

INVESTIGAÇÃO OPERACIONAL

Dezembro 1992

Número 2

Volume 12

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 12 - nº 2 - Dezembro 1992

Publicação semestral

Editor Principal: Joaquim J. Júdice
Universidade de Coimbra

Comissão Editorial

M. Teresa Almeida
Inst. Sup. Economia e Gestão

Jaime Barceló
Univ. de Barcelona

Paulo Barcia
Univ. Nova de Lisboa

Isabel Branco
Univ. de Lisboa

António Câmara
Univ. Nova de Lisboa

C. Bana e Costa
Inst. Superior Técnico

M. Eugénia Captivo
Univ. de Lisboa

Jorge O. Cerdeira
Inst. Sup. de Agronomia

João Clímaco
Univ. de Coimbra

J. Dias Coelho
Univ. Nova de Lisboa

J. Rodrigues Dias
Univ. de Évora

Laureano Escudero
IBM, Espanha

J. Soeiro Ferreira
Univ. do Porto

J. Fernando Gonçalves
Univ. do Porto

Clóvis Gonzaga
Univ. Fed., Rio Janeiro

Luís Gouveia
Univ. de Lisboa

Rui C. Guimarães
Univ. do Porto

J. Assis Lopes
Inst. Superior Técnico

N. Maculan
Univ. Fed., Rio Janeiro

Ernesto Q. Martins
Univ. de Coimbra

Vladimiro Miranda
Univ. do Porto

J. Pinto Paixão
Univ. de Lisboa

M. Vaz Pato
Inst. Sup. Economia e Gestão

Celso Ribeiro
Univ. Católica, Rio Janeiro

A. Guimarães Rodrigues
Univ. do Minho

Mário S. Rosa
Univ. de Coimbra

J. Pinho de Sousa
Univ. do Porto

L. Valadares Tavares
Inst. Superior Técnico

Isabel H. Themido
Inst. Superior Técnico

B. Calafate Vasconcelos
Univ. do Porto

José M. Viegas
Inst. Superior Técnico

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: 5.000\$00

ESTUDO DE EFICIÊNCIA NA BOLSA DE VALORES DE LISBOA: CICLICIDADE E REGRAS DE COMPRA

João A.O. Soares

José A. Assis Lopes

Carlos M.F. Monteiro¹

Secção Autónoma de Economia e Gestão

Instituto Superior Técnico - Univ. Técnica de Lisboa

Abstract

This paper concerns the problem of testing financial markets' efficiency and, particularly, researches the existence of cycles in time series of prices. An algorithm of test is proposed, based on two main steps: 1) spectral analysis for the detection of the dominating frequencies; 2) simulation of trade rules based on the exploitation of 1). This methodology and the alternative Alexander filters are applied to the analysis of the Lisbon Stock Exchange Index "BTA", between 1988 and 1992. The results obtained demonstrate the inefficiency of this Stock Exchange within the above period.

Resumo

Este artigo debruça-se sobre o problema do teste à eficiência dos mercados financeiros, e, em particular, sobre a pesquisa de ciclicidade nas séries de preços. Nesse contexto propõe-se um algoritmo de teste baseado em dois estádios principais: 1) análise espectral para a detecção de frequências predominantes; 2) simulação de regras de compra baseadas na exploração destas. Esta metodologia, bem como a dos Filtros de Alexander, é explicada à análise do índice de cotações Totta e Açores, da Bolsa de Valores de Lisboa, entre 1988 e 1992. Os resultados obtidos permitem concluir pela ineficiência da referida Bolsa durante esse mesmo período.

Keywords

Stock Markets, Efficiency, Alexander Filters, Trade Rules, Forecasting

I. Introdução

A eficiência, ou melhor, a falta de eficiência, da Bolsa de Valores de Lisboa, é assunto frequentemente referido por quantos escrevem sobre o mercado português de capitais. No entanto, os argumentos aduzidos restringem-se, em geral, a uma apreciação qualitativa sobre as pequenas dimensão e liquidez bolsistas, bem como sobre a sua baixa rendibilidade. Uma justificação sustentada estatisticamente resume-se, tanto quanto é do nosso conhecimento, a dois artigos: Crato e Lopes (1989) e Sousa Câmara (1991). Ambos conhecem, contudo, limitações: o primeiro, porque abrange um pequeno período, de reconhecida ineficiência - 1.1.86 a 10.11.87 -; o segundo porque se baseia no estudo exclusivo das autocorrelações de ordem 1 e 30, confundindo indevidamente autocorrelações significativas com ausência de eficiência.

¹ Os autores agradecem as contribuições do Dr. Fábio Palmar, do BTA, e do Eng. Santa Marta Gomes da S.A.E.G. - IST.

Assim, neste trabalho procura-se:

- 1) Em primeiro lugar, rever o conceito de eficiência dos mercados de capitais, associando os diferentes graus de eficiência dos preços com as respectivas formas de os testar - isto é feito no ponto II;
- 2) centrados justificadamente no problema da eficiência fraca, apresentar uma metodologia de teste cuja principal inovação e interesse é a de reunir as vantagens das regras de filtragem e da análises à função de autocorrelação, de que falaremos - ponto III;
- 3) aplicar essa metodologia aos dados do mercado accionista da Bolsa de Valores de Lisboa (B.V.L.), confrontando os resultados com os de outras conhecidas regras (filtro de Alexander) e com o rendimento dos títulos do Estado - ponto IV;
- 4) concluir da eficiência do mercado - ponto V.

II. Definição de Eficiência. Graus de Eficiência dos Preços

Pode definir-se eficiência como a capacidade de afectar um conjunto limitado de recursos por forma a serem usados da maneira mais produtiva. No caso do mercado de capitais, isto pressupõe que os recursos financeiros de agentes com capacidade de poupança sejam canalizados prioritariamente para as empresas com oportunidade de investimento mais rentáveis. Como estas não são infinitas, o aumento da procura destes títulos faz aumentar a sua cotação, baixando conseqüentemente a sua taxa de rendimento. Os recursos financeiros excedentários são então canalizados para outros títulos que, obviamente, terão um preço menor, mas que poderão proporcionar idêntico rendimento (menores lucros sobre menores preços). No limite, podendo os preços flutuar livremente, e de idêntica forma circularem os recursos, espera-se que as taxas de rendimento marginais, ajustadas do seu grau de risco, sejam idênticas para todos os detentores de poupanças e investimentos produtivos.

Assim, para que um mercado de capitais seja eficiente como distribuidor óptimo de recursos, dois aspectos têm de ser cumpridos:

- Em primeiro lugar, que a oferta e a procura tenham uma dimensão considerável, e que haja condições de operacionalidade e regulamentação capazes de assegurar rapidez e concorrência na transacção dos títulos. Também, se é irrealista exigir custos de transacção nulos, espera-se que estes sejam muito reduzidos, compensando exclusivamente os encargos de processamento e risco suportados. Estas características constituem a chamada **eficiência operacional** ou interna.
- Em segundo lugar, que os preços reflectam total e imediatamente a informação que vai estando disponível e que seja relevante para a avaliação das empresas cotadas. De facto, sendo o valor dos títulos determinado pela expectativa dos seus rendimentos futuros, novas informações, que levam a uma alteração dessa expectativa, devem reflectir-se na procura desses títulos, originando a actualização

imediatamente dos seus preços. Esta característica designa-se por **eficiência externa** ou dos preços, e é sobre ela que recaem os diferentes testes.

Estes baseiam-se na ideia de que, se num mercado eficiente os preços reflectem imediatamente toda a informação, isso implica que não é possível concertadamente superar o rendimento médio do mercado. Desta forma, a aferição de eficiência expressa numa série de cotações pode fazer-se por comparação entre o rendimento de políticas de compra e venda baseadas num conjunto de informação, e o resultado de uma estratégia totalmente inactiva - comprar no início e vender no fim: o "buy-and-hold". Resta dizer que a literatura consagrou três graus de eficiência correspondentes a outros tantos conjuntos de informação:

- a) **EFICIÊNCIA FORTE** - Nesta definição nunca é possível superar o "buy-and-hold". Toda a informação, pública e privada, está reflectida nos preços. É a definição mais exigente, sendo usualmente testada por análise de eventuais casos de abuso de informação - "insider trading". Pretende-se desta forma verificar se os investidores bolsistas mais favorecidos no acesso à informação de uma dada empresa - funcionários superior e grandes accionistas -, obtêm lucros acima do normal, nas transações dos títulos dessa empresa. Nas bolsas norte-americanas é publicado um boletim onde essas transações são registadas (o "Official Summary of Insider Trading"), facto que não ocorre em Portugal. É pois impossível, aqui, proceder a essa análise.
- b) **EFICIÊNCIA SEMI-FORTE** - Significa que toda a informação pública (exclui-se, pois, a privada) é imediatamente reflectida nos preços. Advém da existência de ampla informação de carácter financeiro, bem como da capacidade de avaliação e resposta prontá por parte dos agentes económicos. Os estudos que se podem considerar como testes a este grau de eficiência, são-no por uma de duas vias: ou testam a velocidade de reacção das cotações à divulgação de uma notícia que as irá afectar - anúncio do resultado das empresas, alterações às taxas de juro, venda de blocos de acções, etc; ou se debruçam sobre variáveis que sendo pública e antecipadamente conhecidas, possam gozar de uma correlação forte com as cotações, permitindo a sua previsão - cotações de títulos cuja evolução preceda a dos outros, mercados "leaders", etc.
- c) **EFICIÊNCIA FRACA** - Significa que não se consegue superar o "buy-and-hold" com regras de compra baseadas na análise dos preços históricos (um subconjunto da informação pública). Sendo a definição em que o conjunto de informação é mais reduzido, óbvio se torna dizer que uma demonstração da ineficiência a este nível implica a ineficiência nos graus acima.

Há três tipos de teste estatísticos a este grau de eficiência: testes de repetições ("runs analysis"), filtros e estudo das autocorrelações. Os testes de repetições ou "runs" consistem na transformação de uma série de rendimentos noutra correspondente aos sinais daquela: 0, +, -.

Testa-se, então, se esta poderá provir de uma sequência gerada aleatoriamente. As vantagens deste tipo de testes advêm de, sendo não paramétricos, não pressuporem nenhuma hipótese quanto à distribuição estatística dos rendimentos, permitindo, para mais, ultrapassar o problema da heterocedasticidade. Contêm, no entanto, uma desvantagem essencial: constituem uma forma particular e pouco potente de teste à primeira autocorrelação, ignorando defasamentos superiores (ver Taylor [1989, pp.140-141]). Nesse sentido concentrar-nos-emos somente nos dois outros tipos de teste.

Os filtros são essencialmente regras mecânicas para determinação dos pontos de compra e venda. Os mais conhecidos são os baseados nas médias móveis (pontos determinados pelo cruzamento destas com as séries de dados) e o filtro de Alexander. Este pode descrever-se simplifadamente pela seguinte regras: comprar um título sempre que o seu preço sobe mais do que X% relativamente ao mínimo local entre o ponto em análise e a última venda; vendê-lo sempre que desce mais do que X%. A grande vantagem dos filtros, como forma de teste à eficiência, advém do facto de estes fornecerem directamente os rendimentos a comparar com o "buy-and-hold". A sua desvantagem decorre de, não sendo um método analítico, nunca se poder afirmar, face aos resultados negativos com certos filtros, que não existam outros que superem o "buy-and-hold".

Vantagens e desvantagens inversas tem o estudo das autocorrelações. Decorre de um procedimento analítico, baseado na ideia de que a eficiência fraca implica a inexistência de padrões nos dados, capazes de se revelarem através do correlograma dos rendimentos dos títulos. Contudo, a presença de autocorrelações significativamente não nulas não é garantia da existência de uma política activa de transacções que não incorra em custos superiores aos ganhos. Justifica-se, então, uma forma de teste mista, que descreveremos no ponto seguinte.

III. Metodologia para Teste da Eficiência dos Preços

A análise de uma série cronológica pode ser efectuada no domínio temporal ou frequencista. A primeira assenta em boa parte no estudo da função de autocorrelação. A segunda consiste na análise espectral, que permita descrever a variação de uma série através da adição de componentes cíclicas com diferentes frequências - correspondentes para o caso discreto ao conjunto $[1/n, \dots, (n/2)/n]$. É claro que há uma correspondência entre as duas análises, a qual resulta de o espectro ser a Transformada de Fourier da função de autocovariância (ver Soares [1989]). Nesse sentido, o estudo prévio do correlograma de uma série indicia o que se irá encontrar no gráfico do estimador do espectro de potência - o periodograma. Contudo, há vantagens na utilização deste último no estudo da ciclicidade, pois pelo periodograma visualiza-se a decomposição total da variância amostral na contribuição relativa das diversas frequências.

Dada esta justificação, estamos em condições de apresentar o algoritmo de teste:

PASSO 1 - Análise do periodograma da série de rendimentos para detecção de frequências predominantes, que são seleccionadas para o segundo PASSO.

PASSO 2 - Simulação de uma política cíclica de compra e venda baseada nos períodos seleccionados. Para cada período calculam-se os resultados correspondentes às diferentes fases possíveis, isto é, em que a primeira compra ocorre entre os pontos 1 e "p" da série ("p" = período dos ciclos). O melhor resultado, por período, é seleccionado para o terceiro PASSO. Podem ainda simular-se políticas não baseadas em períodos fixos. Para tal há que adicionar sinusóides correspondentes a dois ou mais períodos seleccionados, marcando os pontos de compra e venda de entre os máximos e mínimos.

PASSO 3 - Aos resultados, por período, obtidos anteriormente, há que deduzir os correspondentes custos de transacção. Os rendimentos líquidos são então comparados com a política de "buy-and-hold". Se forem superiores está provada a ineficiência do mercado.

IV. Análise de Eficiência na Bolsa de Valores de Lisboa: Jan.88 - Fev.92

A análise a que vamos proceder incide sobre a vertente accionista da Bolsa de Lisboa. Para esse efeito debruçar-nos-emos, não sobre alguns títulos seleccionados, mas sobre o indicador mais conhecido da evolução do mercado - o Índice do Banco Totta e Açores. Neste não consideraremos o período anterior a 1988, já que até 1986 a actividade bolsista foi bastante reduzida, e em 1987 perdurou, durante quase todo o ano, o impedimento de variações diárias superiores, em valor absoluto, a 5%. Esta regra, que ainda para mais coincidiu com um período de grande especulação bolsista, originou um alisamento artificial da série de valores.

A Figura 1 representa a evolução do índice entre 1.1.87 e 28.2.92. Embora a nossa análise não considere o ano de 1987, decidimos incluí-lo para melhor compreensão da evolução posterior. Assim, nos primeiros nove meses de 1987, o índice aumenta o valor anormal de 458%. Até ao fim do ano dá-se uma descida (reposição) brusca, sendo que entre Janeiro e Dezembro o índice aumentou 145%. Como consequência, em 1988 continua a descida, agora mais suave, registando-se uma variação anual negativa de 25%. O ano seguinte é um ano de retoma (aumento global de 40%), sendo 1990 e 1991 caracterizados, no essencial, por novas descidas. A importância do pico especulativo de 1987 pode ainda ser constatada quando se estabelecem comparações plurianuais. De facto, se considerarmos o período de Jan. 88 a Fev. 92 salienta-se a evolução negativa do mercado (menos 38,7%). Olhando desde Jan. de 87 a evolução é já positiva (56,7%), traduzindo contudo uma reduzida taxa de crescimento anual - cerca de 9,5%.

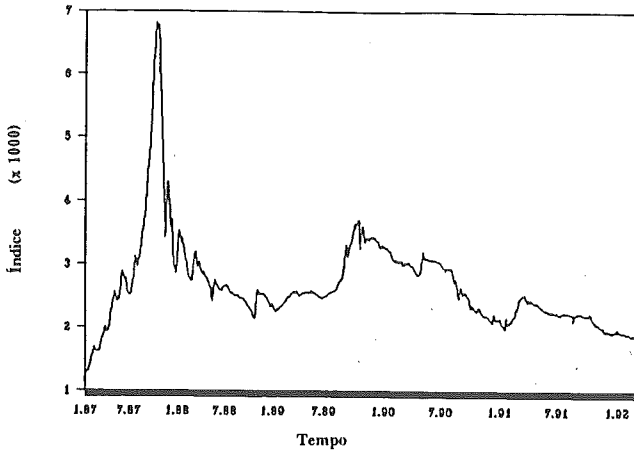


Figura 1 - Índice B.T.A. (Jan. 87 - Fev. 92)

Como é sabido, na análise geral de séries cronológicas, o estudo da função de autocorrelação ou a análise espectral não se fazem sobre uma série com marcadas tendências, mas antes sobre séries diferenciadas. Também aqui passaremos a analisar, equivalentemente, a série de taxas de rendimento diário simples (ver Figura 2). Os factos mais marcantes da mesma são: uma média próxima de zero, mas ligeiramente negativa (-0.04%); um número reduzido de valores mais extremados, com um mínimo de -7.5% e um máximo de 6.8%; uma maior variação do valor médio na parte inicial, com estabilização posterior.

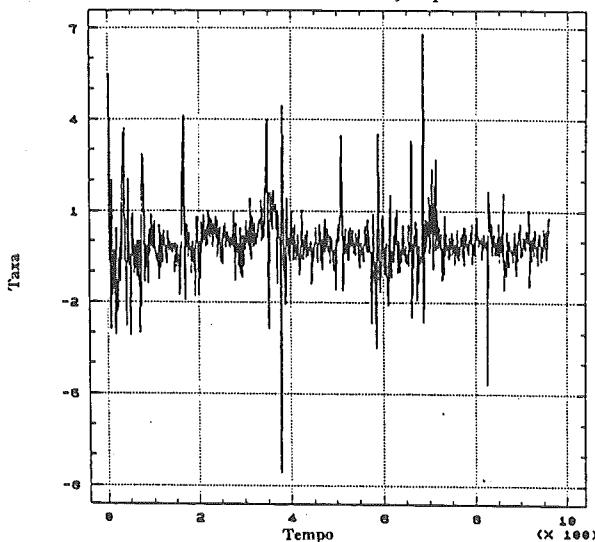


Figura 2 - Taxa de Rendimento Diário do Índice B.T.A. (Jan. 88 - Fev. 92)

Passemos agora à análise do correlograma (Figura 3). Nele destaca-se: o elevado valor da primeira autocorrelação (0.5); um valor ainda alto (0.21) para a segunda, e que é devido ao efeito de arrastamento produzido pela anterior, já que a segunda autocorrelação parcial é

negativa e não significativamente diferente de zero; valores significativos para o desfasamento 9 e 10; finalmente, de entre todos os outros valores não significativos, destacam-se os desfasamentos 25-26 e 32-33. As mesmas conclusões se retiram se se adoptarem valores corrigidos da presença de heteroscedasticidade, para os desvios padrão das autocorrelações (ver, p.ex., Lo e Mackinlay [1988]).

