuing Problems with Heterogeneous arrivals and Service. *Opns. Res.*, 19, pp722-734
- [43] Yoki, T.; Kodaira, K. (1989): Grade of Service in the ISDN Era. *IEEE Comm. Magazine*, April 1989, pp46-51

APÊNDICE - Algumas Propriedades e resultados sobre Processos Pontuais

A.1 - Distribuição de Palm

Consideremos um PP simples e estacionário, e seja f_X a função sobrevivência do intervalo X , definida por:

$$\mathfrak{F}_X(t) = \text{pr}\{X > t\} = \lim_{\partial \rightarrow 0^+} \text{pr}\{N(t) = 0 \mid N(-\partial, 0) > 0\} \quad (\text{A.1.1})$$

prova-se (Cox 80) que:

$$\mathfrak{F}_X(t) \cdot v = -\frac{d}{dt} p_0(t) \quad (\text{A.1.2})$$

em que " v é o parâmetro de ocorrência do PP :

$$v := \lim_{\partial \rightarrow 0^+} \frac{\text{pr}\{\partial > 0\}}{\partial} \quad (\text{A.1.3a})$$

$$e: \quad p_k(t) := \text{pr}\{N(t) = k\} \quad (\text{A.1.3b})$$

Para um processo simples e estacionário,

$$v = \lambda \quad (\text{A.1.4a})$$

sendo λ a intensidade do processo:

$$\lambda := \lim_{\partial \rightarrow 0^+} \frac{E[N(\partial)]}{\partial}$$

Se o processo for múltiplo e as VAs M_i , definidas em (2.3), forem independentes e identicamente distribuídas ($M_i = M$) (i.i.d.), então, pelo teorema de Korolyuk:

$$\lambda = v \cdot E[M] \quad (\text{A.1.4b})$$

Define-se a distribuição de Palm, como sendo a distribuição discreta:

$$\pi_k(t) = \lim_{\partial \rightarrow 0^+} \text{pr}\{N(t) = k \mid N(-\partial, 0) > 0\} \quad (\text{A.1.5})$$

podendo provar-se as equações de Palm-Khintchine:

$$\frac{d}{dt} p_k(t) = -\lambda(\pi_k(t) - \pi_{k-1}(t)) \quad (\text{A.1.6})$$

$$\frac{d}{dt} p_0(t) = -\lambda(\pi_0(t))$$

de (A.1.6) temos que:

$$-\frac{1}{\lambda} \frac{d}{dt} \left(\sum_0^k p_i(t) \right) = -\frac{1}{\lambda} \frac{d}{dt} \text{pr}\{N(t) \leq k\} = \pi_k(t) \quad (\text{A.1.7})$$

de (2.6):

$$\text{pr}\{N(t) \leq k\} = 1 - \text{pr}\{T_{k+1} \leq t\}$$

designando por F_X a função de distribuição (FD) da VA X , temos:

$$F_{N(t)}(k) = 1 - F_{T_{k+1}}(t) \quad (\text{A.1.8})$$

e, então:

$$\frac{d}{dt} F_{N(t)}(k) = -\frac{d}{dt} F_{T_{k+1}}(t) = -\lambda \pi_k(t) \quad (\text{A.1.9})$$

Passaremos a designar F_{T_k} por F_k . De (A.1.9) e (A.1.6), e tendo em consideração

que $k \geq 1 \Rightarrow F_k(0) = 0$, e $F_0(t) = 1$, vem que (com $k = 0, 1, \dots$):

$$p_k(t) = F_k(t) - F_{k+1}(t) \tag{A.1.10}$$

de (A.1.9) temos também que:

$$dF_k(t) = \lambda \pi_{k-1}(t) dt \Rightarrow \int_0^{+\infty} e^{-st} dF_k(t) = \lambda \int_0^{+\infty} e^{-st} \pi_{k-1}(t) dt$$

designando por $f^{(l)}(s)$ a transformada de Laplace (TL) da função f , e por $f^{(ls)}(s)$ a transformada de Laplace-Stieljes (TLS) de f , vem:

$$F_k^{(ls)}(s) = \lambda \pi_{k-1}^{(l)}(s), k \geq 1 \tag{A.1.11}$$

mas, de (A.1.7), para $k \geq 1$:

$$-\lambda \pi_{k-1}^{(l)}(s) = s \sum_0^{k-1} p_i^{(l)}(s) - \sum_0^{k-1} p_i(0)$$

ora:

$$p_i(0) = \text{pr}\{N(0) = i\} = \begin{cases} 1 & \text{se } i = 0 \\ 0 & \text{se } i > 0 \end{cases}$$

então, para $k \geq 1$:

$$f_{k-1}^{(l)}(s) = 1 - s \sum_{i=0}^{k-1} p_i^{(l)}(s) \tag{A.1.12}$$

ou:

$$\left[f_k^{(l)}(s) = 1 - s \sum_0^{k-1} p_i^{(l)}(s) \right] \tag{A.1.13}$$

Usando funções geradoras de probabilidade temos:

$$P(z, t) := \sum_0^{+\infty} z^i p_i(t) \quad \Pi(z, t) := \sum_0^{+\infty} z^i \pi_i(t)$$

e, então, (A.1.7) fica:

$$\frac{d}{dt} P(z, t) = -\lambda(1 - z) \Pi(z, t) \tag{A.1.14}$$

ou, usando as TL das funções geradoras:

$$sP^{(l)}(z, s) - 1 = -\lambda(1 - z) \Pi^{(l)}(z, s) \tag{A.1.15}$$

A.2 - Processo dos intervalos

Analizando o PP pelo prisma do processo dos intervalos (2.5), teremos, admitindo que as VAs X_i são mutuamente independentes:

$$T_k = \sum_0^{k-1} X_i \Rightarrow F_k(t) = F_{x_1} * F_{x_2} * \dots * F_{x_k}$$

em que F_{X_i} é a FD da VA X_i , e * indica "convolução". Passando a TLSs:

$$F_k^{(ls)}(s) = \prod_1^k F_{x_i}^{(ls)}(s) \tag{A.2.1}$$

A.3 - Momentos

A.3.1 - Processo $\{N(t)\}$

Seja $M(t) := E[N(t)]$ a média do PP, em função do tempo. Define-se intensidade local do processo, como sendo a função:

$$\lambda(t) := \lim_{\delta \rightarrow 0^+} E[N(t, t + \delta)] \tag{A.3.1}$$

É evidente que, sendo o processo estacionário, $E[N(t, t + \delta)] = E[N(\delta)]$ e então $\lambda(t) = \lambda$. Utilizando a aditividade das médias vem:

$$M(t) = \int_0^t \lambda(\tau) d\tau \quad (\text{A.3.2a})$$

e, de (A.1.10):

$$M(t) = \int_0^\infty F_k(t) \quad (\text{A.3.2b})$$

$$M^{(ls)}(t) = \int_0^\infty F_k^{(ls)}(t) \quad (\text{A.3.2b})$$

então, sendo o processo estacionário:

$$M(t) = \lambda t \quad (\text{A.3.3})$$

generalizando:

$$E[N(A)] = \lambda \mu(A) \quad (\text{A.3.4})$$

em que $\mu(A)$ é a medida (de Lebesgue) do boreliano A.

Define-se a função intensidade condicional de processo, do seguinte modo:

$$h(t) = \lim_{\substack{\delta_1 \rightarrow 0^+ \\ \delta_2 \rightarrow 0^+}} \frac{\text{pr}\{N(t, t + \delta_2) > 0 \mid N(-\delta_1, 0) > 0\}}{\delta_2} \quad (\text{A.3.5})$$

Num PP simples e estacionário:

$$\lim_{t \rightarrow +\infty} h(t) = \lambda \quad (\text{A.3.6})$$

Demonstra-se (Cox 80) que a variância em função do tempo ($V(t) := \text{var}[N(t)]$), é dada por:

$$V(t) = \int_0^t \int_0^t c(u - z) du \quad (\text{A.3.7})$$

em que a função $c(x)$ (densidade de covariância) é dada por:

$$c(x) = \lambda \delta(x) + \lambda h(x) + \lambda^2 \quad (\text{A.3.8})$$

sendo $\delta(x)$ a função delta de Dirac.