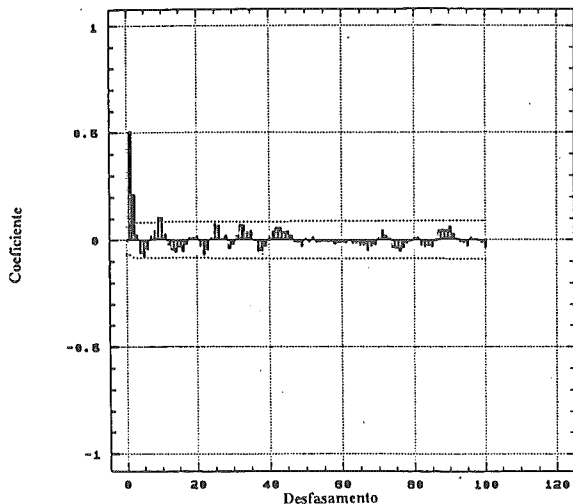


Figura 3 - Autocorrelações Amostrais da Taxa de Rendimento Diário do Índice B.T.A.

Na Figura 4 apresenta-se o periodograma da série. Nele se evidencia o predomínio das baixas frequências (entre 0 e 0.12 ciclos por intervalo amostral), com destaque claro para duas zonas que se tinham demarcado no correlograma: a correspondente a períodos entre 8 e 11 dias, e, com um valor de ordenada inferior, a de 33 dias.

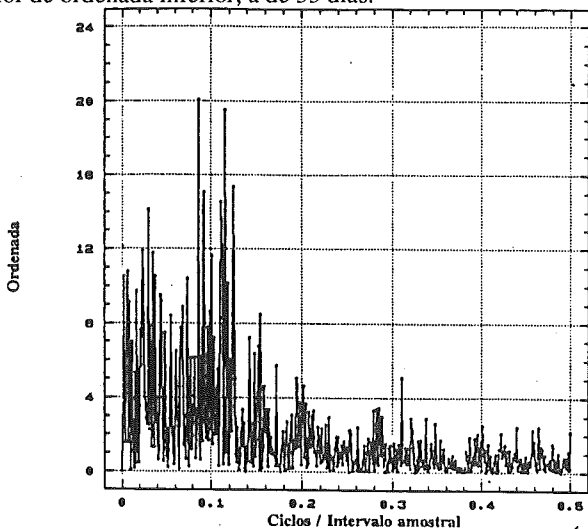


Figura 4 - Periodograma da Taxa de Rendimento Diária do Índice B.T.A.

No passo seguinte efectuaram-se simulações para aqueles períodos, com as diversas fases possíveis. Os valores mais elevados de rendimento foram:

- a) ciclo de 9 dias (compra-venda ao fim de 5 dias - compra passados mais 4 dias) com a primeira compra em 5.1.88;
- b) ciclo de 33 dias (compra-venda ao fim de 16 dias - compra passados mais 17 dias) com a primeira compra em 26.2.88.

Finalmente, no PASSO 3 procedeu-se à avaliação dos resultados. A mesma tem por base o Quadro 1, que passamos a comentar:

- a) As diferentes linhas correspondem às regras de transacção cíclicas de 9 a 33 dias, e, para além do "buy-and-hold", resolvemos incluir os rendimentos dos filtros de Alexander com 10, 5, 1.5 e 1%, bem como os rendimentos dos títulos de dívida pública. Os primeiros permitem avaliar o desempenho das regras cíclicas, com alternativas de período variável. Os segundos, calculados a partir de uma média ponderada dos diferentes títulos, compilada pelo Banco Totta e Açores, foram seleccionados como elemento comparativo, em virtude de não sofrerem o grau de risco associado ao mercado accionista.
- b) Em coluna apresentam-se: I - o número de dias com títulos em carteira (dias "In"); II - o número de dias "out"; III - o número de compras e vendas efectuadas, para cálculo dos custos de transacção; IV - estes mesmos custos tendo por base a taxa actual de corretagem mais elevada (0.5% para operações inferiores a 5000 contos) e a taxa de Bolsa (0.05%); V e VI - as taxas de rendimento brutas e líquidas dos custos de transacção; VII e VIII - as taxas médias diárias para os dias "In", incluídas para complemento da análise das colunas anteriores.
- c) Dos valores obtidos destacam-se os seguintes resultados:
 - c1) Por qualquer dos critérios (V e VII), os rendimentos brutos das regras de transacção cíclicas são superiores ao "buy-and-hold". Após deduzidos os custos de transacção, só a regra dos 33 dias a supera, embora com um rendimento negativo.
 - c2) Os filtros de Alexander com parâmetro inferior a 5% superam claramente as regras cíclicas e, conseqüentemente, o "buy-and-hold" (isto em termos brutos e líquidos). São ainda superiores ao desempenho dos títulos de dívida pública.

	I	II	III	IV	V	VI	VII	VIII
Regras de Filtragem	Dias "In"	Dias "Out"	NR Trans.	Custos de Trans.(%)	Rend. bruto no período	Rend. liq.no período	Rend.Bruto média diária	Rend.Liq. média diária
9 D	535	425	212	116.6%	50.0%	-66.6%	0.0804%	-0.1380%
33 D	453	507	58	31.9%	26.0%	-5.9%	0.0583%	-0.0120%
F.Alex.10%	448	512	16	8.8%	-18.0%	-26.8%	-0.0394%	-0.0590%
F.Alex.5%	363	597	24	13.2%	43.2%	30.0%	0.1047%	0.0680%
F.Alex.1.5%	322	638	62	34.1%	153.2%	119.1%	0.2988%	0.1930%
F.Alex.1%	357	603	76	41.8%	183.2%	141.5%	0.3000%	0.1830%
Buy-and-Hold	960	0	2	1.1%	-38.7%	-39.8%	-0.0415%	-0.0430%
Div.pública	960	0	2	1.1%	105.2%	104.1%	0.0749%	0.0743%

Quadro I - Análise Comparativa de Rendimentos

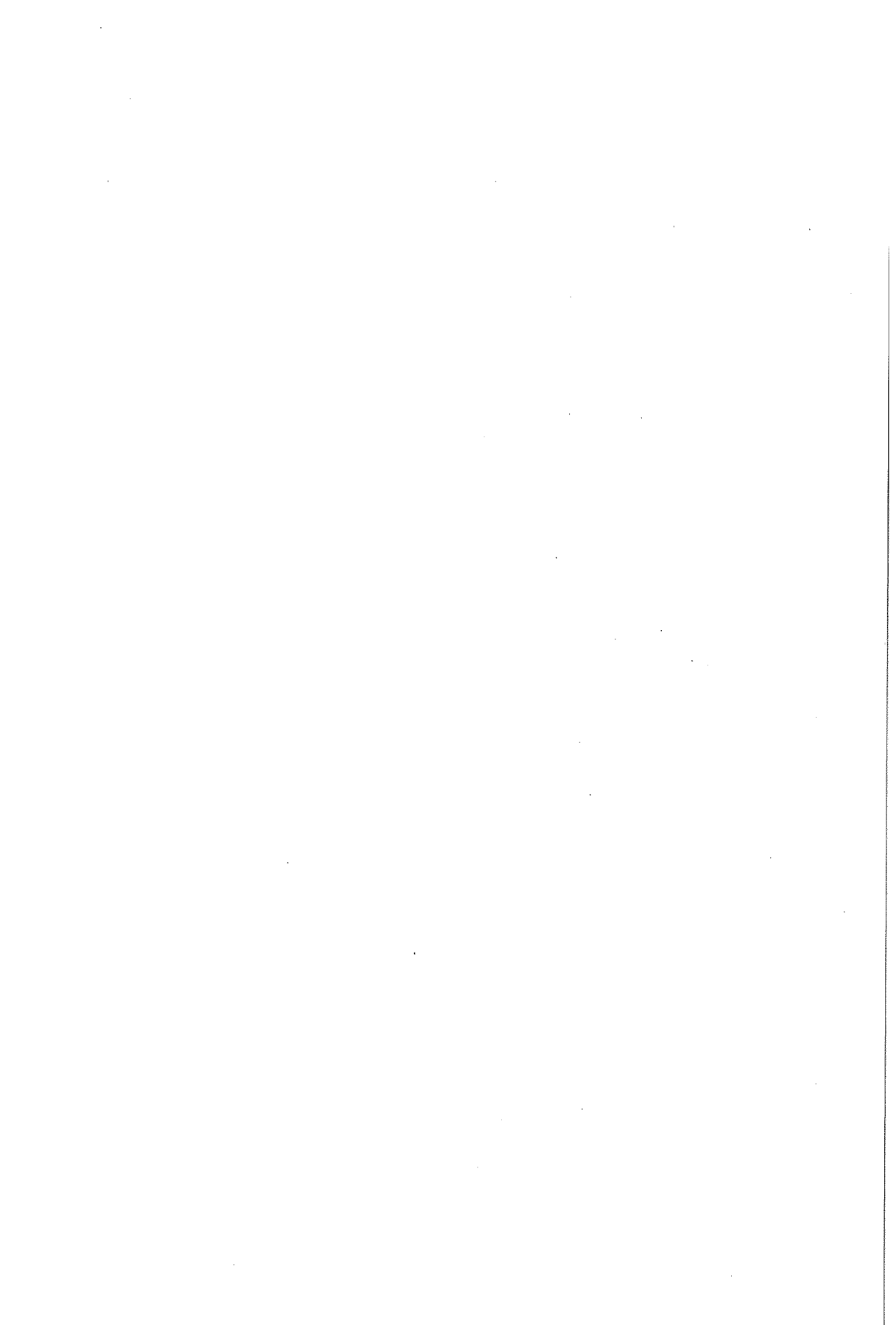
V. Conclusões

As conclusões deste trabalho são as seguintes:

- 1 - A Bolsa de Valores de Lisboa revela ineficiência para o período em análise (Jan.88 - Fev. 92).
- 2 - É possível demonstrá-lo através de uma metodologia que reúne o carácter analítico da análise espectral e a simulação das regras de compra.
- 3 - Apesar de pela análise do correlograma da série das taxas de rendimento, o desfaseamento de 33 dias não ser significativo a ± 2 DP's, a metodologia referida em 2) prova que uma política cíclica de compra e venda, com este período, ultrapassa o "buy-and-hold".
- 4 - A política referida em 3) revela, contudo, resultados bastante inferiores ao rendimento dos títulos de dívida pública e aos filtros de Alexander de 1.5 e 1%.
- 5 - Os filtros de Alexander de 1 e 1.5% mostram ter sido possível, durante o período em análise, uma rendibilidade de cerca de 23% ao ano, isto é, sensivelmente 5 pontos percentuais acima do rendimento dos títulos de dívida pública.
- 6 - É de explorar o prosseguimento da metodologia exposta neste trabalho, construindo regras de transacção baseadas na adição de sinusóides com período correspondente aos picos predominantes do periodograma.

VI. Referências

- [1] Crato, N. e Assis Lopes, J., Forecasting Price Trends at the Lisbon Stock Exchange, in "A Reappraisal of the Efficiency of Financial Markets", edited by Guimarães, Kingsman and Taylor, NATO Asi Series, Vol. F54 (1989), Springer-Verlag, Berlin Heidelberg.
- [2] Lo, A.W. e Mackinlay, A.C., Stock Market Prices Do Not Follow Random Walks> Evidence from a Simple Specification Test, The Review of Financial Studies, 1 (1988) 41-66.
- [3] Soares, J.A.O., Métodos Estatísticos de Análise da Ciclicidade em Séries Económicas, Tese de Mestrado, IST-UTL, 1989.
- [4] Sousa da Câmara, A., A Ineficiência do Mercado - Carteira de Acções na Prática, O Investidor Económico 21 (1991) 4-10.
- [5] Taylor, S., Modelling Financial Time Series, John Wiley & Sons, New York, 1989.



ANÁLISE DE UMA REGRA DE COMPRA DE MERCADORIAS SUJEITAS A FLUTUAÇÕES DE PREÇO

José António Sarsfield Cabral

Rui Campos Guimarães

GEIN - Secção de Gestão e Engenharia Industrial

DEMEGI - Departamento de Engenharia Mecânica e Gestão Industrial

Faculdade de Engenharia da Universidade do Porto

Rua dos Bragas, 4099 Porto

Abstract

Price fluctuation is a major source of complexity for commodity purchasing decisions. In a efficient market it is appropriate to adopt passive management policies, involving an attempt either to transfer the risks associated with price movements to third parties or to buy consistently at or close to the average market price.

This paper describes an alternative type of policy concerning the problem of the day to day buying commodities through the futures markets and deals specifically with a heuristic rule developed by the authors for the 'purchasing under a deadline' scenario. The rule, which aims at reducing the long-run purchasing costs while keeping risks within acceptable limits, is based on short-term forecasts produced by Taylor's price-trend model.

The policy was tested using the 1972-87 and 1982-87 series of CBOT daily corn and soybean futures closing prices. The results suggest that it leads to a significant improvement on the purchasing results even for buyers with a considerable degree of risk aversion.

Resumo

A flutuação dos preços das mercadorias constitui um factor de complexidade nas decisões de compra de matérias primas. Se um mercado for considerado eficiente, fará sentido adoptar políticas de compra passivas, através das quais se procuram transferir para terceiros os riscos associados ao movimento dos preços ou ainda adquirir as mercadorias a um custo tanto quanto possível próximo do preço médio do mercado.

Nesta comunicação descreve-se uma política alternativa que envolve a compra de mercadorias através dos mercados de futuros e com a qual se pretende reduzir, no longo prazo, o custo de aquisição de matérias primas, mantendo os riscos associados à flutuação dos preços dentro de limites aceitáveis. A política baseia-se em previsões dos preços efectuadas com base nos modelos com tendência de Taylor e numa regra heurística para a compra de futuros, desenvolvida pelos autores.

Os resultados do teste desta política na simulação da aquisição de futuros de milho e de soja, na bolsa de Chicago, ao longo dos anos 1977-87 e 1982-87, respectivamente, constituem claros indicadores da possibilidade de se obterem poupanças significativas em relação ao preço médio do mercado, mesmo por compradores com razoável aversão ao risco.

Keywords

commodities, futures markets, time series analysis, forecasting, purchasing rules.

1. Introdução

Em diversas indústrias transformadoras, as matérias-primas correspondem a uma parcela muito significativa do custo dos produtos finais. Sucede que os preços de tais matérias primas são frequentemente determinados em bolsas de mercadorias, alterando-se de dia para dia e mesmo de minuto para minuto. Em alguns casos não é possível repercutir, nem sequer parcialmente, a flutuação daqueles preços no preço de venda dos produtos acabados.

Nestas circunstâncias, o problema do aprovisionamento de matérias primas complica-se substancialmente. De facto, para além das outras tarefas inerentes ao processo de aquisição, o responsável pelas compras deverá decidir sobre a quantidade de mercadoria a adquirir em cada

oportunidade de compra. O preço que pagará pelo produto poderá ser significativamente afectado pelas decisões adoptadas e, mesmo quando é possível transferir para o cliente o risco associado à variação dos preços, os ganhos que possam resultar de boas decisões de compra podem justificar a adopção de políticas que visem a redução do custo global de aquisição das matérias primas.

O problema analisado nesta comunicação diz respeito ao que, na terminologia de Kingsman [1985], corresponde ao último estágio do processo de decisão associado à compra de mercadorias sujeitas a flutuações de preço. Tal estágio consiste no desenvolvimento de regras de compra que, a partir de previsões de preços a curto prazo, permitam definir as quantidades a adquirir nas sucessivas oportunidades de compra.

Estas regras (ou políticas) são definidas tendo em conta o cenário de compra, isto é, o conjunto de parâmetros e restrições que condicionam o processo de aquisição da mercadoria em causa. O cenário considerado corresponde ao que vulgarmente se designa por "compra com data limite", que é o típico dos processos de compra que envolvem sistemas de transporte descontínuos, como é o caso das mercadorias agrícolas (Cabral e Guimarães [1988] e Guimarães e Kingsman [1990]). Nos termos de Kingsman [1985] esse cenário ("Type 2 Tactical Buying Situation") define-se da forma seguinte: "Aquisição de uma determinada quantidade de uma mercadoria antes de uma data limite, existindo várias oportunidades de compra. Em cada oportunidade é cotado apenas um preço perante o qual o comprador pode adquirir toda ou parte da quantidade global requerida ou ainda esperar pela oportunidade de compra seguinte".

No problema em causa, admite-se ainda que a compra diz respeito não à mercadoria "física" necessária para o processo produtivo, mas a uma determinada quantidade de futuros dessa matéria-prima. Considera-se que a compra dos futuros faz parte do procedimento habitual de aquisição de mercadorias sujeitas a flutuações de preço (ver Guimarães [1981]). Em cada oportunidade do "período activo de compra" (período alocado para a compra da mercadoria, tipicamente de 1 a 2 meses), o comprador poderá adquirir os futuros que pretender, desde que na data limite disponha do número de futuros equivalente à quantidade de matéria-prima "física" requerida inicialmente. Não se admite a possibilidade da venda de futuros e presume-se que as transacções são efectuadas ao preço das cotações diárias de fecho dos futuros envolvidos.

2. A Eficiência dos Mercados no Contexto do Problema de Compra

Vários autores têm proposto diferentes definições de eficiência. Em todas elas se associa às bolsas de mercadorias (e de valores) um mecanismo capaz de processar a informação (passada e presente) de uma forma tão rápida e eficaz que os preços, a cada momento, traduziriam o "justo" valor do que lá fosse transaccionado. Por outras palavras, cada preço cotado num tal mercado reflectiria o ponto de equilíbrio entre a oferta e a procura e incorporaria toda a

informação relevante disponível naquele momento. Por exemplo, segundo Fama [1970], a "eficiência perfeita" (na forma fraca, semi-forte ou forte, consoante se considere apenas a informação contida nos preços do passado e do presente, a informação pública ou toda a informação pública e privada) implicaria uma reacção instantânea e total aos choques provocados pela informação, o que, por sua vez, determinaria que, em relação aos preços, os mercados não possuiriam "memória".

Para alguns autores (ver Guimarães et al [1989]) a impossibilidade de, consistentemente, se obterem lucros num mercado deveria ser considerada uma condição necessária para a sua eficiência. Por consequência, se a um qualquer mercado for aplicada uma política de actuação que prove ser capaz de gerar lucros (após se descontarem o factor de risco e os custos associados às transacções), esse mercado não deverá ser considerado eficiente. A questão da eficiência é colocada assim na aceitação ou rejeição da prova que for feita sobre a significância dos resultados obtidos com a política adoptada.

No contexto dos processos de compra, parece razoável aceitar-se que a eficiência de um mercado depende não só da forma como a informação se reflecte nos preços, mas também das restrições que são impostas pelo processo particular de compra de uma determinada mercadoria e ainda do critério através do qual o comprador avalia as várias políticas alternativas. Deste modo, para um determinado processo de compra de mercadorias, um mercado será considerado eficiente (na sua forma fraca) se os resultados obtidos pela utilização dos preços do presente e do passado não forem, do ponto de vista do comprador, melhores do que aqueles que seriam obtidos na ausência de qualquer informação. Esta definição de eficiência (ver Cabral e Guimarães [1989]), que toma por base a apresentada em Jensen [1978], será a adoptada a partir daqui. Nesta definição, a eficiência depende de cada comprador em particular e do respectivo processo de compra. É neste aspecto que ela se distingue da de Jensen.