A covariância entre as VAs de contagem em dois borelianos arbitrários A e B, vem dada por (Cox 80):

$$\begin{aligned} \text{cov}[N(A), N(B)] &= \int_A dz \int_B c(u - z) du = \lambda \mu(A \cap B) + \\ &+ \lambda \int_A \int_B h(u - z) du - \lambda^2 \mu(A) \mu(B) \end{aligned} \quad (\text{A.3.9})$$

A.3.2 - Processo $\{X_i\}$

Vamos aqui considerar um PP com estacionaridade na sequência $\{X_i\}$ dos intervalos. De acordo com o discutido em 2.4, isto corresponde a termos um PP estacionário abordado de acordo com a condição evento na origem. Prova-se que (ver eq. A.3.5), nestas condições:

$$E[X_i] = 1/\lambda = E[X], \quad i = 0, 1, \dots \quad (\text{A.3.10})$$

Define-se o coeficiente de correlação de ordem $k > 0$, ρ_k , como sendo:

$$\rho_k := \text{corr}[X_i, X_{i+k}] = \text{corr}[X_1, X_{1+k}] := \frac{\text{cov}[X_1, X_{1+k}]}{s_{x_1} s_{x_{1+k}}} \quad (\text{A.3.11})$$

$i \geq 1$. De notar que $X_i = X$, $\forall i \in \mathbb{N} \Rightarrow \rho_k$ constante.

Consideremos agora a sequência de VAs $\{S_k\}$, $S_0 = 0$ (ver 2.4). Define-se o índice de dispersão Z_k por:

$$Z_k := \frac{\text{var}[S_k]}{k(\text{E}[X])^2} \quad (\text{A.3.12})$$

Prova-se que, se $\text{var}[X]$ é finita e $\text{var}[X] = c^2/\lambda^2$ (sendo c o coeficiente de variação) se tem:

$$\text{var}[S_k] = \frac{c^2}{\lambda^2} [k + 2 \sum_{i=1}^{k-1} (k-i)\rho_i] \quad (\text{A.3.13})$$

De (A.3.12) e (A.3.13) pode verificar-se que $Z_k = c^2$, para todo $k \geq 1$, sse $\rho_k = 0$, $k \geq 1$. Z_k está relacionado com o índice de dispersão ($Z(t) := V(t)/M(t)$) das VAs de contagem, pela relação assintótica:

$$\lim_{t \rightarrow +\infty} Z(t) = \lim_{k \rightarrow +\infty} Z_k \quad (\text{A.3.14})$$

A.4 - Outras relações e grandezas

Mais uma vez, vamos considerar o PP estacionário.

De (A.1.9) e (A.1.10) tiramos a importante relação:

$$\frac{d}{dt} F_{T_{n+1}}(t) = f_{T_{n+1}}(t) = \lambda(F_{S_n}(t) - F_{S_{n+1}}(t)) \quad (\text{A.4.1})$$

integrando por partes, conclui-se que:

$$E[T_{n+1}^{r-1}] = \frac{\lambda}{r} E[S_{n+1}^r - S_n^r], \quad r \geq 2, n \geq 0 \quad (\text{A.4.2})$$

em particular, fazendo $r = 2$, podemos obter:

$$E[T_{n+1}] = \frac{1+c^2}{2\lambda} + \frac{c^2}{\lambda} \sum_{i=1}^n \rho_i + \frac{n}{\lambda}, \quad n \geq 1 \quad (\text{A.4.3})$$

$$E[X_0] = E[T_1] = (1 + c^2)/(2\lambda)$$

como $T_{n+1} = X_0 + \dots + X_n$, para o caso de escolha arbitrária da origem, no tempo, e de (2.5b):

$$E[X_n] = (1 + c^2 \rho_n)/\lambda, \quad n \geq 1 \quad (\text{A.4.4})$$

A.5 - Importância de $P^{(l)}(z, s)$

A TL da função geradora de probabilidade da distribuição $\{p_k(t)\}$ de um PP estacionário resume grande parte da informação correspondente ao processo. Resulta isto do facto de que o momento factorial de ordem r ($E[(X)_r] := E[X(X-1)\dots(X-r+1)]$, $r \geq 1$) de uma VA discreta (X) ser igual à derivada de ordem r da sua função geradora de probabilidade, no ponto $z = 1$; e do facto de que o momento simples de ordem r de uma VA contínua, ser igual ao produto de $(-1)^r$ pela derivada de ordem r da TLS da sua FD, no ponto $s = 0$.

Então, da definição da distribuição $\{p_k(t)\}$, e do acima dito:

$$\mathfrak{E}\{E[(N(t))_r]\} = \frac{\partial^r P^{(1)}(z,s)}{\partial z^r} \Big|_{z=1} \quad (\text{A.5.1})$$

Achando as transformadas inversas de (A.5.1), encontraremos os momentos factoriais da VAs $N(t)$, e daqui todos os outros. Em particular, sendo o processo estacionário, $E[N(t)] = \lambda t$; então de (A.5.1) com $r = 1$:

$$\lambda = s^2 \frac{\partial P^{(1)}(z,s)}{\partial z} \Big|_{z=1} \quad (\text{A.5.2})$$

Por outro lado, de (A.1.13):

$$\begin{aligned} F_{k+1}^{(1s)}(s) - F_k^{(1s)}(s) &= 1 - s \sum_0^k p_i^{(1)}(s) - 1 + s \sum_0^{k-1} p_i^{(1)}(s) \\ &= -s p_k^{(1)}(s) = -s \frac{1}{k!} \frac{\partial^k P^{(1)}(z,s)}{\partial z^k} \Big|_{z=0} \end{aligned}$$

Então:

$$\begin{aligned} E[X_k] &= E[T_{k+1} - T_k] = -\frac{d}{ds} (F_{k+1}^{(1s)}(s) - F_k^{(1s)}(s)) \Big|_{s=0} = \\ &= \frac{d}{ds} (s p_k^{(1)}(s)) \Big|_{s=0} = \frac{1}{k!} \frac{\partial^{(k)} P^{(1)}(z,0)}{\partial z^k} \Big|_{z=0} \end{aligned} \quad (\text{A.5.3})$$

(convencionam-se que a derivada de ordem zero de uma função é a própria função).

De (A.4.3) e (A.5.3), com $k = 0$:

$$c^2 = 2\lambda P^{(1)}(0, 0) - 1 \quad (\text{A.5.4})$$

(A.5.3) e (A.4.4) podem ser conjuntamente usados para a obtenção dos coeficientes de correlação $\{\rho_n\}$.

A covariância do número de eventos em intervalos adjacentes ($C(t) = \text{cov}[N(t), N(2t) - N(t)]$, $t > 0$) também pode ser determinada através de $P^{(1)}(z, s)$, tendo em consideração que, da definição podemos tirar que:

$$C(t) = (\text{var}[N(2t)] - \text{var}[N(t)])/2 \quad (\text{A.5.5})$$

A COMPLEMENTARIDADE LINEAR NA RESOLUÇÃO DE ESTRUTURAS COM APOIOS UNILATERAIS

Fernanda M. Pires
Universidade do Algarve

A.A.Serra Neves
Faculdade de Engenharia,
Universidade do Porto

Resumo: Apresenta-se uma aplicação do problema linear complementar à resolução duma estrutura elástica linear sob a acção de forças exteriores sofrendo pequenas deformações na presença de apoios unilaterais rígidos e sem atrito. Obtêm-se vários problemas lineares complementares com a mesma matriz e diferentes termos independentes, em cujas soluções as variáveis positivas são praticamente as mesmas embora com diferentes valores. Usa-se um algoritmo pivotal principal que, contrariamente aos métodos tradicionais de restrições activas, pode começar com qualquer solução básica. Essa facilidade permite usar como solução inicial dum problema a solução do problema anterior. Apresenta-se experiência computacional que mostra a vantagem de tal facilidade.

1 - Introdução

Nos problemas de análise de estruturas de Engenharia Civil a situação mais corrente corresponde a conhecer-se completamente as condições de contorno da estrutura, traduzidas no conhecimento dos valores de algumas variáveis envolvidas nessas zonas. Pode no entanto surgir a situação de à partida não se conhecer a extensão das próprias zonas de contacto, devido ao facto de existirem variáveis relativas a tais regiões com restrição de sinal. É o chamado problema do contacto unilateral, problema não linear apesar dos meios estruturais em estudo serem elásticos lineares e as deformações serem infinitesimais. Por outro lado nos problemas de análise estrutural correntes no domínio da Engenharia Civil a mesma estrutura tem que ser estudada para as diversas combinações de

do outro, relacionam-se entre si pelas chamadas relações constitutivas do material estrutural.