Verificando-se a hipótese fraca de eficiência, a melhor informação que um negociante pode extrair dos preços sobre aquilo que se passará no futuro está contida no preço prevalecente no momento. Nestas condições, resultarão infrutíferas quaisquer tentativas de desenvolver e aplicar com proveito políticas de compra que, a partir de previsões de preços (tendo por base as cotações do passado e do presente), tenham por objectivo reduzir os custos de aquisição de uma mercadoria. Verificando-se a "eficiência perfeita", fosse qual fosse a informação utilizada por qualquer política de compra, tal política estaria condenada a não ter sucesso no longo prazo. Numa situação deste tipo, o comprador optará por uma política de compra "passiva", cujo fito é o de transferir para terceiros os riscos e os custos associados aos movimentos dos preços ou o de comprar consistentemente ao preço médio do mercado. Em contraste com esta situação, políticas activas de compra, tais como a que se apresenta neste trabalho, só farão sentido se a eficiência do mercado for posta em causa, implicando que se verifiquem as seguintes condições:

- (i) existência de tendências persistentes, detectáveis através da observação dos preços históricos;
- (ii) capacidade de desenvolver ou aplicar uma regra de compra a partir da qual se possam obter reduções, estatisticamente significativas, no custo global médio de compra; e
- (iii) possibilidade de que as poupanças sejam geradas dentro dos limites de risco toleráveis pelo comprador.

Através dos modelos com tendência propostos por Taylor, tem sido possível não só demonstrar a existência de tendências significativas em diversas séries de preços mas também fundamentar regras de compra capazes de gerar poupanças significativas (ver, por exemplo, Taylor [1980, 1986, 1988], Taylor e Tari [1989], Guimarães e Kingsman [1990]). Ficam assim preenchidas as duas primeiras condições. Resta, porém, demonstrar a terceira.

3. Regra de Compra com Controlo do Risco

A política de compra com controlo do risco baseia-se na política proposta por Guimarães [1981] e recorre, igualmente, a previsões de preços para o curto prazo obtidas a partir dos modelos de Taylor. Na sua origem está a ideia de que, a verificarem-se tendências no curto prazo (tal se admite nos modelos de Taylor), as compras só se devem efectuar quando se prevêem tendências positivas. A quantidade a comprar deverá ser proporcional à confiança que se deposita na subida dos preços, confiança essa medida pela "tendência normalizada" (ver Taylor [1978]). Na base da reformulação da regra de Guimarães esteve o reconhecimento de que nela não se tomava em consideração a atitude do comprador face ao risco.

A política proposta será apresentada de acordo com a seguinte notação:

- N: número total de oportunidades incluídas no período activo de compra (dias)
- Q: quantidade total de mercadoria (futuros) que deverá ser adquirida durante o período de compra
- n: índice denotando cada um dos dias do período activo; em cada dia, o valor de n corresponde ao número de oportunidades que restam até à data limite ($n = N, N-1, \dots, 1$)
- A_n : quantidade que, de acordo com a política, deverá ser adquirida no dia n
- q_n : quantidade por adquirir quando faltam n oportunidades para a data limite (note-se que $q_n = Q$ e que $q_n = q_{n+1} - A_{n+1}$)
- $\hat{\mu}_{n-1}$: estimativa da tendência para o dia n-1, calculada depois de se conhecer a cotação do preço no dia n
- $\hat{\sigma}_{n-1}$: estimativa do desvio do padrão de $\hat{\mu}_{n-1}$, calculada depois de se conhecer a cotação do preço no dia n
- c_n : estimativa da "tendência normalizada" obtida no dia n para o dia seguinte ($c_n = \hat{\mu}_{n-1}/\hat{\sigma}_{n-1}$)
- T: parâmetro controlador de compras

A quantidade a adquirir em cada oportunidade é especificada de acordo com as seguintes expressões:

- Para $T > 0$

$$\text{para } n \geq 2: A_n = \begin{cases} 0, & \text{se } -T > c_n \\ q_n/n, & \text{se } -T \leq c_n \leq T \\ \text{Min}[q_n, \text{Max}(A_n^*, q_n/n)] & \text{se } T < c_n \end{cases}$$

para $n = 1: A_1 = q_1$

onde

$$A_n^* = f(n) c_n q_n.$$

- Para $T = 0$

Adopta-se a regra original de Guimarães, para a qual se efectuam compras na quantidade A_n^* sempre que c_n é positiva e se adiam as compras quando $c_n \leq 0$.

A regra heurística que está na base da política de compra toma em consideração as seguintes situações:

- (i) c_n possui um elevado valor positivo
O sinal (positivo) da tendência é previsto com grande confiança e, consequentemente, uma elevada proporção de q_n (quantidade que falta comprar) deverá ser adquirida. Quanto maior for c_n tanto maior será aquela proporção.
- (ii) c_n possui um elevado valor negativo
O sinal (negativo) da tendência é previsto com elevada confiança e, consequentemente, não se deverão fazer compras.
- (iii) c_n possui um pequeno valor absoluto
Nesta situação, em que o sinal da tendência não se pode prever com confiança, adoptar-se-á a política alternativa face à qual a regra é avaliada ("benchmark"). No contexto do problema em análise, a política "benchmark" consiste na compra de igual quantidade de futuros em cada oportunidade, conduzindo a um custo de aquisição igual ao preço médio do mercado no período activo de compra. Isso significa que no dia n se adquirirá a quantidade q_n/n .

A distinção entre quaisquer duas destas situações pode ser feita à custa do parâmetro controlador de compra T (ver Cabral e Guimarães [1989]). Se for escolhido um valor elevado para T (por exemplo, 3), a regra conduzirá praticamente à adopção da "benchmark" (isto é, a compra da mesma quantidade todos os dias), para a qual a poupança em relação ao preço médio do mercado será nula. Esta situação ocorre porque c_n é aproximadamente $N(0,1)$, sendo portanto altamente improvável que tome um valor absoluto superior a 3. À medida que se reduz T , o valor esperado da poupança deverá aumentar até a um valor máximo em $T = 0$, mas, simultaneamente, o risco de, num período de compra, se obter uma poupança negativa subirá também, até a um limite superior à volta de 50% (quanto $T = 0$).

A função $f(n)$ reflecte a influência exercida pelo número de oportunidades que restam para a data limite sobre a quantidade a adquirir. Considerem-se dois dias diferentes no período activo, por exemplo $n = 25$ e $n = 3$, nos quais a confiança na subida dos preços é idêntica ($c_{25} = c_3 > 0$) e em que a quantidade que falta adquirir, q_n , é a mesma ($q_{25} = q_3 > 0$). Nesta situação, quanto maior for o número de oportunidades disponíveis até ao fim do período activo de compra tanto maior será a probabilidade de que as tendências se modifiquem ou mesmo se invertam e que, portanto, se venham a verificar preços inferiores ao preço corrente. Assim, nas

condições acima descritas, a quantidade a comprar quando restam apenas 3 dias deverá ser maior do que quando ainda restam 25.

A função $f(n)$ tem que ser calculada empiricamente, à custa de dados do passado ou recorrendo-se a preços simulados, maximizando uma medida de eficiência. As suas únicas características conhecidas a priori são as de que nunca pode ser negativa e deverá crescer monotonamente à medida que n diminui (Guimarães e Kingsman [1990], Cabral e Guimarães [1989]).

Guimarães [1981] propôs a adopção da "poupança relativa normalizada", PRN, como medida de eficiência das regras de compra. Tal medida define-se da seguinte forma:

$$PRN = \frac{(PMM - PC)/PMM}{MVR}$$

onde

- PMM : preço médio dos futuros ao longo do período activo de compra (seria o preço de compra se fosse adoptada a política "benchmark")
- PC : preço médio unitário pago pelos futuros, de acordo com a política adoptada
- MVR : medida da volatilidade relativa, expressa através da média das estimativas do desvio absoluto médio dos retornos diários obtidas durante o período activo (um retorno diário, $x_t = \log(z_t) - \log(z_{t-1})$, reflecte aproximadamente a variação relativa entre preços consecutivos).

Quando o custo de posse dos futuros não é considerado, o numerador de PRN representa a poupança relativa obtida com a política proposta em relação à "benchmark". A diferença entre os custos de posse dos futuros para as duas políticas pode ser desprezada. Por exemplo, considerando taxas de juro correntes e assumindo que no acto de compra de futuros o depósito exigido pelo "broker" é de 10% do valor dos futuros, aquela diferença de custos em períodos activos de 1 mês não ultrapassará 0.03% do preço médio dos futuros ao longo do período de compra (Cabral e Guimarães [1988]).

O denominador de PRN representa um factor de escala e foi introduzido com o objectivo de estabilizar a distribuição desta medida de eficiência ao longo do tempo. A invariabilidade de PRN no tempo foi testada (Cabral e Guimarães [1989], Cabral [1990]) e, para efeitos práticos, confirmada.

As estimativas das tendências e dos seus desvios padrão utilizadas no cálculo de c_n (recorde-se que $c_n = \hat{\mu}_{n-1}/\hat{\sigma}_{n-1}$), são obtidas recorrendo às expressões definidas em Taylor [1980, 1986]. Por sua vez, as estimativas de μ_n e de v_n dependem dos valores encontrados para os parâmetros da função autocorrelação dos retornos diários, A e p ($\rho_i = Ap^i$, para $A > 0$ e para $0 < p < 1$). O parâmetro A reflecte a magnitude das tendências (expressa em proporção da magnitude dos retornos) e $m = (1-p)^{-1}$ traduz a duração média (ou estabilidade) das tendências. Num mercado não-eficiente aqueles valores devem ser significativamente diferentes de zero. A estatística U^* (Taylor [1980]) permite testar $\rho_i = 0$ em alternativa a $\rho_i = Ap^i$.

O valor do parâmetro controlador do risco, T , é determinado dinamicamente de acordo com o procedimento formal proposto pelos autores. Tal procedimento, descrito em Cabral e

Guimarães [1988], é baseado no conceito de utilidade de von Newman e Morgenstern [1967] e assegura que, no longo prazo, o valor esperado da utilidade do comprador é maximizada. Se a função utilidade for linear com PRN (isto é, se o comprador for neutro face ao risco), então o valor óptimo de T é zero e, na prática, a política toma a forma originalmente proposta por Guimarães [1981].

3. Compra com Data Limite de Futuros de Milho através da Regra com Controle do Risco

As cotações de fecho de todos os contratos de futuros de milho entre o primeiro dia útil de 1977 e o último de 1987 na CBOT (Chicago Board of Trade) constituíram o conjunto de preços utilizado no teste da regra de compra. Este conjunto foi dividido em séries consecutivas de 21 e de 42 dias, a partir do primeiro dia útil de 1977. Estas séries representam os períodos activos de compra utilizados na simulação.

No teste efectuado, admitiu-se que o contrato a negociar em cada período activo de 21 ou de 42 dias dependia da data limite (isto é, do último dia do período), de acordo com o procedimento definido em Guimarães [1981].

Os valores adoptados para os parâmetros A e m correspondem às estimativas obtidas por Guimarães [1981] com os preços do período 1963-76 ($\hat{A} = 1.49\%$, $\hat{m} = 26$ dias). A função $f(n)$ foi substituída pela aproximação curvilinear proposta por Cabral e Guimarães [1989], $f(n) = \tau(n-1)^{-\omega}$, calibrada recorrendo a preços gerados artificialmente. Admitiu-se que o comprador exprime a sua aversão ao risco (em função da poupança relativa) de acordo com a curva apresentada no diagrama (i) da Figura 1. A escolha do parâmetro T foi efectuada em função das estimativas diárias dos desvios padrão dos retornos de acordo com a curva apresentada na Figura 1 (ii). Nestas condições, a simulação foi efectuada apenas com base em informação anterior ao período 1977-87.

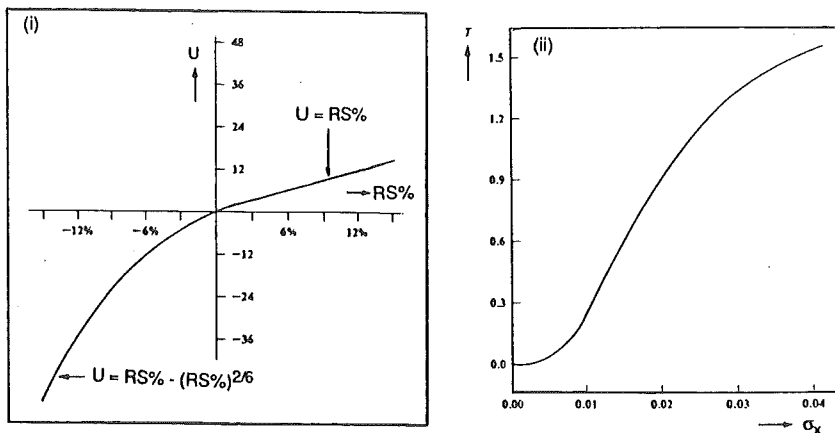


Figura 1: Exemplo de função utilidade e dos correspondentes valores óptimos do parâmetro T (Fonte: Cabral e Guimarães [1988])

Uma vez adoptado este procedimento, a regra é avaliada em termos de valor médio da utilidade - valores positivos de \bar{U} correspondem a desempenhos positivos e valores negativos de \bar{U} correspondem a desempenhos negativos.

Na Tabela 1 apresentam-se os resultados referentes a períodos de 21 e 42 dias, obtidos adaptando dinamicamente T e fixando-o no valor $T = 0$ (política original de Guimarães). Em relação à regra de compra com controlo de risco, destaca-se em primeiro lugar que a utilidade média apresenta valores positivos tanto para períodos de 21 como de 42 dias. No contexto do problema que vem sendo tratado, este resultado assume uma enorme importância: ele significa que a regra seria atractiva mesmo para um comprador razoavelmente avesso ao risco e, portanto, seria preferida à "benchmark". Note-se que os valores médios de \overline{PR} e de \overline{PRN} são mais baixos dos que se obteriam se se fizesse $T = 0$. Esta redução corresponde ao preço a pagar para se conseguir limitar a frequência de situações em que a poupança é negativa.

DIMENSÃO DO PERÍODO DE COMPRA (Dias)	DIMENSÃO DA AMOSTRA	\bar{U} T adapt.	\overline{PR} (%)		\overline{PRN}	
			T adapt.	T = 0	T adapt.	T = 0
21	132	0.37	0.50%	0.52%	0.49	0.54
42	65	0.93	1.09%	1.15%	1.00	1.10

Parâmetros da regra de compra: $\hat{A} = 1.49\%$, $m = 26$; $\tau = 14$, $\omega = 1.4$

Tabela 1: Resultados da aplicação da regra com controlo do risco e da regra original de Guimarães ($T = 0$) em períodos de 21 e 42 dias, entre 1977 e 1987

Cabral e Guimarães [1988] demonstraram que as poupanças relativas médias obtidas com $T = 0$ são estatisticamente significativas. Em Cabral [1990] mostra-se ainda que \overline{PRN} é pouco sensível aos erros cometidos nas estimativas dos parâmetros da regra de compra.

Na Figura 2 ilustra-se graficamente o funcionamento da regra proposta, comparando-a com a regra original de Guimarães ($T = 0$) e com a política "benchmark", recorrendo-se a compras simuladas num período de 21 dias com data limite em 87/12/04. Nos primeiros sete dias do período de compra os valores de c_n estão compreendidos entre $-T$ e T . Nesta situação, a regra estabelece que se adopte a "benchmark" (ou seja comprar q_n/n em cada oportunidade). A partir de $n = 14$, as quantidades a adquirir em cada oportunidade são proporcionais a $f(n) \cdot c_n \cdot q_n$, até que toda a quantidade inicialmente prevista esteja comprada.

No caso da regra de Guimarães ($T = 0$), as aquisições começam apenas quando c_n é positiva. Assim, só a partir de $n = 18$ se desencadeiam as compras, inicialmente a um ritmo extremamente baixo, compatível com valores de c_{18} , c_{19} e c_{20} muito próximos de zero. Aplicando a regra com controlo do risco, ao fim do sexto dia, $n = 16$, estavam já adquiridos

cerca de 29% da quantidade total, enquanto que de acordo com a regra original de Guimarães apenas se tinham comprado, até esse momento, cerca de 4%.

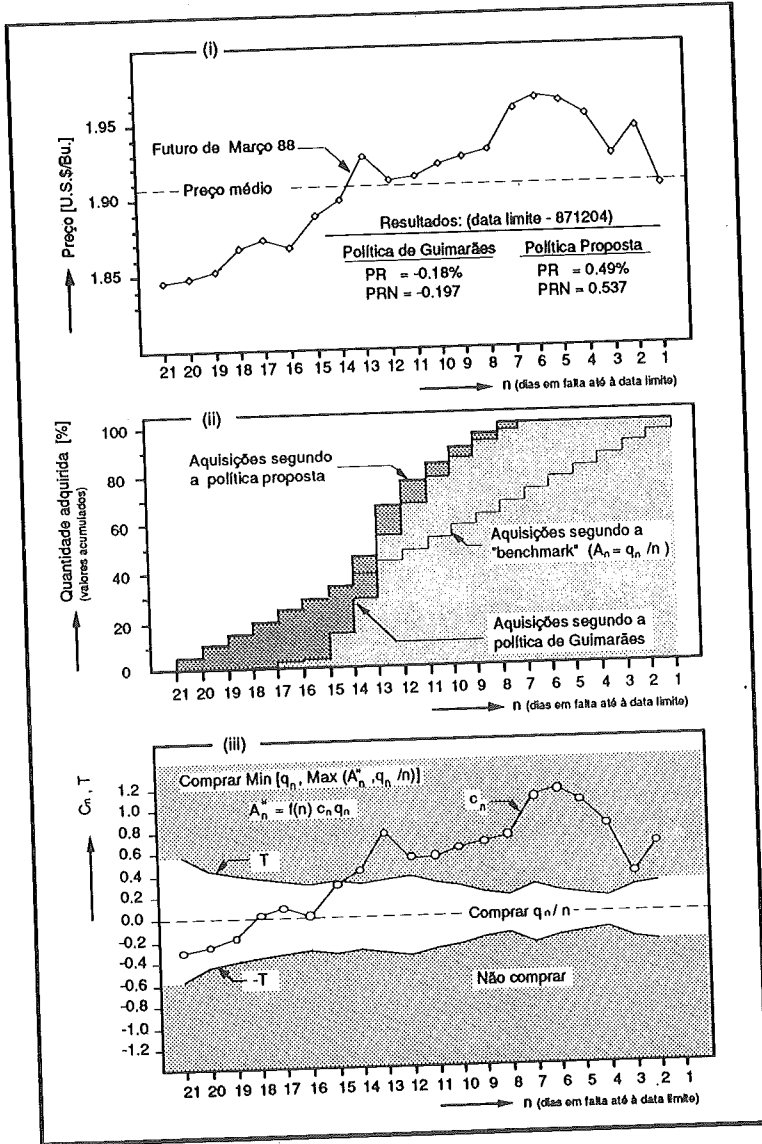


Figura 2: Compras em período de 21 dias com data limite em 87/12/04 (Fonte Cabral [1990])
 Com a análise descrita neste ponto completa-se uma sequência de estudos relativos à compra de futuros de milho em Chicago. No âmbito do cenário seleccionado (compra com data limite), os resultados obtidos, tanto com preços simulados (Cabral e Guimarães [1989]) como

com 25 anos de preços diários reais (Guimarães e Kingman [1990] e Cabral e Guimarães [1989]), permitem rejeitar claramente a tese de que não é possível (no caso dos futuros de milho da Chicago Board of Trade) bater o mercado de uma forma consistente.