2.1 - Equações de Equilíbrio e Compatibilidade

Aplicando a metodologia do Método dos Elementos Finitos [3] aos campos de tensão e deformação, é possível obter a partir das condições diferenciais de equilíbrio

$$-b(x) = D^T \delta(x) \quad (2)$$

e de compatibilidade

$$\epsilon(x) - \epsilon_0(x) = D u(x) \quad (3)$$

relações algébricas envolvendo um número finito de parâmetros, as variáveis discretas nos pontos nodais

$$\epsilon^c - \epsilon_0^c = C u^\alpha \quad (4)$$

Além disso

$$0 = f^\alpha = C^T \delta^c + f_0^\alpha \quad (5)$$

onde D é um operador diferencial auto-adjunto de 1ª ordem,

ϵ^c - vector das demonstrações generalizadas

ϵ_0^c - vector das demonstrações generalizadas iniciais

δ^c - vector das tensões generalizadas

u^α - vector dos deslocamentos nodais

f^α - vector das forças nodais equivalentes

e C é um operador diferencial relacionando os deslocamentos nodais com as deformações generalizadas. Nas relações (4) e (5) é suposto terem sido contempladas as condições fronteira que não digam respeito às zonas de contacto unilateral.

2.2 - Relações Constitutivas

Tendo por hipótese o material um comportamento elástico linear, os campos de deformações e de tensões relacionam-se entre si em cada ponto através da Lei de Hooke Generalizada

$$\delta(x) = R(x) \epsilon(x) \quad (6)$$

onde $R(x)$ representa a matriz de rigidez do material estrutural no ponto x .

Aplicando o processo de discretização do Método dos Elementos Finitos [3] obtém-se a forma discreta daquelas relações

$$\delta^c = R^c \epsilon^c \quad (7)$$

acções que sobre ela podem actuar, previstas nos regulamentos da especialidade e que se traduzem numa alteração parcial nas condições fronteira. É suposto não existir atrito nas ligações unilaterais, pelo que a condição de contacto pode ser expressa unicamente em função dos deslocamentos.

Os exemplos apresentados dizem respeito a meios contínuos, estados planos de tensão e lajes, pelo que se torna necessário recorrer a um processo de discretização estrutural. Adoptou-se a formulação dos deslocamentos do Método dos Elementos Finitos.

O sistema governativo do problema em análise contendo as relações de equilíbrio, compatibilidade, constitutivas e fronteira identifica-se com um conjunto de Problemas Lineares Complementares Generalizados (GLCP) cada um correspondente ao termo independente associado às diversas combinações de acções consideradas. Pretende-se pois resolver problemas do tipo:

$$w^k = q^k + Mz^k, \quad z_j^k \geq 0, \quad w_j^k \geq 0, \quad w_I^k = 0, \quad z_i^k w_i^k = 0, \quad i \in J \quad (1)$$

com q^k , $k = 1, 2, \dots$, vectores dados, M uma matriz quadrada de ordem n também dada e $I = \{i \in \{1, \dots, n\} : z_i \text{ sem restrição de sinal}\}$, $J = \{1, \dots, n\} - I$

Usa-se um algoritmo que tem a vantagem de poder partir de qualquer solução básica, o que é particularmente importante no caso presente. Com efeito para cada caso estudado resolvem-se diferentes GLCP da forma (1) com a mesma matriz cujas soluções apresentam conjuntos de variáveis básicas semelhantes. Assim, a base da solução de um GLCP pode ser utilizada como base inicial para o GLCP seguinte com economia no número de iterações necessário para a sua solução. Apresentam-se alguns resultados que atestam a eficiência do algoritmo escolhido.

A estrutura do artigo é a seguinte. A descrição do modelo aparece nas secções 2 e 3. Na secção 4 é apresentado o método de resolução. Finalmente a resolução de dois casos concretos é discutida na última secção do artigo.

2 - Relações Fundamentais

O objetivo da análise estrutural é fundamentalmente conhecer os campos de tensão $\delta(x)$, de deformação $\epsilon(x)$ e deslocamentos $u(x)$ verificados no meio estrutural quando sujeito às acções aplicadas.

O estado de tensão é caracterizado pelas variáveis estáticas que se relacionam entre si por relações que traduzem o equilíbrio, e o estado de deformação é caracterizado pelas variáveis cinemáticas que se relacionam entre si por expressões que traduzem a compatibilidade. Estes estados, embora possam ser descritos de forma independente um

2.3 - Relações na Superfície de Contacto Unilateral

Para cada ponto da superfície de contacto unilateral há ainda a considerar as seguintes restrições:

- a) A componente normal dos deslocamentos na superfície de contacto $u_n(x)$ só pode apresentar valores não negativos

$$u_n(x) \geq 0 \quad (8)$$

- b) A componente normal da tensão na superfície de contacto não pode ser de tracção

$$\delta_n(x) \geq 0 \quad (9)$$

- c) A tensão na superfície de contacto unilateral só se desenvolve quando há contacto

$$u_n(x) \delta_n(x) = 0 \quad (10)$$

Passando estas relações à forma discreta, obtém-se:

$$\begin{aligned} u_n &\geq 0 \\ f_n &\geq 0 \\ u_n^T f_n &= 0 \end{aligned} \quad (11)$$

onde, u_n e f_n são respectivamente deslocamentos e forças nodais orientadas perpendicularmente à superfície de contacto unilateral.

3 - Sistema Governativo

Associando as relações de equilíbrio, compatibilidade, constitutivas e de contacto unilateral, obtém-se:

$$\begin{aligned} \epsilon^c - \epsilon_0^c &= C u^\alpha & \text{a)} \\ 0 &= f^\alpha = C^T \delta^c + f_0 & \text{b)} \\ \delta^c &= R^c \epsilon^c & \text{c)} \\ u_n &\geq 0 & \text{d)} \\ f_n &\geq 0 & \text{e)} \\ u_n^T f_n &= 0 & \text{f)} \end{aligned} \quad (12)$$

onde se pode constatar a linearidade de todas as relações, com excepção da relação de complementaridade (12, f).

Admitindo, por simplicidade, que a orientação da zona de apoio unilateral é tal que u_n e f_n são subvectores de u^α e f^α ,

$$u^\alpha = \begin{bmatrix} u \\ u_n \end{bmatrix} \quad (13) \quad f^\alpha = \begin{bmatrix} 0 \\ f_n \end{bmatrix} \quad (14)$$

e procedendo à substituição da expressão (12, a) em (12, c) e desta em (12, b), obtém-se:

$$\begin{aligned} 0 &= E_{11} u + E_{12} u_n + f_0 & \text{a)} \\ f_n &= E_{21} u + E_{22} u_n + f_{0n} & \text{b)} \\ u_n \geq 0, f_n \geq 0, u_n^T f_n &= 0 & \text{c)} \end{aligned} \tag{15}$$

onde

$$E = C^T R^C C \tag{16}$$

representa a matriz de rigidez da estrutura. Essa matriz é simétrica positiva definida e, por consequência, E_{11} e E_{22} também o são.

O sistema governativo assim obtido constitui um Problema Linear Complementar Generalizado (GLCP). Como a matriz E é simétrica positiva definida (pois C é de característica completa), então o GLCP tem solução única [1]. Uma vez resolvido o GLCP (15), obtém-se o vector u que caracteriza o campo de deslocamentos de todo o meio estrutural. Se se deseja determinar o campo de tensões, basta substituir os valores encontrados nas relações (12, a) e (12, c).

4 - Resolução do Problema Linear Complementar

O Problema Linear Complementar consiste em determinar vectores z e w de R^n tais que

$$w = q + Mz, w \geq 0, z \geq 0, w^T z = 0 \tag{17}$$

em que $q \in R^n$ e $M \in R^{n \times n}$ são dados.

O Problema Linear Complementar Generalizado (GLCP) aparece quando algumas componentes de z não têm restrição de sinal. Considerando os conjuntos I e J já definidos atrás, o GLCP consiste em determinar vectores

$$z = \begin{bmatrix} z_I \\ z_J \end{bmatrix} \quad \text{e} \quad w = \begin{bmatrix} w_I \\ w_J \end{bmatrix}$$

tais que

$$\begin{aligned} w &= q + Mz \\ z_J \geq 0, w_J \geq 0, z_J^T w_J &= 0, w_I = 0 \end{aligned}$$

Neste problema específico será:

$$z = u^\alpha, w = f^\alpha, M = E, q = \begin{bmatrix} f_0 \\ f_{0n} \end{bmatrix}$$

Define-se o conjunto das inadmissibilidades $H = \{i \in I : z_i < 0 \text{ ou } w_i < 0\}$. É claro que quando for $H = \emptyset$ se terá a solução do GLCP em causa, uma vez que as variáveis z_i são sem restrição de sinal.