A extraordinária estabilidade que se verificou nos parâmetros de autocorrelação entre 1963 e 1987 permitiu que a regra de compra fosse aplicada sem se proceder à sua recalibração. Num mercado em que os parâmetros A e m se alterem significativamente ao longo do tempo será necessário encontrar procedimentos para, por um lado, detectar essas alterações e, por outro, recalibrar os parâmetros da regra de compra.

4. Procedimento de Recalibração dos Parâmetros da Regra de Compra

O mercado dos futuros de soja em Chicago constitui um exemplo no qual os parâmetros A e m não são estáveis. Por essa razão, será utilizada uma série de cotações diárias destes futuros para ilustrar procedimentos de calibração da regra de compra a adoptar numa tal situação.

Neste estudo foram utilizadas as séries construídas a partir das cotações diárias de fecho de todos os contratos de soja negociáveis na Bolsa de Chicago, entre o primeiro dia útil de 1972 e o último de 1987. Seguindo o procedimento adoptado para o milho, dividiu-se o período 1972-87 em dois subperíodos: 1972-81 (calibração) e 1982-87 (teste).

Como o objectivo de calibrar a regra de compra, o subperíodo 1972-81 foi seccionado em séries consecutivas de 21 dias, o que conduziu a uma amostra de 117 períodos activos de compra. De forma idêntica ao procedimento adoptado para a compra de futuros de milho, fez-se depender o contrato a negociar em cada período activo da respectiva data limite, de acordo com a regra apresentada em Cabral [1990].

Os valores obtidos para os parâmetros de autocorrelação A e m em 1972-81 foram, respectivamente, de 2.61% e de 26 dias. A estatística U^* atingiu o valor altamente significativo de 10.91. Adoptada a aproximação curvilínea de $f(n)$, a sua calibração conduziu a $\omega = 1.3$ e a $\tau = 26$. Tal como se aguardava (atente-se na magnitude das estimativas de A e de m), os resultados da aplicação da regra no período de calibração foram muito animadores. Fazendo, por exemplo, $T = 0$, a poupança relativa média em períodos de 21 dias atingiu 1.35% com $\overline{PRN} = 0.76$.

Uma vez calibrada, a regra foi simulada no período 1982-87. Este teste conduziu a resultados extremamente pobres. A eficiência da regra ficou muito aquém do esperado, registando-se valores de \overline{PRN} quase vinte vezes inferiores aos encontrados no período de calibração e não significativamente diferentes de zero. A discrepância entre os resultados obtidos nos períodos de calibração e teste tem origem na alteração dos parâmetros de autocorrelação do processo estocástico dos retornos diários (ver Cabral [1990]). De facto, Cabral e Guimarães [1988, 1989] demonstraram que quando tais parâmetros são estáveis também o é a distribuição de PRN .

A variação do padrão de comportamento no mercado coloca dificuldades adicionais na adopção de regras activas de compra que se baseiam em métodos estatísticos de previsão para o curto prazo. Essencialmente, aquelas dificuldades são de dois tipos:

- (i) detecção da magnitude da modificação do comportamento do mercado e do momento em que tal modificação ocorre; e
- (ii) neste contexto, estimação dinâmica dos parâmetros a utilizar na regra de compra.

Segundo Guimarães e Kingsman [1990], a recalibração dos parâmetros A e m deve ser efectuada em simultâneo com a recalibração da função $f(n)$. Segundo os mesmos autores, não sendo possível executar simultaneamente a recalibração dos parâmetros e da função é preferível manter todas as estimativas anteriores. Atendendo a que, por um lado, são necessários, no mínimo, cerca de 1500 retornos (seis anos de preços diários) para obter razoáveis estimativas de A e m e a que, por outro, se exige um razoável esforço computacional na calibração de $f(n)$, a actualização dos parâmetros de autocorrelação e da regra de compra não constitui uma tarefa simples.

Os parâmetros A e m são estimados (Taylor [1986]) a partir da minimização da função

$$S(A,p) = \sum_{i=1}^I (r_{i,y} - A \cdot p^i)^2,$$

onde $r_{i,y}$ denota as autocorrelações estimadas dos retornos reescalados y_t (que se obtêm dividindo os retornos x_t pela estimativa corrente do seu desvio padrão) e, de acordo com Taylor [1986], I toma o valor 30.

Por sua vez, os coeficientes $r_{i,y}$ são estimados recorrendo à expressão

$$r_{i,y} = \frac{\sum_t (y_t y_{t+1})}{\sum_t (y_t^2)} \cdot N/(N-iC),$$

na qual se admite que o valor esperado dos retornos reescalados é nulo e onde a soma do numerador inclui apenas os termos correspondentes aos retornos y_t e y_{t+1} pertencentes ao mesmo contrato. N representa o número total de retornos reescalados e C o número de contratos.

O procedimento de recalibração dinâmica que a seguir se apresenta inclui uma carta de controle CUSUM para a monitorização do processo gerador dos retornos diários e, no processo de estimação de A e m , recorre à ponderação dos coeficientes amostrais dos retornos reescalados. Com este procedimento procura-se ultrapassar as dificuldades de recalibração anteriormente apontadas.

A aplicação do CUSUM (Page [1954]) basear-se-á no seguinte princípio: dado um processo gerador de retornos com parâmetros de autocorrelação A e m , a aplicação regra de compra (com $T = 0$) em períodos activos de dimensão fixa deverá conduzir a valores de PRN

com valor esperado constante. Um desvio significativo na média de PRN corresponderá a uma modificação do valor de A ou de m . A recalibração deve ser efectuada sempre que tais alterações sejam identificadas através de uma carta de controle CUSUM, construída a partir de valores de PRN provenientes de compras efectuadas de facto ou de compras simuladas. O procedimento de recalibração deverá, assim, ser mantido activo quer se executem compras quer não.

Ao CUSUM associa-se a regra de decisão descrita em Wetherill [1977] e Murdoch [1979], que se baseia na estatística "intervalo de decisão" (h) e permite concluir se as alterações na inclinação do CUSUM são significativas e, conseqüentemente, se a média de PRN se modificou.

O CUSUM é calculado a partir da média de PRN correspondente a quatro períodos consecutivos de 21 dias. A utilização desta variável têm como principal vantagem a redução da ligeira assimetria à direita que se verifica na distribuição de PRN (ver Cabral [1990]). Tal assimetria poderia aumentar a probabilidade de assinalar uma mudança da média quando esta efectivamente não se modificou. Na prática, do ponto de vista do procedimento de compra, o facto de se calcular cada valor do CUSUM com base em quatro períodos (em vez de um) não tem conseqüências sensíveis (Cabral [1990]).

Na Figura 3 ilustra-se a utilização do CUSUM e do respectivo intervalo de decisão (para detecção de descidas da média) obtidos no decurso da simulação da compra de futuros de soja na CBOT de 1978 a 1987, em períodos de 21 dias.

No final de Fevereiro de 1982 (12ª observação) é detectada uma primeira diminuição significativa da média de PRN, que, no entanto, se mantém positiva (≤ 0.575). De acordo com os pressupostos admitidos tal modificação significa que os parâmetros A e m se alteraram, implicando a sua recalibração (segundo um procedimento que se apresentará em seguida). A partir da 12ª observação, reinicia-se o processo, até que, no final de Julho de 1985 (22ª observação) o CUSUM e o intervalo de decisão permitem concluir que $E(\text{PRN}) \leq 0$. Nestas condições deve-se abandonar a política de compra proposta e adoptar a "benchmark".

Uma das vantagens adicionais do gráfico CUSUM de $\overline{\text{PRN}}$ resulta da facilidade de interpretação e acompanhamento dos resultados. À medida que o CUSUM vai sendo desenhado, o comprador dispõe de uma estimativa corrente do valor esperado de PRN. Por exemplo, a inspecção visual do CUSUM apresentado na Figura 3 mostra que, no período em causa, o valor esperado de PRN se situaria muito abaixo do valor inicialmente estimado durante o período de calibração 1972-77 ($\overline{\text{PRN}} = 1.15$). Independentemente da detecção automática de mudanças nos parâmetros do processo gerador dos retornos, a observação do gráfico poderia, por si só, sugerir a suspensão da regra de compra e o retorno à "benchmark".

Uma vez detectada uma alteração significativa no padrão de autocorrelação dos retornos, é necessário recalibrar A e m. No método de calibração habitual (utilizado, por exemplo no caso do milho) é idêntica a importância atribuída a cada retorno do período de calibração, quer este seja mais antigo ou mais recente. Daqui podem resultar inconvenientes na aplicação da regra de compra, que serão tanto mais acentuados quanto mais brusca for a variação dos parâmetros de autocorrelação.

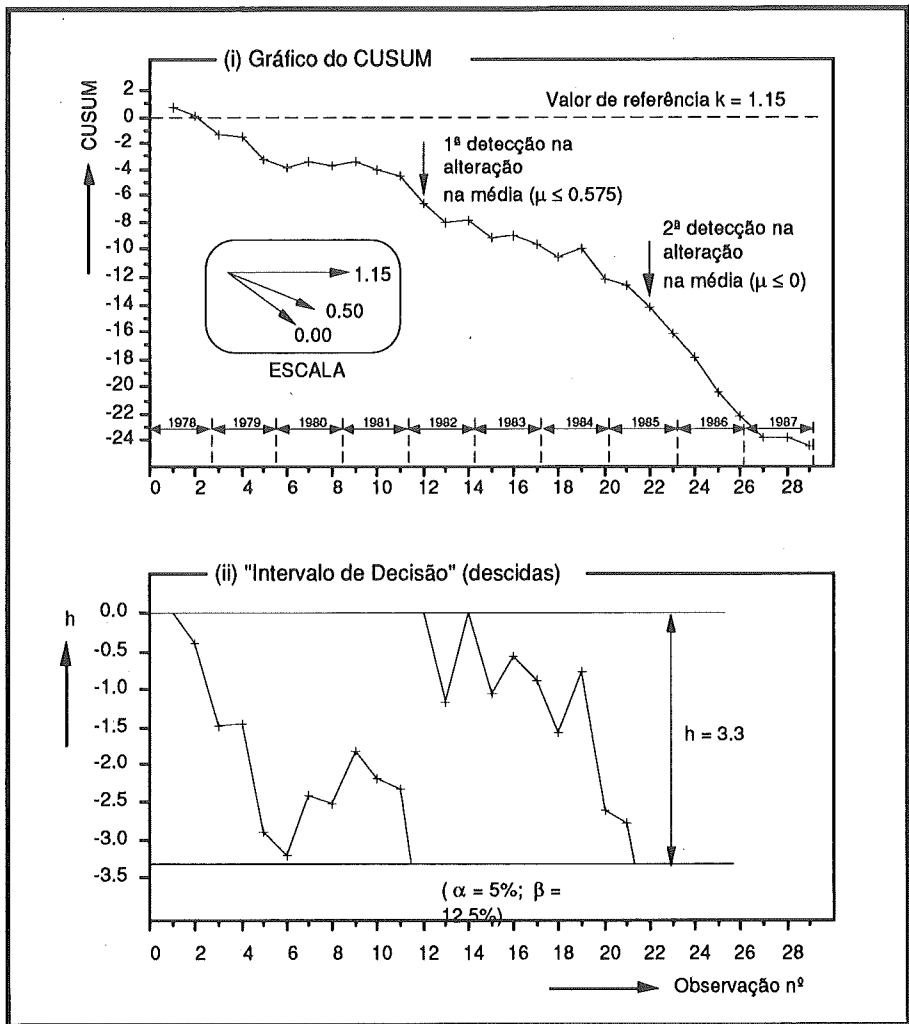


Figura 3: Gráficos do CUSUM e do respectivo intervalo de decisão (compra de futuros de soja na CBOT, períodos de 21 dias, calibração em 1972-77) (Fonte: Cabral [1990])

Esta insuficiência pode ser ultrapassada sem, no entanto, alterar a essência da técnica utilizada na estimação de A e m. Tal pode ser conseguido através do seguinte procedimento:

- (i) para cada ano j de um determinado período de calibração constituído por J anos, calculam-se as estimativas $r_{i,y,j}$ dos coeficientes de autocorrelações dos retornos y_t ;
- (ii) as estimativas globais $r_{i,y}^*$ correspondentes ao período de calibração são obtidas recorrendo a uma média ponderada dos valores de $r_{i,y,j}$ na qual os pesos decaem exponencialmente à medida que estes valores se referem a anos mais antigos:

$$r_{i,y}^* = \sum_{j=1}^J \pi_j r_{i,y,j}$$

onde os pesos podem ser expressos em função de uma constante de amortecimento exponencial, θ , de acordo com

$$\pi_j = \theta \cdot (1 - \theta)^{J-j} \cdot [1 - (1 - \theta)^J]^{-1} \quad (0 < \theta < 1);$$

- (iii) finalmente, as estimativas de A e de m são obtidas por minimização de S, após se ter substituído na respectiva expressão os coeficientes $r_{i,y}$ pelas estimativas $r_{i,y}^*$.

Adoptando para J (duração do período de calibração) o valor 10 (tal como no caso do milho), os pesos correspondentes às observações mais antigas do que os 10 anos são nulos, ficando os mais recentes dependentes do valor do parâmetro θ . Cabral [1990] sugere que se atribua a θ um valor situado entre 0.2 e 0.3.

O processo global de recalibração combina a carta de controle CUSUM com uma calibração em "rolamento" utilizando autocorrelações anuais ponderadas exponencialmente. O "rolamento" consiste em estimar A e m no início de cada ano recorrendo à técnica que se acaba de descrever (utilizando a informação dos dez anos anteriores). As estimativas assim obtidas deverão ser mantidas até ao final de cada ano, a menos que haja evidência de que a média de PRN é nula. A carta CUSUM permite, por um lado, avaliar a estabilidade ou instabilidade da função autocorrelação dos retornos e, desse modo, facilitar a escolha do valor de θ e, por outro, fundamentar a opção entre a regra proposta e a "benchmark".

Para se poder aplicar a regra de compra resta ainda decidir sobre a forma de calibrar a função $f(n)$ neste novo contexto. Atendendo à relativa insensibilidade da eficiência da regra de compra em relação a $f(n)$, calibra-se esta função da forma habitual, utilizando para esse fim séries de preços gerados artificialmente com as estimativas dos parâmetros A e m que vão resultando do "rolamento".

Na Tabela 2 apresentam-se os resultados da compra efectuada com base neste procedimento de recalibração (com $\theta = 0.2$ e $\theta = 0.3$), simulando a sua aplicação à compra de futuros de soja na Bolsa de Chicago, entre 1982 e 1987.

(i) $\theta = 0.2$								
	$\overline{\text{PRN}}$		$\hat{A}[\%]$	\hat{m}	$f(n)$			
	21 Dias	42 Dias			τ	ω		
1982	0.73	1.86	(72-81)	1.47	9	19	2.0	
1983	0.31	0.20	(73-82)	1.55	12	30	2.0	
1984	0.83	-0.09	(74-83)	1.77	14	30	1.4	
1985 ^(a)	0.52	1.57	(75-84)	1.16	11	30	2.0	
1986	0	0	-	-	-	-	-	
1987	0	0	-	-	-	-	-	
Média de $\overline{\text{PRN}}$	0.40	0.59						
Utilidade média	0.42	0.62						

(ii) $\theta = 0.3$								
	$\overline{\text{PRN}}$		$\hat{A}[\%]$	\hat{m}	$f(n)$			
	21 Dias	42 Dias			τ	ω		
1982	0.75	1.89	(72-81)	1.47	9	19	2.0	
1983	0.29	0.16	(73-82)	0.93	12	30	2.0	
1984	0.83	-0.11	(74-83)	1.68	14	30	1.4	
1985 ^(a)	0.82	2.13	(75-84)	1.05	9	15	2.0	
1986	0	0	-	-	-	-	-	
1987	0	0	-	-	-	-	-	
Média de $\overline{\text{PRN}}$	0.45	0.68						
Utilidade média	0.47	0.70						

(^a) Adopção da "benchmark" a partir de Agosto

Tabela 2: Resultados que derivam do procedimento de recalibração proposto na compra de futuros de soja no período 1982-87 (Fonte: Cabral [1990])

Apresentam-se ainda os valores dos parâmetros A, m, τ e ω utilizados em cada ano, bem como as utilidades médias que seriam obtidas se fosse adoptada a curva na Figura 1 (i), fazendo $T = 0$.

O valor médio de PRN no período de teste (1982-87) é agora positivo, contrariamente ao que se passaria se as compras naqueles anos fossem efectuadas com as estimativas obtidas no período inicial de calibração 1972-81. Por outro lado, verifica-se que a utilização de $\theta = 0.3$ conduziria a melhores resultados globais. No entanto, as diferenças que resultam da adopção de um ou outro valor de θ só são relevantes em 1985, que, de acordo com o CUSUM, corresponde a um ano precedido por uma alteração significativa dos parâmetros A e m. Este resultado ilustra a vantagem de adoptar um valor elevado para θ em períodos de instabilidade de A e m.

5. Conclusões

Nesta comunicação procurou descrever-se o resultado de uma longa sequência de trabalhos relacionados com a compra de mercadorias através do mercado de futuros. De tais trabalhos – englobando quer os processos de previsão de preços quer o desenvolvimento de heurísticas capazes de converter tais previsões em decisões de compra – emerge um resultado central: existem mercadorias para as quais a política de compra proposta permite obter reduções significativas no seu custo de aquisição, mesmo por compradores razoavelmente avessos ao risco. Põe-se assim em causa o "dogma" da eficiência dos mercados, demonstrando-se a pertinência da adopção de políticas "activas" de compra.

Com o procedimento de recalibração dinâmica apresentado, este resultado torna-se extensível a mercados nos quais os processos geradores dos preços são instáveis.