Para a resolução deste GLCP usou-se o algoritmo bloco tipo Bard descrito em [1]. Esse algoritmo utiliza em cada iteração soluções básicas do GLCP. À semelhança da programação linear uma tal solução é constituída por variáveis básicas e não básicas e tem associado dois conjuntos F e T tais que

$$F = \{i : z_i \text{ é básica}\}, T = \{i : w_i \text{ é básica}\}$$

e

$$F \cap T = \emptyset, F \cup T = \{1, \dots, n\}$$

As variáveis não básicas $w_i, i \in F$ e $z_i, i \in T$ são nulas. Além disso as variáveis básicas $z_i, i \in F$ e $w_i, i \in T$ têm valores \bar{q}_i , onde \bar{q} é calculado a partir da resolução do sistema

$$M_{FF} \bar{q}_F = -q_F \quad (18)$$

e de

$$\bar{q}_T = q_T + M_{TF} \bar{q}_F \quad (19)$$

O algoritmo utilizado pode ser esquematizado nos seguintes passos:

Passo inicial: $F = I; T = J$

Passo geral: Calcular \bar{q}_F e \bar{q}_T a partir de (18) e (19). Seja $H = H_1 \cup H_2$, onde

$$H_1 = \{i \in F : \bar{q}_i < 0\}; H_2 = \{i \in T : \bar{q}_i < 0\}$$

a) Se $H = \emptyset$ pára. A solução é

$$z_F = \bar{q}_F, z_T = 0, w_F = 0, w_T = \bar{q}_T$$

b) Caso contrário faz $F = F - H_1 \cup H_2, T = T - H_2 \cup H_1$ e repete o passo geral.

Este algoritmo é heurístico, isto é, não é possível estabelecer teoricamente a sua convergência. A prática mostra que para evitar possíveis entradas em ciclo é suficiente aceitar apenas trocar uma única variável básica com uma não básica do mesmo índice nos casos em que o número de elementos de H é muito pequeno.

5 - Aplicações

Com base na formulação apresentada foi elaborado um código de cálculo automático de análise de lajes e estados planos de tensão.

O código foi desenvolvido para uma discretização da estrutura em elementos finitos isoparamétricos de 8 nós, sendo as tensões e deformações generalizadas localizadas em pontos de Gauss correspondentes à integração numérica reduzida. Para as lajes foi adoptada a teoria de Mindlin [2].

Ensaiou-se um problema plano de tensão correspondente a uma viga de grande altura apoiada numa superfície plana através de contacto unilateral, sendo solicitado por 3 forças concentradas de igual valor. A malha de elementos finitos utilizada bem como a configuração da deformada são apresentados na figura 1.

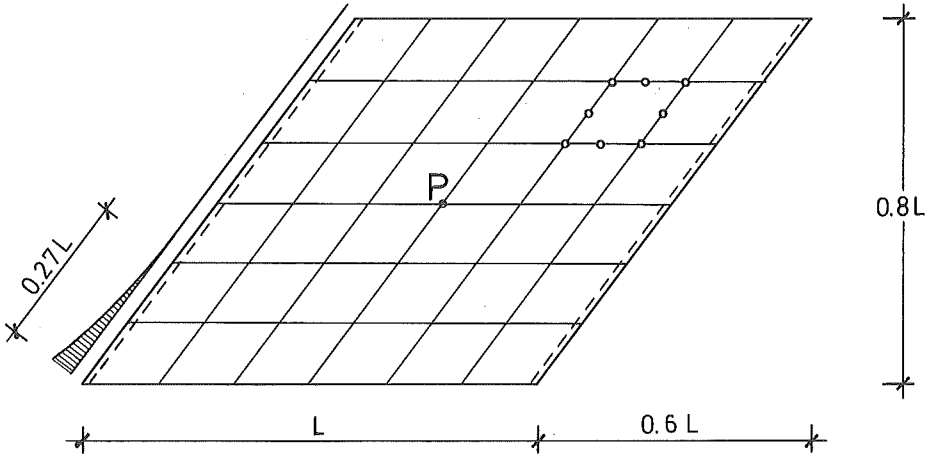


Fig. 1 - Estado plano de tensão

Ensaiou-se depois uma laje enviesada apoiada em dois bordos opostos unilaterais. A solicitação foi constituída por uma carga concentrada aplicada no centro da laje. Nestas condições foi possível determinar a extensão dos cantos levantados da laje, como se representa na figura 2.

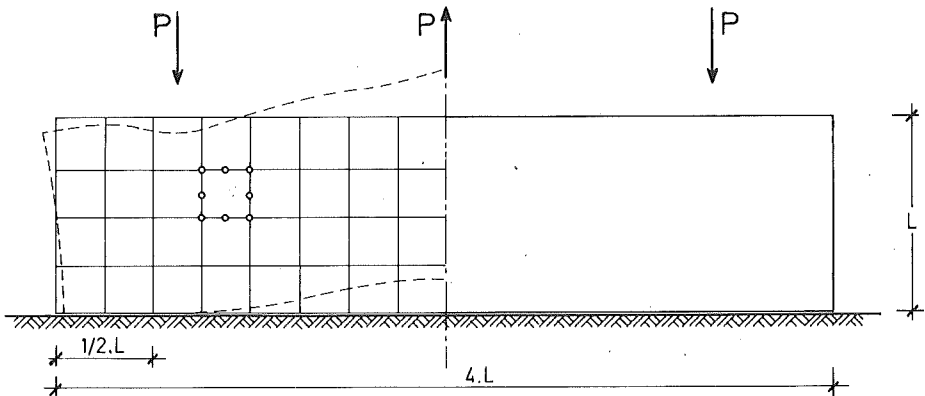


Fig. 2 - Laje enviesada

Na resolução dos GLCP necessários à análise das estruturas representadas nas figuras 1 e 2 obtiveram-se os seguintes resultados:

| N | STOR | #F | IT | NO | T |
|-----|-------|-----|----|-----|------|
| 242 | 11838 | 229 | 9 | 23. | 6.63 |
| | | 230 | 2 | 9.6 | 1.32 |
| | | 229 | 2 | 9.9 | 1.17 |
| 306 | 10106 | 293 | 7 | 42. | 12.0 |
| | | 293 | 4 | 23. | 4.55 |
| | | 294 | 4 | 14. | 2.57 |

onde

N = dimensão da matriz

STOR = número de posições de memória necessárias para números reais

#F = número de variáveis z básicas na solução

IT = número de iterações

NO = número de produtos/divisões*10⁴

T = tempo total em segundos CPU num computador CYBER 180-830.

Cada problema foi resolvido para três termos independentes diferentes correspondendo às diferentes combinações das solicitações que é necessário estudar. De salientar o facto de grande parte do trabalho efectuado na primeira resolução ser aproveitado para as seguintes. Como se pode ver pelo quadro, o número de elementos do conjunto F é pouco alterado em cada subproblema embora a composição do conjunto seja ligeiramente diferente. Daí ainda serem necessárias algumas iterações, mas esse número é inferior ao do GLCP inicial. Os resultados destes dois exemplos mostram a grande eficiência do algoritmo para o fim em vista.

6 - Referências

- [1] Júdice, J.J. e Pires, F.M. - *Bard-type Methods for the Linear Complementarity Problem with Symmetric Positive Definite Matrices*, IMA J.of Math.Appl. in Business & Industry 2 (1988/9) 51-68.
- [2] Mindlin, R.D. - *Influence of Rotary Inertia and Shear on Flexural Motions of Isotropic Elastic Plates*, J.Appl.Mech. (1951) 18-31.
- [3] Serra Neves, A.A. - *Análise de Placas e Cascas em Regime Elastoplástico Aplicando a Metodologia da Programação Matemática*, Tesé de Doutoramento, Universidade do Porto, 1988.

**Fotografia, Montagem
Impressão e Acabamentos**
Tip. Nocamil
COIMBRA

ÍNDICE

Jorge R.G.Freire de Sousa, José A.Vasconcelos Ferreira

Contributo para a Definição de Estratégias no Desenvolvimento de SADs 3

C.Henggeler Antunes, Maria J.Alves, Ana L.Silva, João N.Clímaco

Algumas Reflexões sobre uma Base de Métodos de Programação Linear
Multicritério 19

Maria Cândida Mourão, Maria João Ralha

Árvores de Suporte com Restrições de Grau 37

José Fernando Gonçalves, Manuel Pina Marques

Heurística para o Sequenciamento de Lotes Económicos de n Produtos
em m Máquinas Idênticas 47

Lino M.Tralhão, J.Craveirimha, J.P.Paixão

Modelização Estocástica de Tráfego em Redes Multiserviço 63

Fernanda M.Pires, A.A.Serra Neves

A Complementariedade Linear na Resolução de Estruturas
com Apoios Unilaterais 89



Associação Portuguesa para o Desenvolvimento
da Investigação Operacional

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