6. Referências

- [1] Cabral, J.A.S., *Análise, Desenvolvimento e Teste de Regras de Compra de Mercadorias Sujeitas a Flutuações de Preço*, Tese de Doutoramento, Faculdade de Engenharia da Universidade do Porto, 1990.
- [2] Cabral, J.S.A. e Guimarães, R.M.C., *Are Commodity Futures Markets Really Efficient? A Purchasing-Oriented Study of the Chicago Corn Futures Market*, *The Review of Futures Markets*, 7 (Supplement) (1988) 598-619 (Chicago Board of Trade).
- [3] Cabral J.A.S. e Guimarães, R.M.C., *Purchasing Maize Futures Under a Deadline: Testing and Risk-Yield Evaluation of a Price-Trend Buying Policy*, *A Reappraisal of the Efficiency of Financial Markets*, Eds. R.M.C.Guimarães, B.G.Kingsman e S.J.Taylor, NATO ASI Series, Vol. F54 (1989) 699-733 (Springer-Verlag, Berlin).
- [4] Fama, E.F., *Efficient Capital Markets: A Review of Theory and Empirical Work*, *Journal of Finance*, 25 (1970) 383-417.
- [5] Guimarães, R.M.C., *National Planning of Commodity Import Operations: The Case of Portugal's Feed Grain Imports*, Tese de PhD., Universidade de Lancaster, Reino Unido, 1981.
- [6] Guimarães, R.M.C. e Kingsman, B.G., *A Day to Day Buying Policy for Commodities – A Study of Purchasing Maize*, *Journal of the Operational Research Society*, 41, nº 1, (1990) 25-37.
- [7] Guimarães, R.M.C, Kingsman, B.G. e Taylor, S.J., *The NATO Workshop on a Reappraisal of the Efficiency of Financial Markets*, *A Reappraisal of the Efficiency of Financial Markets*, eds. R.M.C.Guimarães, B.G.Kingsman e S.J.Taylor, NATO ASI Series, Vol.F54 (1989) 1-24 (Springer-Verlag, Berlin).
- [8] Jensen, M.C., *Some Anomalous Evidence Regarding Market Efficiency*, *Journal of Finance Economics*, 6 (1978) 95-101.
- [9] Kingsman, B.G., *Raw Materials Purchasing – An Operational Research Approach* (Pergamon Press, Oxford, England), 1985.
- [10] Murdoch, J., *Control Charts* (the Macmillan Press, London), 1979.
- [11] Page, E.S., *Continuous Inspection Schemes*, *Biometrika*, 41 (1954) 100-115.
- [12] Taylor, S.J., *Time Series Properties and Models of Commodity Prices*, Tese de PhD., Universidade de Lancaster, Reino Unido, 1978.
- [13] Taylor, S.J., *Conjectured Models for Trends in Financial Prices, Tests and Forecasts*, *The Journal of the Royal Statistical Society*, 143A (1980) 338-362.
- [14] Taylor, S.J., *Modelling Financial Time Series* (John Wiley & Sons, Chicester), 1986.
- [15] Taylor, S.J., *How Efficient are the Most Liquid Futures Contracts? A Study of Treasury Bond Futures*, *The Review of Futures Markets*, 7 (Supplement), (1988) 598-619 (Chicago Board of Trade).
- [16] Taylor, S.J. e Tari, A., *Further Evidence Against the Efficiency of Futures Markets*, *A Reappraisal of the Efficiency of Financial Markets*, ed. R.M.C.Guimarães, B.G.Kingsman e S.J.Taylor, NATO ASI Series, Vol.F54 (1989) 577-605 (Springer-Verlag, Berlin).
- [17] Wetherill, G.B., *Sampling Inspection and Quality Control* (Chapman and Hall, London), 1977.
- [18] von Neumann, J. e Morgenstern, O., *Theory of Games and Economic Behaviour* (John Wiley & Sons, New York), 1967.

SIMULATED ANNEALING ALGORITHM FOR THE SIMPLE PLANT LOCATION PROBLEM: A COMPUTATIONAL STUDY

Maria Leonor Alves
Instituto Superior de Engenharia de Coimbra

Maria Teresa Almeida
Instituto Superior de Economia e Gestão (UTL)

Abstract

The simulated annealing algorithm has recently been used in the approximate solution of several combinatorial optimization problems. In some of them it outperformed more traditional heuristics whereas in other cases its performance was much poorer.

In this study we compare the results of four versions of the simulated annealing algorithm with those of some well known heuristic methods on a set of simple plant location problems. The quality of the simulated annealing solutions is very good but the computing time is significantly larger when compared to the others.

Resumo

O algoritmo 'simulated annealing' tem vindo a ser usado na resolução aproximada de vários problemas de optimização combinatoria. Nalguns casos mostrou-se superior a heurísticas mais tradicionais enquanto noutros não foi competitivo.

Neste estudo são apresentados resultados computacionais para um conjunto de problemas de localização simples retirados da literatura. São comparados os resultados para quatro versões do algoritmo 'simulated annealing' com os obtidos com métodos heurísticos já conhecidos. As soluções obtidas por 'simulated annealing' são de boa qualidade mas requerem maior tempo de computação.

Keywords

Simulated annealing; simple plant location problem; heuristics.

1 - Introduction

The simulated annealing algorithm was developed by Kirkpatrick, Gelatt and Vecchi (1983) and by Cerny (1985) for the approximate solution of combinatorial optimization problems. It is based on the analogy that can be established between the search for a local optimal solution of a combinatorial problem and the behaviour of a physical system when its temperature is first raised to a very high level and then successively lowered until the system is frozen. This analogy explains why the typical simulated annealing terminology includes expressions such as 'temperature', 'cooling schedule', etc.

However the simulated annealing method can also be interpreted as a modified version of the traditional local search procedures, aimed at avoiding the poor local optima these methods may get trapped in.

The algorithm has been applied to a vast number of typical combinatorial optimization problems: the travelling salesman problem, the quadratic assignment problem, the graph colouring problem, etc. It has also been applied to engineering problems such as VLSI design

and image processing problems, Aarts and Korst (1989). For the combinatorial optimization problems computational studies are required to access the performance of simulated annealing compared to the performance of other heuristics. This comparison has been made for a number of problems with quite different outcomes depending on the problem considered, Jonhson et al. (1991).

In this paper we present a computational study of the simulated annealing algorithm applied to the simple plant location (SPL). The SPL is a NP-Hard problem, Garey and Jonhson (1979), for wich very effective approximate methods are known, Erlenkotter (1978), Bilde and Krarup (1977). So it seems an appealing ground for the investigation on the new algorithm's performance. In our study we compare the results given by four versions of the simulated annealing algorithm, on a set of test problems taken from the literature, with those given by some other heuristics.

It should be noted that, both in terms of computing time and of solution quality, the results obtained with the simulated annealing algorithm are strongly dependent on the choices made for a set of parameters (to be described later). Although we devoted some effort to the tuning of such parameters we can by no means claim we have achieved the best tuning. With our implementations we were able to produce very good quality solutions for every instance tried but with computing times larger than those of other methods.

In section 2 we present a brief description and a formulation of the SPL. In section 3 we describe the simulated annealing algorithm and the details of its application to the SPL. In section 4 we present the computational results and in section 5 the concluding remarks

2 – Simple Plant Location Problem

The simple plant location problem (SPL) is that of choosing, among m possible sites, the locations for a set of facilities to service a set of n customers, minimizing the total cost incurred. The total cost includes fixed costs for locating (or opening) the facilities and variable costs for servicing the customers from the open facilities.

Let $I = \{1, 2, \dots, m\}$ be the set of possible locations and $J = \{1, 2, \dots, n\}$ be the set of customers. Let c_{ij} ($i \in I, j \in J$) be the cost of supplying customer j from a facility located at i (it may include production costs, trnsportation costs, etc.) and let f_i ($i \in I$) be the fixed cost for locating a facility at location i .

Define:

$$y_i = \begin{cases} 1, & \text{if a facility is located at } i \\ 0, & \text{otherwise} \end{cases}$$

$$x_{ij} = \begin{cases} 1, & \text{if customer } j \text{ is supplied from } i \\ 0, & \text{otherwise} \end{cases}$$

The SPL may be formulated as:

$$\min \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} + \sum_{i \in I} f_i y_i \quad (1)$$

$$\text{s. t. } \sum_{i \in I} x_{ij} = 1 \quad j \in J \quad (2)$$

$$y_i - x_{ij} \geq 0 \quad i \in I, j \in J \quad (3)$$

$$y_i \in \{0, 1\} \quad i \in I \quad (4)$$

$$x_{ij} \in \{0, 1\} \quad i \in I, j \in J \quad (5)$$

Constraints (2) ensure that each customer is supplied from one facility and constraints (3) ensure that customers are supplied only from open facilities. Constraints (4) and (5) define the variables as binary and (1) represents the total cost to be minimized.

Erlenkotter developed two heuristic methods that produce very good feasible solutions for the SPL. The first method is a primal-dual constructing type method. The second is an improvement type method based on the dual solution. Often it can be shown that the solution obtained is optimal. If it is not the case a branch and bound algorithm that embodies the bounds already at hand completes the search for an optimal solution. A similar method, without the improvement heuristic, was independently presented by Bilde and Krarup (1977). For a survey on location problems and a detailed presentation of these methods see Alves (1991).

3 – Simulated Annealing Algorithm for the Simple Plant Location Problem

The simulated annealing algorithm was imported to the Combinatorial Optimization field from the Statistical Mechanics area. Behind it is the idea of establishing an analogy between two activities apparently with no point in common: the growing of a crystal and the search for an optimal solution.

To grow a crystal one first raises the temperature to a very high level to obtain a melt that is then cooled, or annealed, until the freezing temperature is reached. If the procedure is performed very quickly (rapid quenching) the resulting structure tends to be very rough (glass) and its energy level tends to be higher. On the other hand, if the procedure is performed slowly with the temperature level going down step by step and allowing for equilibrium to be achieved at each level (careful annealing) then a better structured crystal tends to be obtained (and therefore a lower energy level). In the first case the system is rushed to a state from which no way out is possible, even if it is well away from the global minimum energy value. In the second case the system is given the opportunity to try its way to a low energy state, even if at first sight some of the moves do not seem rewarding.

The traditional local search procedures can be interpreted as a rapid quenching. The equivalent to a careful annealing will then be a procedure that applies a less greedy, more wise, search strategy.

Such a procedure can be sketched as follows:

Initialization

Find an initial solution, S , and its cost, $c(S)$.

Choose an initial temperature, T_0 , and a temperature length, L .

$T := T_0$;

Iteration

WHILE not frozen **DO**

BEGIN

$\lambda := 1$;

REPEAT

Choose at random a solution, S' , in a neighbourhood of S ;

Compute its cost, $c(S')$;

IF $\Delta = c(S') - c(S) \leq 0$ **THEN**

$S := S'$;

ELSE

$S := S'$ with probability $p = \exp(-\Delta/T)$;

ENDIF

$\lambda := \lambda + 1$;

UNTIL ($\lambda = L$)

Reduce the temperature;

END

The real temperature and its physical lowering mechanism as well as the time spent each temperature level are represented by parameters which keep their original names to emphasize the analogy. The freezing point is said to have been reached when the chance of being able to move away from the current solution seems to get too small (we will elaborate more on that later). The added flexibility on the search strategy comes from the randomization of the solution acceptance criterion. A new solution is accepted if its cost is better than the current one and may also be accepted if it is not. Increasing cost moves are accepted with a probability that decreases

as the temperature goes down and that also depends on the magnitude of the solution value increase.

To implement the algorithm for the solution of a particular combinatorial problem it is necessary to specify what a solution and its costs are and to define the neighbourhood to be considered. It is also necessary to set the values of the parameters and to adopt a stopping criterion.

Solutions

For the SPL it seems adequate to consider only feasible solutions. This may not be the case for other problems, Johnson et al. (1991). A feasible solution, S , may be defined as the set of open facilities, i.e., $S = \{i \in I: y_i = 1\}$ as the corresponding x_{ij} variables values are trivial to obtain from it, assigning each customer j to the nearest open facility.

Neighbourhoods

As the cardinality of the set of open facilities in the optimal solution is not known in advance we adopt a definition of neighbourhood that allows for different cardinalities. Given a feasible solution, S , its neighbourhood, $\mathcal{N}(S)$, is defined as

$$\chi(S) = \{S' \subset I : |S' - S| \leq 1 \wedge |S - S'| \leq 1\}$$

This means that a solution S' is in the neighbourhood of a solution S if it can be obtained from S by closing a facility, by opening a facility or by closing a facility and opening a new one simultaneously.

In practice every time a new solution $S' \in \mathcal{N}(S)$ is needed we first randomly choose which of the three strategies to follow and then randomly select the facility to be opened/closed.

Cooling Schedule

We adopted the geometric temperature reduction scheme. The temperature level was first set at an initial value, T_0 , and then successively reduced after L iterations, according to the formula $T_k = T_0 \times r^k$, where k represents the number of temperature levels already considered and $r < 1$ is a reduction factor. This reduction scheme was proposed by Kirkpatrick et al. (1983). Other reduction schemes have also been tried. However according to our previous experience as well to the one reported in Johnson et al. (1989) there seems to be no reason to prefer them.

The choice for the values of the parameters T_0 , r and L was made based on some computational experiments to be reported at the beginning of next section.

4 – Computational Results

In our computational study we used three sets of three problems each, taken from the literature.

The three problems in the first set are identified as A_1 , A_2 and A_3 and were taken from Karg and Thompson (1964). They have 33 customers and 33 possible locations and fixed costs of

2000.0, 2500.0 and 3000.0, respectively. The three problems in the second set (B_1 , B_2 and B_3) have 57 customers and 57 possible locations and fixed costs 3000.0, 4000.0 and 5000.0, respectively. They were also taken from Karg and Thompson (1964).

The last three problems (C_1 , C_2 and C_3) have 100 customers and 100 possible locations and fixed costs of 1000.0, 2000.0 and 3000.0, respectively. They were built from the set of 100 European cities used by Aarts and Korst (1989) computing the distance matrices as in Grottschel and Holland (1991).

Next we present some aspects of the parameter tuning. For more details see Alves (1991).

4.1. Cooling Schedule Design

Initial Temperature

The choice of the initial temperature, T_0 , is usually made according to the acceptance ratio $\chi(T_0)$, it yields. For a temperature level T the acceptance ratio is defined as:

$$\chi(T) = \left(\frac{\text{number of solutions accepted at } T}{\text{number of solutions generated at } T} \times 100 \right) \%$$

Our previous computational experience indicated that temperatures yielding acceptance ratios greater than 50% cause an increase in computing time that is not rewarded by better quality solutions. For the same initial temperature the acceptance ratio depends on the initial solution considered. We tried two alternatives for the initial solution: a randomly generated solution (as in most implementations we are aware of) and the solution produced by Erlenkotter's primal-dual heuristic. Tables I and II show the average acceptance ratios (five runs) obtained for a set of eight temperatures in the range [1000, 15000].

T_0	A_1	A_2	A_3	B_1	B_2	B_3	C_1	C_2	C_3
1000	29.1	22.5	19.7	15.8	14.8	11.5	11.6	7.4	6.2
2000	35.0	32.2	26.6	22.9	20.5	20.1	19.7	12.5	10.3
4000	45.5	45.9	39.5	32.8	33.9	33.1	35.3	21.1	18.1
6000	56.3	50.6	46.9	39.7	37.3	36.5	44.7	28.9	28.8
8000	58.1	59.4	53.0	48.1	46.6	43.5	44.3	35.0	32.5
10000	63.7	59.4	60.7	52.9	50.2	47.7	52.2	43.1	38.8
12000	68.4	68.2	65.5	55.8	51.9	48.9	53.9	46.4	42.5
15000	72.3	70.1	69.0	58.5	57.2	57.7	56.4	54.6	48.2

Table I – $\chi(T_0)$

Initial solution: randomly generated

T_0	A_1	A_2	A_3	B_1	B_2	B_3	C_1	C_2	C_3
1000	12.1	14.6	16.7	10.7	8.6	5.9	2.0	6.1	1.1
2000	25.0	19.8	19.3	19.5	11.7	12.4	3.6	9.3	3.0
4000	40.4	32.7	37.3	29.2	21.8	20.5	11.3	16.8	6.5
6000	49.7	44.2	42.0	32.4	31.4	29.3	18.2	18.3	11.6
8000	50.9	54.0	45.6	42.8	38.6	32.8	28.7	24.0	17.6
10000	53.4	52.3	55.3	42.6	43.7	37.8	35.7	31.5	19.4
12000	62.3	64.1	60.6	49.8	43.4	42.6	38.8	32.3	24.9
15000	70.7	66.8	61.8	52.9	52.1	49.3	39.5	40.0	30.7

Table II – $\chi(T_0)$

Initial solution: primal-dual heuristic

As expected, for the same temperature, ratios are higher for randomly generated initial solutions.

Temperature Reduction Factor and Temperature Length

The temperature reduction factor, r , controls the rate of decrease of the temperature. To avoid poor local optima the procedure must be performed smoothly, i.e., r shall not be too small. According to our experience the best values are $0.70 \leq r \leq 0.95$. Values smaller than 0.70 yield poorer solutions and values greater than 0.95 cause larger computing times but do not provide in general significantly better solutions.

The temperature length, L , determines the number of solutions generated at each temperature level. We tried two alternatives for it. In the first alternative we set $L = |\mathcal{N}(S)|$ independently of the temperature level. In the second alternative we increased the temperature length progressively, setting $L = |\mathcal{N}(S)| \times (1.1)^k$ where k is the number of temperature levels already considered. The results obtained, for randomly generated initial solutions, considering 5, 10 and 15 temperature levels and four different values for the reduction factor are presented below.

In 2 out of 3 cases the values obtained after considering 10 temperature levels and $L = |\mathcal{N}(S)| \times (1.1)^k$ are at least as good as the values obtained after considering 15 temperature levels and $L = |\mathcal{N}(S)|$. Similar results were obtained when the initial solution was the one given by the primal-dual heuristic. This seems to indicate that increasing moderately the temperature length as the temperature goes down may significantly reduce the number of temperature levels required to obtain good quality solutions.

Problem	Temperature Reduction Factor				k
	0.70	0.80	0.90	0.95	
A ₁	21285	20783	21219	21073	5
A ₁	20265	20756	21037	20285	10
A ₁	20509	20756	20933	20825	15
A ₂	22989	23020	22861	23381	5
A ₂	22989	22361	22391	22418	10
A ₂	22127	22361	22391	22418	15
A ₃	24599	24858	24858	24858	5
A ₃	23999	24294	23474	23474	10
A ₃	23474	23627	23474	23474	15
B ₁	33724	32684	33061	33593	5
B ₁	32547	32684	32562	33066	10
B ₁	32547	32684	32562	32881	15
B ₂	36354	35909	35962	36354	5
B ₂	35623	35909	35962	36354	10
B ₂	35623	35909	35692	35649	15
B ₃	38968	39876	39934	39070	5
B ₃	38617	39344	38918	39070	10
B ₃	38617	38742	38881	38070	15
C ₁	37094	37813	37528	37813	5
C ₁	36601	37813	37528	37813	10
C ₁	34641	37813	37303	37813	15
C ₂	47740	47287	48118	49114	5
C ₂	46844	47283	48118	46845	10
C ₂	46844	46466	48118	46845	15
C ₃	55469	56872	57219	56655	5
C ₃	55469	55953	55858	56655	10
C ₃	55469	54481	55843	55822	15

Table III – Solution costs

$$L = |N(S)|$$

Problem	Temperature Reduction Factor				k
	0.70	0.80	0.90	0.95	
A ₁	21285	20783	21219	21073	5
A ₁	20265	20726	20813	20652	10
A ₁	20363	20393	20633	20609	15
A ₂	22989	23020	22861	23381	5
A ₂	22127	22252	22474	22474	10
A ₂	22127	22252	22252	22299	15
A ₃	24599	24858	24858	24858	5
A ₃	23474	23474	23474	23474	10
A ₃	23474	23474	23474	23474	15
B ₁	32814	32804	33097	32779	5
B ₁	32156	32658	32290	32779	10
B ₁	32136	32547	32156	32779	15
B ₂	35909	35962	36270	36309	5
B ₂	35547	35777	36270	36309	10
B ₂	35547	35617	35623	35617	15
B ₃	39344	38742	38617	39070	5
B ₃	38742	38742	38617	38981	10
B ₃	38617	38547	38617	38881	15
C ₁	37800	37813	36717	37813	5
C ₁	37800	36795	36717	37585	10
C ₁	37800	36795	36717	37585	15
C ₂	48538	48946	48894	48238	5
C ₂	45619	46557	47552	47821	10
C ₂	45619	46557	47181	47821	15
C ₃	55464	55634	56136	57572	5
C ₃	55464	55379	54880	56425	10
C ₃	55464	53641	54779	56202	15

Table IV – Solution costs

$$L = |\mathcal{N}(S)| \times (1.1)^k$$

Freezing Temperature

The freezing point, i.e., the stopping criterion, is usually established based on the value of the acceptance ratio: when a temperature level is reached for which this ratio drops below a given threshold the system is considered frozen and the procedure stopped. However in our computational experiments we observed that often a decrease in the temperature is not followed by a decrease in the acceptance ratio. So we adopted a more conservative criterion: we declare the system frozen only after three consecutive temperature levels with acceptance ratios less than or equal to a threshold of 0.02.

4.2. Final Results

Tables V and VI show the final results obtained with the different versions of the simulated annealing algorithm, computed as:

$$\left(\frac{Z_H - Z_B}{Z_B} \times 100 \right) \%$$

where Z_B represents the optimum cost (except for problems C2 and C3 for which we do not know the optimum and therefore we use a lower bound) and Z_H represents the cost given by the simulated annealing version considered.

In implementation ANNEAL1 the temperature length is $L = |N(S)|$ and 10 runs were executed. In implementation ANNEAL2 the temperature length is $L = |N(S)| \times (1.1)^k$ and 5 runs only were executed. Table V presents the results using a randomly generated initial solution and Table VI presents the results for the primal-dual heuristic initial solution.

In each case the best (BST), the average (AVR) and the worst (WST) results over the 10 or 5 runs are shown as well as the average computing time (in seconds) per run on a VAX3600 machine.

Prob.	ANNEAL1				ANNEAL2			
	BST	AVR	WST	time	BST	AVR	WST	time
A ₁	0.00	1.32	3.12	8.7	0.00	0.30	1.21	17.0
A ₂	0.00	1.21	4.32	5.4	0.00	0.49	1.06	10.4
A ₃	0.00	0.25	1.35	5.8	0.00	0.13	0.65	11.9
B ₁	0.00	0.85	2.43	47.2	0.00	0.58	1.28	67.1
B ₂	0.20	0.79	2.17	35.8	0.00	0.66	2.31	49.0
B ₃	0.00	0.57	1.44	32.0	0.51	0.71	1.24	43.3
C ₁	0.82	4.50	8.34	641.5	2.59	4.94	8.22	949.9
C ₂	0.23	2.62	5.60	501.2	1.77	4.25	7.44	488.1
C ₃	2.14	3.76	8.87	330.4	0.08	1.96	3.58	643.9

Table V – Solution cost gaps
Initial solution: randomly generated

ANNEAL1					ANNEAL2			
Prob.	BST	AVR	WST	time	BST	AVR	WST	time
A ₁	0.00	1.02	3.68	10.0	0.00	0.89	3.17	17.1
A ₂	0.00	1.04	1.52	5.1	0.00	0.30	0.78	10.6
A ₃	0.00	0.27	1.38	5.8	0.00	0.28	1.38	10.2
B ₁	0.77	1.41	2.71	44.8	0.00	0.82	2.22	69.3
B ₂	0.00	0.54	1.45	36.9	0.00	0.00	0.00	63.3
B ₃	0.00	0.78	2.28	30.8	0.00	0.23	0.53	54.8
C ₁	1.92	4.22	7.00	427.2	0.52	3.38	6.98	707.9
C ₂	1.21	3.33	7.93	457.2	1.06	3.89	7.44	488.1
C ₃	1.78	3.88	7.62	277.2	0.08	1.46	3.21	643.9

Table VI – Solution cost gaps
Initial solution: primal-dual heuristic

Implementation ANNEAL2 produces, on average, better quality WST solution but is more time consuming.

As already mentioned the simulated annealing algorithm may be considered a randomized version of the traditional local search technique aimed at avoiding poor local optima. In tables VII and VIII the results of the simulated annealing are compared with those of a similar local search algorithm where only improved solutions are accepted.

ANNEAL1					LOCALOPT			
Prob.	BST	AVR	WST	time	BST	AVR	WST	time
A ₁	0.00	1.32	3.12	8.7	0.00	2.21	3.95	0.7
A ₂	0.00	1.21	4.32	5.4	0.74	1.92	3.26	0.5
A ₃	0.00	0.25	1.35	5.8	1.35	2.39	3.89	0.5
B ₁	0.00	0.85	2.43	47.2	0.94	2.98	7.60	3.0
B ₂	0.20	0.79	2.17	35.8	0.57	2.18	4.42	2.5
B ₃	0.00	0.57	1.44	32.0	0.26	2.31	7.43	2.3
C ₁	0.82	4.50	8.34	641.5	1.56	2.73	4.41	46.5
C ₂	0.23	2.62	5.60	501.2	0.15	1.67	3.66	30.1
C ₃	2.14	3.76	8.87	330.4	0.89	2.76	4.28	25.2

Table VII – Solution cost gaps
Initial solution: randomly generated

Prob.	ANNEAL1				ANNEAL2			
	BST	AVR	WST	time	BST	AVR	WST	time
A ₁	0.00	1.02	3.68	10.0	0.00	0.77	2.13	0.8
A ₂	0.00	1.04	1.52	5.1	0.00	0.68	1.55	0.6
A ₃	0.00	0.27	1.38	5.8	1.65	1.95	2.36	0.6
B ₁	0.77	1.41	2.71	44.8	0.00	0.55	0.87	3.7
B ₂	0.00	0.54	1.45	36.9	0.00	0.37	1.17	3.1
B ₃	0.00	0.78	2.28	30.8	0.00	0.50	1.79	2.5
C ₁	1.92	4.22	7.00	427.2	0.00	0.10	0.29	53.1
C ₂	1.21	3.33	7.93	457.2	0.12	0.15	0.25	39.6
C ₃	1.78	3.88	7.62	277.2	0.08	0.20	0.30	30.8

Table VIII – Solution cost gaps
Initial solution: primal-dual heuristic

LOCALOPT globally produced better results but failed to find some global optima reached by ANNEAL1. The performance of LOCALOPT is possibly due to the simplicity of the constraints that define the feasible region of the SPL.

The improvement method developed by Erlenkotter (1978) rather than searching neighbourhoods generates new primal solutions based on violated complementary conditions. In table IX we compare the improvement, over the initial solution cost, achieved by an enhanced version of it, Van Roy and Erlenkotter (1982), with the improvement over the same solution cost, achieved by the simulated annealing algorithm. The results are favourable to the annealing but at a high computational burden.

Problem	ANNEAL1	ANNEAL2	DUAL ADJ
A ₁	4.12	4.12	3.98
A ₂	4.37	4.37	2.04
A ₃	13.56	13.56	13.56
B ₁	3.63	4.37	4.37
B ₂	2.48	2.48	2.48
B ₃	1.07	1.07	0.87
C ₁	1.67	1.67	1.84
C ₂	8.00	8.13	7.35
C ₃	0.11	1.77	0.96

Table IX – Solution Cost Improvement
($\Delta\%$ over primal-dual initial solution cost)

5 – Conclusions

Computational studies are required to assess the performance of the simulated annealing for the solution of combinatorial optimization problems.

In this paper we made a computational study of two implementations of the simulated annealing algorithm for the Simple Plant Location Problem, both using randomly generated and primal-dual heuristic initial solutions.

We tested them on three sets, of three problems each, taken from the literature.

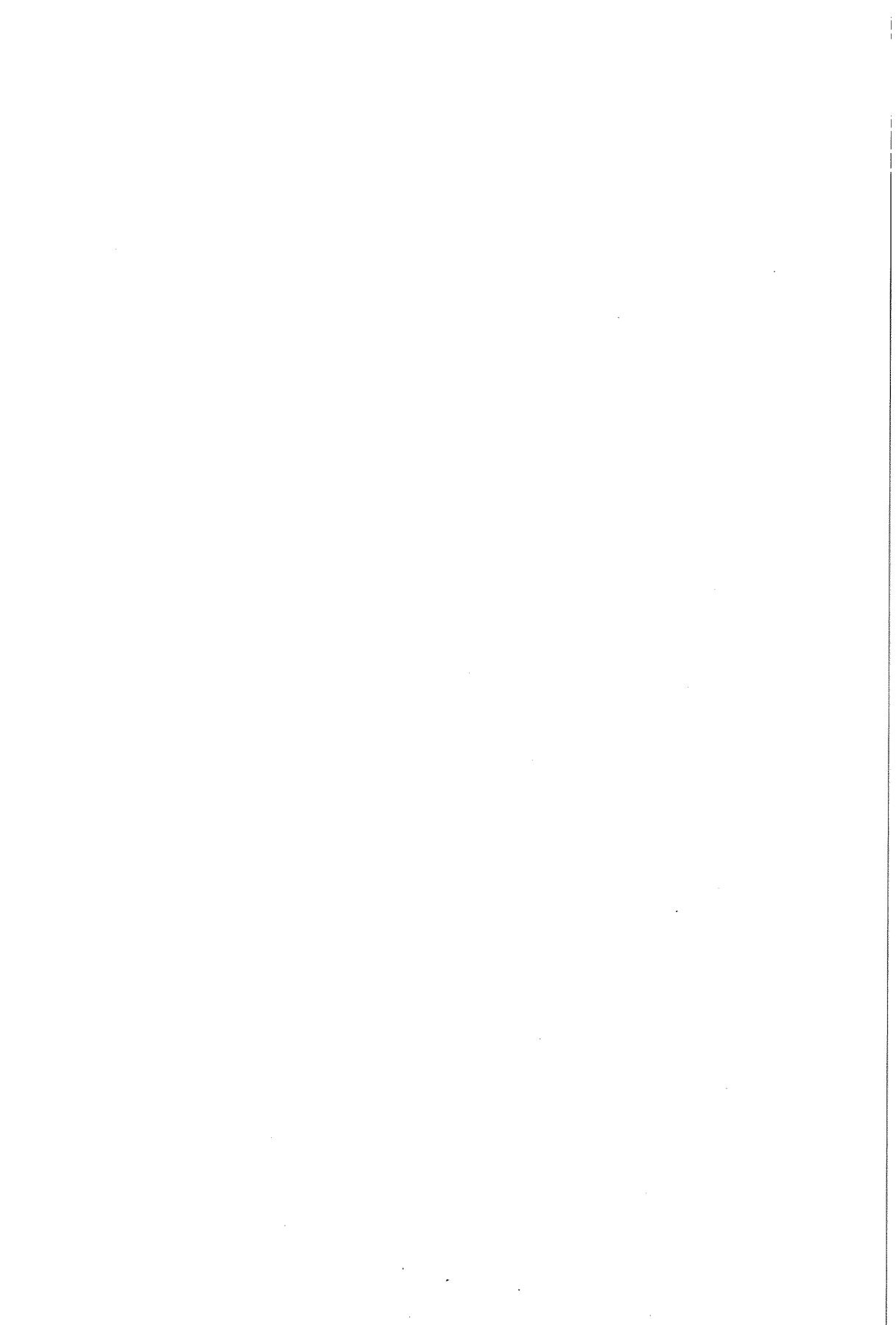
Our limited experience produced mixed results. As far as solution quality is concerned the results were very positive. For every instance tried, both using a randomly generated and a primal-dual initial solution, at least one of the annealing implementations found a solution within 2% of the optimum. These results surpassed, in several cases, those of other heuristics, known to be very effective for the SPL. The results were not equally satisfactory in what computing time is concerned.

Although large computing time were also observed for other problems, Johnson et al. (1991), we believe there is still room for improvement. Besides a better parameter tuning we intend to try to combine this approach with some tabu search rules in order to prevent solution repetitions unavoidable by the simulated annealing methods on its own.

Acknowledgment – This research was partially supported by Instituto Nacional de Investigação Científica.

References

- [1] Aarts, E. and Korst, J., *Simulated Annealing and Boltzman Machines - A Stochastic Approach to Combinatorial Optimization and Neural Computing*, J.Wiley & Sons (1989).
- [2] Alves, M.L., *Soluções Aproximadas para o Problema de Localização Simples - Algoritmo Simulated Annealing*, Master Dissertation, Instituto Superior de Economia e Gestão, Universidade Técnica de Lisboa (1991).
- [3] Bilde, O. and Krarup, J., *Sharp Lower Bounds and Efficient Algorithms for the Simple Plant Location Problem*, *Annals of Discrete Mathematics* (1977) 79-97.
- [4] Cerny, V., *Thermodynamical Approach to the Travelling Salesman Problem: an Efficient Simulation Algorithm*, *Journal of Optimization Theory and Applications* 45 (1985) 41-51.
- [5] Erlenkotter, D., *A Dual Based Procedure for Uncapacitated Facility Location*, *Operations Research* 26 (1978) 992-1010.
- [6] Garey, M. and Johnson, D., *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman (1979).
- [7] Grotschel, M. and Holland, O., *Solution of Large-Scale Symmetric Travelling Salesman Problems*, *Mathematical Programming* 51 (1991) 141-202.
- [8] Johnson, D.; Aragon, C.; McGeoch, L. and Schevon, C., *Optimization by Simulated Annealing: An Experimental Evaluation. Part I, Graph Partitioning*, *Operations Research* 37 (1989) 865-892.
- [9] Johnson, D.; Aragon, C.; McGeoch, L. and Schevon, C., *Optimization by Simulated Annealing: An Experimental Evaluation. Part II, Graph Coloring and Number Partitioning*, *Operations Research* 39 (1991) 378-406.
- [10] Karg, L. and Thompson, G., *A Heuristic Approach to Solving Travelling Salesman Problem*, *Management Science* 10 (1964) 225-248.
- [11] Kirkpatrick, S.; Gelatt, C. and Vecchi, M., *Optimization by Simulated Annealing*, *Science* 220 (1983) 671-680.
- [12] Van Roy, T. and Erlenkotter, D., *A Dual-Based Procedure for Dynamic Facility Location*, *Management Science* 28 (1982) 1091-1105.



THE MULTI-DEPOT TRAVELING SALESMAN PROBLEM

J. Orestes Cerdeira
Instituto Superior de Agronomia
Tapada da Ajuda
1399 Lisboa Codex

Abstract

The multi-depot traveling salesman problem is the special case of the multi-depot vehicle routing problem, where only one vehicle of unlimited capacity is stationed in each depot. The objective is to design a route for each vehicle in such a way that each customer is visited exactly once, and that the total distance traveled is minimum.

Here we present a way for determining lower bounds on the distance of the optimal solutions that may be viewed as a generalization, for the case of more than one depot, of the 1-tree procedure for obtaining lower bounds to the traveling salesman problem. We also describe how to assign customers to depots in $O(n^2 + k^2n)$ time, where n and k denote the number of customers and depots, respectively, proving that if the triangle inequality holds, using the clusters defined from that assignment it is possible to determine a feasible solution whose total distance is not greater than $2(n-k+1) / (n-k+2)$ times the distance corresponding to the optimal solution. Finally, we report some computational experiments in order to evaluate the quality of these bounds.

Resumo

O problema do caixeiro viajante com múltiplos depósitos é o problema de optimização de percursos de distribuição no caso particular de um único veículo, de capacidade ilimitada, estar estacionado em cada depósito. O objectivo é definir uma rota para cada um dos veículos de forma a que cada cliente seja visitado uma única vez e a distância total seja mínima.

Apresenta-se uma forma de determinar minorantes das distâncias correspondentes às soluções óptimas do problema que pode ser interpretada como uma generalização, para o caso de múltiplos depósitos, do método, habitualmente designado por 1-tree, para a obtenção de minorantes para o problema do caixeiro viajante. Descreve-se um algoritmo para a afectação de clientes a depósitos, de complexidade $O(n^2 + k^2n)$, em que n e k representam respectivamente o número de clientes e depósitos e prova-se que, sendo válida a desigualdade triangular, a partir da afectação obtida é possível determinar uma solução admissível a que corresponde uma distância total não superior a $2(n - k + 1) / (n - k + 2)$ vezes a distância da solução óptima. Por último, apresentam-se alguns resultados computacionais com vista a avaliar a qualidade dos minorantes e majorantes propostos.

Keywords

Lagrangean Relaxation, approximation algorithms, traveling salesman problem, trees, matchings.

1. Introduction

The multi-depot traveling salesman problem (TSP) is the special case of the multi-depot vehicle routing problem where there is only one vehicle with unlimited capacity stationed in each depot. The objective is to design a route for each vehicle so as to minimize the total distance traveled, ensuring that each customer is visited exactly once.

Despite the fact of being NP-hard, even when restricted to any fixed number of depots, a well-solved case of the problem arises when the number of depots is equal to the number of

customers. In this case the optimal solution may be found by duplicating the edges of the minimum cost perfect matching on a bipartite graph. The edges of that graph representing connections of depots to customers, and each edge having a cost equal to the distance between the corresponding depot and customer.

In this paper we describe a way for obtaining lower and upper bounds on the distance of the optimal solutions of the multi-depot TSP. The lower bounds result from a relaxation of the problem that we call the k -tree problem. This relaxation generalizes, to the case of k depots, the successfully used 1-tree relaxation of the TSP [2]. The upper bounds are obtained by solving the TSP in each set of a feasible partition of the set of k depots and n customers. We show that a certain partition, defined by an $O(n^2 + k^2 n)$ algorithm, allows us to determine feasible solutions of the multi-depot TSP whose total distances are less than or equal to $(1 + \frac{n-k}{n-k+2})$ times the distance of the optimal solutions, whenever the triangle inequality holds. We also present some computational results in order to evaluate the quality of these bounds.

2. Lower Bounds

Let $G = (V, E, d)$ be a complete undirected weighted graph describing each instance of the multi-depot TSP. Nodes 1 up to k are used to represent the depots, and the remaining nodes of the node set $V = \{1, 2, \dots, k + n\}$ denoting the customers. E is the edge set, and associated to each e of E , we have the distance d_e of traveling directly between the two locations represented by the extreme nodes of edge e . If an edge e corresponds to locations that are not directly connected, we assigned to d_e the value $+\infty$.

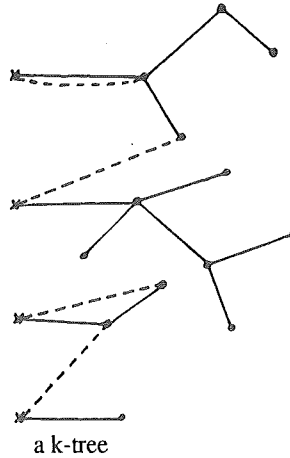
The multi-depot TSP consists of determining k simple cycles in graph G such that each node is included in exactly one cycle, no two nodes from set $\{1, 2, \dots, k\}$ appear in the same cycle, and the sum of distances of all edges in those cycles is minimum.

A concept that has been successfully used in determining lower bounds to the symmetric TSP, special case of the multi-depot TSP where the number of depots k is equal to one, is the concept of 1-tree introduced by Held and Karp [6]. A 1-tree in graph G is a spanning tree of G where node 1 has degree equal to one, plus an extra edge incident in this node. Seeking to generalize this concept so as to obtain lower bounds to the multi-depot TSP for values of k greater than one, we have to take into account that the duplication of edges connecting nodes in set $\{1, 2, \dots, k\}$ to nodes in set $\{k + 1, k + 2, \dots, k + n\}$ is now allowed. Thus, we will define a k -tree in graph G as being a multi-graph $T_k = (V, C_k \cup M_k)$, where the subsets of edges C_k and M_k of E satisfy the following conditions:

- (i) C_k is a cover of the nodes in set V by k node-disjoint trees;
- (ii) each node of set $\{1, 2, \dots, k\}$ is covered by exactly one edge of C_k ;
- (iii) with the edges of C_k there is no path connecting any two nodes in set $\{1, 2, \dots, k\}$;

- (iv) M_k is a matching of cardinality k , where each edge has one extreme node in set $\{1, 2, \dots, k\}$ and the other in set $\{k + 1, k + 2, \dots, k + n\}$.

In figure 1 is represented a k -tree in a 17 node graph where the value of k is equal to 4.



x nodes 1 up to k — edges of C_k
 • nodes $k+1$ up to $k+n$ --- edges of M_k

Fig. 1

Notice that each feasible solution of the multi-depot TSP in graph G may be represented by a k -tree. Moreover, if we want a k -tree T_k in graph G having the configuration of a feasible solution of multi-depot TSP, we only have to impose, in addition to (i) – (iv), the two following conditions:

- (v) T_k has exactly two edges incident in each node of the set $\{k+1, k+2, \dots, k+n\}$;
- (vi) each connected component of T_k includes no more than one node from the set $\{1, 2, \dots, k\}$;

Let us denote by T_k the set of k -trees in graph G , and by \hat{T}_k the subset of k -trees satisfying the condition (vi). We will also use $X(T_k)$ ($X(\hat{T}_k)$) to denote the set of incidence vectors of k -trees (k -trees that verify the condition (vi)). Thus, the multi-depot TSP may be formulated as follows:

$$\text{Minimize}_{x \in X(T_k)} \sum_{j \in E} d_e x_e \tag{2.0}$$

$$\text{subject to} \quad \sum_{j \in E_i} x_e = 2 \quad i = k + 1, \dots, k + n \tag{2.1}$$

$$x \in X(\hat{T}_k) \tag{2.2}$$

where E_i denotes the set of edges incident in node i .

Consider now the following relaxation of this problem obtained by ignoring the condition (2.2), and taking the constraints (2.1) into the objective function in a Lagrangean fashion:

$$L(\lambda) = \min_{x \in X(T_k)} \sum_{e \in E} (d_e - \lambda_{ie} - \lambda_{je}) x_e + 2 \sum_{i=1}^{k+n} \lambda_i \quad (2.3)$$

where i_e, j_e are the extreme nodes of edge e , and $\lambda = (\lambda_i)_{i=1, \dots, k+n}$ is the vector of multipliers associated with the constraints (2.1), with the convention that $\lambda_i = 0$, for $i = 1, \dots, k$.

It is well known [5, 3] that for each value of λ , $L(\lambda)$ given by expression (2.3), is a lower bound on the distance of the optimal solution of the multi-depot TSP. Problem (2.3) consists of determining the minimum cost k -tree in graph G with respect to a symmetric $(k+n) \times (k+n)$ cost matrix $[\bar{d}_{ij}]$, where $\bar{d}_{ij} = d_{[i,j]} - \lambda_i - \lambda_j$. The minimum cost k -tree in graph G may be obtained by solving the two independent problems:

- (a) to determine the minimum cost set of edges C_k that satisfies the conditions (i) – (iii);
- (b) to determine the minimum cost matching M_k that verifies the condition (iv).

There is an $O(n^2 + k^2 n)$ time algorithm that solves the relaxed version of problem (a) obtained by ignoring the conditions (ii) [1]. Forcing the conditions (ii) to be satisfied may be simply achieved by adding to all edges with one extreme node in set $\{1, 2, \dots, k\}$ the same value great enough to ensure that in the optimal solution no more than one edge is incident in each node of set $\{1, 2, \dots, k\}$. Thus, with this modification on the cost matrix, the above mentioned algorithm can also be used to solve problem (a). Problem (b) can be solved by an $O(k^2 n)$ time matching algorithm for weighted bipartite graphs (see for example [9]). Therefore, we have an $O(n^2 + k^2 n)$ time algorithm to obtain, for any value of λ , the lower bound $L(\lambda)$ on the distance of the optimal multi-depot TSP solution. Obviously, the best of the bounds $L(\lambda)$ given by expression (2.3) will be

$$L(\hat{\lambda}) = \max_{\{\lambda\}} L(\lambda) \quad (2.4)$$

The value $L(\hat{\lambda})$ may be determined by the subgradient method [7], which is a procedure often used to solve the Lagrangean dual of combinatorial optimization problems.

3. Upper Bounds

Let C_k^* be the subset of edges of graph $G = (V, E, d)$ such that

$$d(C_k^*) = \min_{\{C_k\}} \sum_{e \in C_k} d_e x_e \quad (3.0)$$

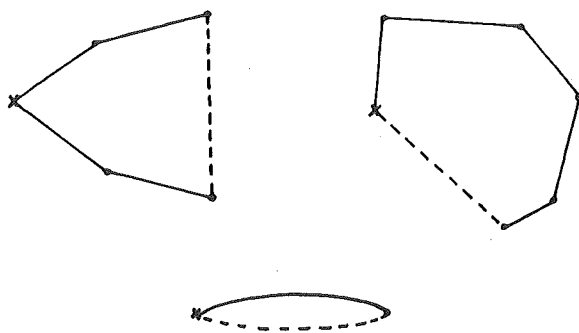
$$\text{subject to: (i), (iii)} \quad (3.1)$$

C_k^* induces a feasible assignment of the n customers to the k depots. Let us denote by $S_1^*, S_2^*, \dots, S_k^*$ the partition of the node set V associated to that assignment. Thus, each set S_i^* , $1 \leq i \leq k$, includes exactly one node from the set $\{1, 2, \dots, k\}$. Let now A be an algorithm for the TSP with time complexity $\xi_A(v)$, where v is the number of nodes of a complete graph. Let us assume that if $v = 2$ the algorithm A returns the cycle consisting of two copies of the unique edge of the graph. If we let algorithm A run k times, having each time as input the subgraph of G induced by the set of nodes S_i^* , $i = 1, \dots, k$, we will obtain a feasible solution F of the multi-depot TSP in graph G and therefore an upper bound on the distance of the optimal solution 0 . Since problem (3.0), (3.1) may be solved in $O(n^2 + k^2 n)$ time [1], $O(n^2 + k^2 n) + \sum_{i=1}^k \xi_A(|S_i^*|)$ is the time complexity of the above described algorithm to solve the multi-depot TSP.

We cannot be very optimist when evaluating the quality of the solutions obtained this way. In fact the negative result of Sahni and Gonzalez [10], stating that the existence of a polynomial time α -approximate algorithm to the TSP would imply that $P = NP$ can be easily generalized, for any given value of k , to the multi-depot TSP. However it can be proved that if the algorithm A has the property of determining cycles whose distances are not greater than twice the distances of the minimum spanning trees, then the following inequality holds

$$d(F) \leq \left(1 + \frac{n - k}{n - k + 2}\right) d(0) \tag{3.2}$$

Notice that deleting in each cycle i , $i = 1, \dots, k$, of any solution S of the multi-depot TSP the greatest distance edge Y_k^i in this cycle, we obtain a feasible solution C_k of problem (3.0), (3.1). Let us denote by Y_k the set of all deleted edges and by C_k^i the set of the remaining edges in cycle i , $i = 1, \dots, k$ (see figure 2).



a multi-depot TSP solution

- x nodes 1 up to k — edges of C_k
- nodes $k+1$ up to $k+n$ --- edges of M_k

Fig. 2

Since Y_k^i is the greatest distance edge among the edges in cycle i , we have the following inequalities:

$$d(Y_k^i) \geq \frac{1}{|C_k^i|} d(C_k^i) \quad i = 1, \dots, k$$

Adding these k inequalities, and taking into account that $|C_k^i| \leq n-k+1$ ($i = 1, \dots, k$) we have that

$$d(Y_k) \geq \frac{1}{n-k+1} d(C_k),$$

and therefore

$$d(S) \geq \left(1 + \frac{1}{n-k+1}\right) d(C_k),$$

defining a lower bound on the distance of the solutions of the multi-depot TSP in terms of a certain feasible solution of problem (3.0), (3.1). Given that C_k^* is an optimal solution of problem (3.0), (3.1) it is now easy to conclude from the above inequality that

$$2 d(C_k^*) \leq \left(1 + \frac{n-k}{n-k+2}\right) d(S) \quad (3.3)$$

Assuming that the algorithm A produces cycles whose distances are not greater than twice the distances of the minimum spanning trees, and given that C_k^* is the union of k minimum spanning trees each on the graph induced by S_i^* ($i = 1, \dots, k$), we finally obtain from (3.3) the inequality (3.2).

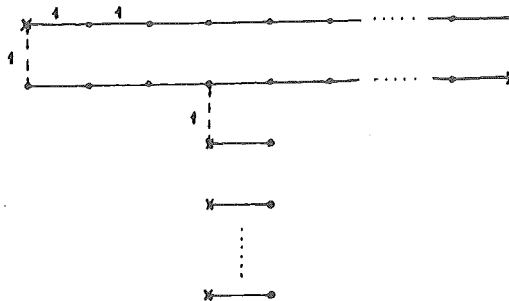
The existence of algorithms to the TSP validating the assumption imposed on algorithm A that allow us to establish expression (3.2), is ensured if we restrict the set of instances of the multi-depot TSP to those with distance matrices satisfying the triangle inequality [8]. Therefore, for any choice of an algorithm A, determining Hamiltonian cycles whose distances are not greater than twice the distances of the minimum spanning trees, the above described method is a $\frac{n-k}{n-k+2}$ -approximation algorithm for the triangle inequality multi-depot TSP. Notice also that it is possible to choose such an algorithm in order that the time complexity of the presented method does not exceed $O(n^2 + k^2 n)$, the time complexity of the algorithm used to define the clusters $S_1^*, S_2^*, \dots, S_k^*$ [8].

It is possible to show that for any given value of $k > 1$ and independently of the algorithm used to solve the TSP, there are instances in which the presented worst-case bound is almost achieved. Consider the Euclidean instance of the multi-depot TSP represented in figure 3. Each set S_i^* , $i = 1, \dots, k$ of the node set partition associated to the optimal solution C_k^* of problem (3.0), (3.1) is the set of nodes on the figure connected by continuous lines. The feasible solution F of the multi-depot TSP obtained by solving, using an exact method, the TSP in the subgraphs induced by the sets S_i^* , $i = 1, \dots, k$, has a total distance given by

$$d(F) = 2n.$$

Yet the optimal solution 0, which is obtained by constructing two node-disjoint rectangles using the nodes of the two first rows in the figure, and duplicating all the remaining continuous edges, is such that

$$d(0) = n + k.$$



an Euclidean instance of the multi-depot TSP

- x nodes 1 up to k
- nodes k+1 up to k+n

Fig. 3

Notice that, for any given value of $k \geq 2$, as n goes to infinity the ratio $d(F) / d(0)$ and the worst-case bound $1+(n-k) / (n-k+2)$ have the same limite value of 2.

Therefore, for any value of $k > 1$ and independently of the algorithm used to solve the TSP, the presented worst-case bound cannot be significantly improved.

4. Computational Results

In the computational tests realized, we used graphs with 50, 60, 70, 80, 90 and 100 nodes, In each of the six cases, ten different distance matrices satisfying the triangle inequality were considered. Each of these matrices results from applying Floyd's algorithm [4] for finding the shortest paths between all pairs of nodes to a symmetric matrix with integer entries randomly selected in the range [0, 100]. For each graph we consider the three problems emerging by letting the number of depots k be equal to $\lceil n/15 \rceil$, $\lceil n/10 \rceil$ and $\lceil n/5 \rceil$.

The lower bounds LB on the distance of the optimal solutions were obtained according to the method described in section 2, at the end of $2(k+n)$ iterations of the subgradient method [7].

The upper bounds UB were determined by the method presented in section 3. The algorithm A used to solve the TSP works as follows: first a Hamiltonian cycle is determined by the so-called minimum spanning tree algorithm [8]. Next, the 3-optimal method appearing in [11, pg. 377-379] is used to improve that cycle. During the 15 first iterations of the subgradient procedure we also applied the same algorithm on the subgraphs induced by each set of the node set partition defined in each iteration. This way we obtained several alternative feasible

solutions to each multi-depot TSP considered. We denote by SUB the distance of the smallest distance solution among the solutions thus determined.

The main results of our computational experiments are shown in figure 4.

k+n	k	aver ($\frac{UB}{LB}$)	worst ($\frac{UB}{LB}$)	aver ($\frac{SUB}{LB}$)	worst ($\frac{SUB}{LB}$)
50	3	1.1010	1.2043	1.0201	1.0604
	5	1.0798	1.1739	1.0194	1.0608
	10	1.0784	1.1597	1.0313	1.0745
60	4	1.0967	1.1438	1.0140	1.0222
	6	1.1092	1.1484	1.0158	1.0279
	12	1.0853	1.1428	1.0289	1.0569
70	4	1.0829	1.1181	1.0212	1.0444
	7	1.1067	1.1562	1.0273	1.0574
	14	1.0884	1.1190	1.0387	1.0672
80	5	1.1219	1.1597	1.0352	1.0642
	8	1.1132	1.1818	1.0373	1.0609
	16	1.1025	1.1538	1.0424	1.0625
50	6	1.1263	1.1885	1.0481	1.1078
	9	1.1342	1.1718	1.0521	1.1000
	18	1.1123	1.1583	1.0470	1.0666
100	6	1.1204	1.1636	1.0434	1.0727
	10	1.1305	1.1727	1.0453	1.0776
	20	1.0848	1.1221	1.0372	1.0542

Fig. 4

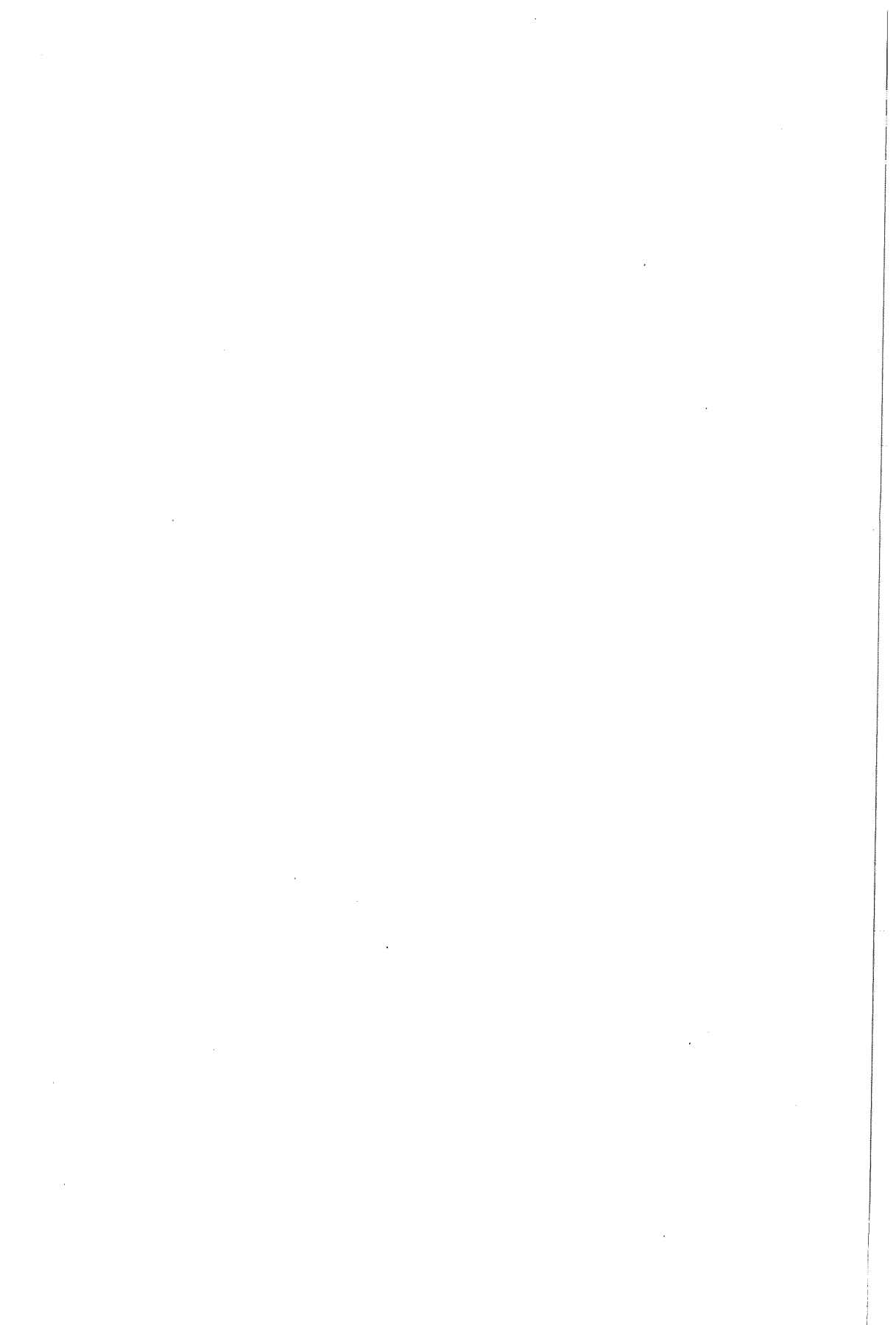
For the ten problems having the same node set and the same value of k, we report in columns 3 and 5 the averages of the ratios UB/LB and SUB/LB, which are upper bounds on the values of UB/d(0) and d(0)/LB, respectively. The greatest values of the ratios UB/LB and SUB/LB among these ten problems are displayed in columns 4 and 6.

Acknowledgement

The author wishes to thank to Prof. Ulrich Faigle for reading and commenting a preliminary version of this paper.

References

- [1] S.Baas and J.O.Cerdeira, Covering Nodes by K Node-Disjoint Trees, with Additional Cosntraints, Memorandum n° 693, Faculteit der Toegespaste Wiskunde, Universiteit Twente, 1978.
- [2] E.Balas and P.Toth, Branch and Bound Methods, in the Traveling Salesman Problem, ed. E.L.Lawler, J.K.Lenstra, A.H.G.Rinnooy Kan, D.B.Shmoys, John Wiley & Sons, (1985), 361-401.
- [3] M.L.Fisher, The Lagrangian Relaxation Method for Solving Integer Programming Problems, Manag.Sci. 27 (1981) 1-18.
- [4] R.W.Floyd, Algorithm 97: Shortest Path, Comm. ACM 5 (1962) 345.
- [5] A.M.Geoffrion, Lagrangian Relaxations and Its Uses in Linear Programming, Mah. Program. Study 2 (1974) 82-114.
- [6] M.Held and R.M.Karp, The Traveling-Salesman Problem and Minimum Spanning Trees, Oper. Res. 18 (1970) 1138-1162.
- [7] M.Held, P.Wolfe and H.P.Crowder, Validation of Subgradient Optimization, Math. Program. 6 (1974) 62-88.
- [8] D.S.Johnson and C.H.Papadimitriou, Performance Guarantees for Heuristics, in The Traveling Salesman Problem, Ed. E.L.Lawler, J.K.Lenstra, A.H.G.Rinnooy Kan, J.B.Shmoys, John Wiley & Sons, (1985) 145-180.
- [9] E.L.Lawler, Combinational Optimization: Networks and Matroids, Holt Rinehart and Winston, New York, 1976.
- [10] S.Sahni and T.Gonzalez, P-Complete Approximation Problems, J. ACM. 23 (1976) 555-565.
- [11] M.M. Syslo, N. Deo and J.S. Kowalik, Discrete Optimization Algorithms with Pascal Programs, Prentice-Hall Inc., Englewood Cliffs, N.J., 1983.



PROBLEMAS DE EMPACOTAMENTO TRIDIMENSIONAL

Maria Helena Correia
António Miguel Gomes
INESC Porto

José Fernando Oliveira
José Soeiro Ferreira
INESC Porto
DEEC, Faculdade de Engenharia
Universidade do porto

Abstract

The three dimensional packing problem consists, basically, in finding efficient packing patterns so that boxes, with known dimensions and quantities, can be packed inside bigger empty spaces. The main and most usual objective is the minimisation of the unused space. This problem is usually found in the context of loading freight containers (or trucks) in which an efficient space utilisation must be achieved.

This paper presents a short review of George & Robinson (1980) heuristic, one of the few known technics concerning the 3D packing problem, together with a study about a parameter of the algorithm, which has shown up as fundamental for the heuristic global performance. It's also described an implementation of a heuristic based on the George & Robinson approach which easily allows the inclusion of practical constraints such as box orientation, weight limitation and loading stability. Finally the solution of a real case is presented and discussed.

Resumo

Nos problemas de empacotamento tridimensionais (3D) é comum o objectivo de determinar padrões eficientes de utilização de objectos grandes por objectos menores, de dimensões e quantidades conhecidas. Frequentemente, e por outras palavras, pretende-se saber como colocar caixas (objectos pequenos) no interior de contentores ou de camiões de forma a obter, nomeadamente, um bom aproveitamento do espaço.

Neste artigo é descrita uma implementação para resolução de problemas de empacotamento tridimensionais baseada na heurística de George & Robinson (1980), uma das poucas técnicas conhecidas para o problema de empacotamento 3D, bem como ao estudo efectuado sobre um parâmetro fundamental no algoritmo, e que é determinante na qualidade do seu desempenho. Por outro lado, descreve-se uma abordagem a este problema que contém uma estrutura de dados permitindo a inclusão de um conjunto de restrições como a orientação de caixas, limitação de peso e equilíbrio de carga, as quais necessitam de ser satisfeitas para que as soluções obtidas possam ser aplicadas a casos reais. Finalmente é apresentada a resolução de um problema que nos foi proposto por uma empresa de transitários.

Keywords: packing, container loading, boxes, heuristic.

Introdução

Com a resolução do problema de Corte e Empacotamento (Cutting & Packing) pretende-se proceder à afectação de vários itens a um espaço n -dimensional de forma a obter uma solução que proporcione um aproveitamento eficiente do espaço. Assim, para além de se determinar o que cortar/empacotar, pretende-se, também, seleccionar a disposição dos objectos a cortar/empacotar. É esta componente geométrica da solução que caracteriza este tipo de problemas, tornando-os numa classe particular dos problemas de optimização combinatoria.

Uma formulação simples e geral dos problemas de Corte e Empacotamento é muito difícil de conseguir pois os contextos em que estes podem surgir e as formas que podem assumir são muito variados. Visto tratar-se de um problema de natureza combinatoria, a sua complexidade irá depender, entre outros factores, do número de dimensões em causa (1D, 2D, 3D, ...) e do número de objectos a colocar.

Dentro desta classe de problemas, os mais simples de resolver são, como é óbvio, os de uma dimensão. Surgem quando se pretende cortar tubos, rolos ou barras, e sempre que apenas uma dimensão (comprimento) seja significativa. Quando se trata de cortar materiais planos, placas ou chapas, em figuras geométricas mais pequenas, regulares (rectangulares) ou irregulares, estamos perante um problema de duas dimensões. Os problemas de três dimensões surgem sobretudo quando se pretende empacotar um conjunto de peças num espaço maior de dimensões fixas. No entanto, com a normalização de uma ou mesmo duas dimensões consegue-se diminuir a dimensionalidade de problemas que aparentemente seriam tridimensionais, pelo que estes problemas ocorrem menos frequentemente.

Para além dos casos acima mencionados em que podemos facilmente reconhecer problemas de Corte e Empacotamento, outros há que, apesar dessa associação não ser tão evidente, podem ser formulados como tal. É o caso do projecto de circuitos VLSI, problemas económicos de aplicação de capitais, balanceamento de linhas de produção, e ainda da gestão de memória por blocos, alocação de memória e escalonamento de sistemas multiprocessados.

Qualquer que seja o número de dimensões envolvidas existe sempre um objectivo básico, que consiste em maximizar/minimizar o espaço ocupado/desperdiçado. Dependendo do tipo de problema poderão surgir outros objectivos e restrições que irão tornar cada caso num problema particular dentro da área de Corte e Empacotamento.

O problema do Empacotamento 3D

Empacotamento em Contentores

Sob a designação de problema de Empacotamento 3D (de peças tridimensionais) encontram-se todos os casos em que se pretende seleccionar a disposição de diferentes tipos de objectos rígidos, em várias quantidades, num espaço de dimensões fixas, obtendo uma solução que proporcione não só um aproveitamento volumétrico eficiente do espaço, mas também a satisfação de outros parâmetros que poderão condicionar a aplicabilidade da solução obtida.

Procede-se desta forma à afectação do espaço disponível a peças mais pequenas. Em geral, é inevitável o aparecimento de "buracos" no empacotamento. Este espaço residual, não ocupado, é considerado desperdício e a selecção do posicionamento das diferentes peças deve sempre considerar a sua minimização.

O objectivo (maximização do valor da carga, cálculo do número mínimo de contentores, ...), as restrições (fragilidade, orientação dos objectos, estabilidade do empacotamento, ...) e mesmo o tempo disponível para a obtenção da solução (encontrar a primeira solução admissível, encontrar a melhor solução possível, encontrar a melhor solução num determinado tempo, ...) são factores variáveis neste tipo de problemas. Devido a este facto a designação de problemas de empacotamento tridimensional não é suficiente para definir um caso concreto, podendo este assumir diversas formas, cada uma das quais requerendo variantes na abordagem utilizada para a sua resolução.

Os objectos a empacotar poderão apresentar as mais diversas formas - regulares (cúbica, cilíndrica, ...) ou irregulares. Uma simplificação do problema consistirá em considerar apenas objectos de forma paralelepédica, assumindo, para aqueles que não a apresentem, a sua envolvente segundo essa forma.

O objecto do nosso estudo é o empacotamento de carga com estas características, isto é, problemas em que tanto o contentor como as peças a empacotar têm forma paralelepédica, considerando apenas soluções em que as caixas sejam empacotadas paralelamente aos lados do contentor. Este problema é conhecido na literatura pela designação de Empacotamento em Contentores.

Apesar da simplificação considerada, o problema é ainda bastante complexo devido à variedade de formas que pode assumir. Na prática, soluções para este tipo de problemas poderão ser requeridas em várias situações, de entre as quais destacamos:

- quando, na presença de uma dada carga, se pretende determinar o número mínimo de contentores necessário para empacotar essa carga;
- quando, na presença do contentor e da carga a empacotar, se pretende:
 - i) encontrar uma solução viável para o empacotamento da totalidade da carga no contentor;
 - ii) encontrar uma solução que proporcione um aproveitamento eficiente do espaço no contentor, considerando que o espaço restante será posteriormente ocupado com outra carga;
 - iii) empacotar obrigatoriamente uma parte da carga, sendo o empacotamento completado com o máximo da restante carga;
 - iv) obter um empacotamento que maximize o valor da carga empacotada.

Um empacotamento eficiente pode, ainda, ser determinado pela satisfação de outros parâmetros, alguns dos quais são bastante difíceis de explicitar matematicamente:

- limite e distribuição do peso no contentor;

- fragilidade da carga;
- ordem de descarga das caixas;
- estabilidade da carga;
- restrições na orientação das caixas a empacotar;
- distância mínima entre caixas que contenham produtos cuja reacção, em caso de fuga, poderá ter consequências indesejáveis;
- etc.

Métodos de resolução do problema de Empacotamento em Contentores

O problema do Empacotamento em Contentores, constitui um problema tipicamente combinatorio para o qual não são conhecidas soluções computacionais que garantam um resultado óptimo. A aplicação de um método baseado numa pesquisa exaustiva da solução óptima seria impraticável sempre que se tratasse de carregamentos com um número considerável de caixas. Este número varia com o caso considerado, podendo ir desde algumas dezenas até alguns milhares. Sendo o problema de tipo combinatorio, a aplicação deste tipo de métodos é limitada pelas capacidades de *hardware* e pelo tempo de processamento disponível para a sua resolução. Numa tentativa de ultrapassar estas dificuldades e conseguir técnicas de interesse prático, foram desenvolvidas várias heurísticas com as quais têm sido conseguidas soluções de boa qualidade.

Uma heurística destinada a resolver o problema do Empacotamento em Contentores, na sua generalidade, deverá considerar todas as formas sob as quais este problema pode surgir, as quais já foram descritas anteriormente. Na base de tal heurística deverá estar sempre um conjunto de regras que vise obter apenas um aproveitamento eficiente do espaço. Após ter sido desenvolvido este método básico, poderão ser feitas alterações tendo em vista a incorporação de restrições específicas de cada tipo de problema. É devido à importância que tem o método de empacotamento considerando apenas a minimização do espaço desperdiçado, que grande parte das heurísticas propostas considera unicamente este caso particular do problema de Empacotamento. Conseguindo um método que proporcione uma boa ocupação do espaço, teremos uma boa base para o desenvolvimento de sistemas de resolução de problemas mais complexos e que satisfaçam as restrições que surgem nos problemas reais.

Em contraste com o grande número de trabalhos versando o problema bidimensional, poucas publicações têm surgido sobre o Empacotamento tridimensional. Tal facto não deve ser interpretado como uma manifestação de desinteresse por estes problemas. Na origem desta situação estão factores como o grande interesse comercial de aplicações computacionais deste tipo, o que vem desencorajar a sua divulgação, e a possibilidade de resolução de problemas intrinsecamente tridimensionais pela aplicação de técnicas bidimensionais e mesmo unidimensionais através da normalização de uma ou duas dimensões, respectivamente.

Um procedimento comum às heurísticas propostas para a resolução do problema de Empacotamento tridimensional consiste na construção de camadas ao longo de uma das direcções do contentor. Estas podem ser obtidas pela construção de conjuntos de caixas de um mesmo tipo - pilhas, módulos em forma de L - ou ainda pela adaptação de técnicas já existentes para problemas bidimensionais. Como exemplo pode referir-se o caso da heurística proposta por Bischoff & Dowsland [1] em que, sendo o empacotamento feito por camadas de um mesmo tipo de caixas (caixas do mesmo tamanho), a selecção do posicionamento das caixas na camada é feita pela aplicação de uma técnica bidimensional - a abordagem de Bischoff & Dowsland para problemas de corte bidimensional. Também no trabalho de De Sha [3] é utilizada uma técnica bidimensional para determinar a colocação de pilhas de caixas na base do contentor.

A maior parte dos trabalhos publicados sobre empacotamento 3D ocupa-se do caso específico do Empacotamento em Contentores. Nalguns casos, pretende-se determinar uma forma de empacotar uma dada carga num contentor, outros tratam da minimização do número de contentores necessários para empacotar uma carga fixa.

Problema abordado

O problema aqui considerado é o do Empacotamento em Contentores. Dentro desta área, abordamos o problema particular do empacotamento de um contentor por uma carga. A solução obtida, deve proporcionar não só um aproveitamento eficiente do espaço mas também satisfazer um conjunto de restrições (orientação das caixas, ordem de descarga, limite e distribuição do peso no contentor) necessárias para que as soluções obtidas possam ser aplicadas a casos reais.

Em face dos objectivos atrás referidos e após a análise de vários métodos propostos na literatura para a resolução do problema, baseámo-nos numa heurística desenvolvida por George & Robinson [5]. Esta heurística, uma das poucas conhecidas para a resolução do problema, só considera a minimização do espaço desperdiçado. Apesar desta limitação, a qualidade dos resultados e a facilidade com que se podem incluir outras restrições do problema, justificam a escolha efectuada.

Descrição da heurística de George & Robinson

A abordagem proposta por George & Robinson, baseia-se no conceito de preenchimento do contentor por fases, construindo sucessivas camadas ao longo da profundidade do contentor. Uma camada, tal como a figura 1 ilustra, é uma secção transversal do contentor, com uma profundidade determinada pela primeira caixa a ser colocada nessa camada. A escolha da caixa inicial de uma camada é de crucial importância no desempenho do algoritmo, não só para obter camadas de profundidade adequada, nem muito estreitas nem muito profundas, mas também porque se tenta preencher a camada com colunas de caixas de um mesmo tipo, o inicial. Os restantes tipos de caixas só são considerados para preencher espaços dentro de uma camada ou então espaços resultantes da junção de espaços contíguos.

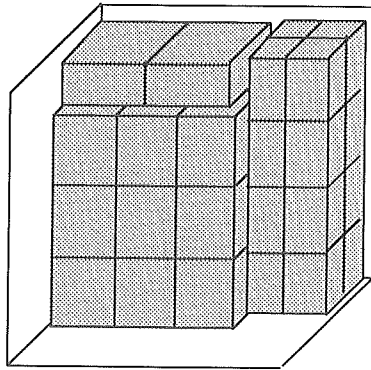


Fig. 1 - Construção de Camadas

O critério utilizado na escolha do tipo de caixa para iniciar uma camada, dá preferência aos tipos de caixas já utilizados. De entre estes é escolhido o tipo de caixa com o maior número de caixas por empacotar. A utilização deste critério contribui decisivamente para que caixas de um mesmo tipo tendam a ficar juntas, satisfazendo-se uma característica desejável em muitas situações reais.

Quando se pretende iniciar a construção de uma camada e não existem caixas de tipos já utilizados, situação que ocorre no início do empacotamento (e não só), escolhe-se para iniciar a camada o tipo de caixa que na altura estiver no início da lista ordenada de tipos de caixas. O critério utilizado na ordenação dos vários tipos de caixas privilegia os mais difíceis de empacotar (com o maior volume, com a maior das menores dimensões, com o maior número de caixas por empacotar, etc).

A profundidade da camada deverá ainda ser inferior a um parâmetro K , impedindo-se deste modo a formação de camadas demasiado profundas. Como se irá mostrar mais à frente, o critério de ordenação das caixas e o valor do parâmetro K têm uma grande influência na qualidade das soluções obtidas.

No preenchimento dos espaços vazios que aparecem no interior das camadas, escolhe-se o tipo de caixa que proporcione o melhor aproveitamento do espaço. Espaços onde não seja possível colocar qualquer tipo de caixa são rejeitados. Os espaços são criados (figura 2) após a colocação de um conjunto de caixas, por cortes ortogonais do espaço original segundo os planos que contêm as faces do conjunto de caixas aí colocado (cortes tipo guilhotina).

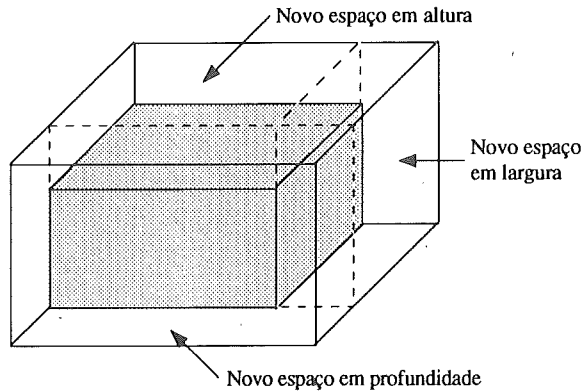


Fig. 2 - Criação de Espaços

Para finalizar esta descrição sumária da heurística de George & Robinson, falta referir o método que permite amalgamar espaços. Esta técnica contribui decisivamente para a eficiência da heurística e está ilustrada na figura 3.

A cada espaço está associado um parâmetro, "largura flexível", que é utilizado na limitação do número de caixas a colocar nesse espaço. Na figura 3, o número de colunas de caixas a colocar para iniciar a nova camada é limitado a três pela "largura flexível" (LF1) associada a este espaço: apenas uma coluna de caixas pode ultrapassar esta largura. A profundidade do espaço *bcef*, gerado pela colocação destas caixas, é aumentada pela amalgamação de espaço livre existente na camada anterior, passando a considerar-se para empacotamento o espaço *acdf*. Da mesma forma, este espaço terá associado uma "largura flexível" (LF2). Esta técnica de preenchimento permite obter empacotamentos compactos e de elevado aproveitamento do espaço disponível.

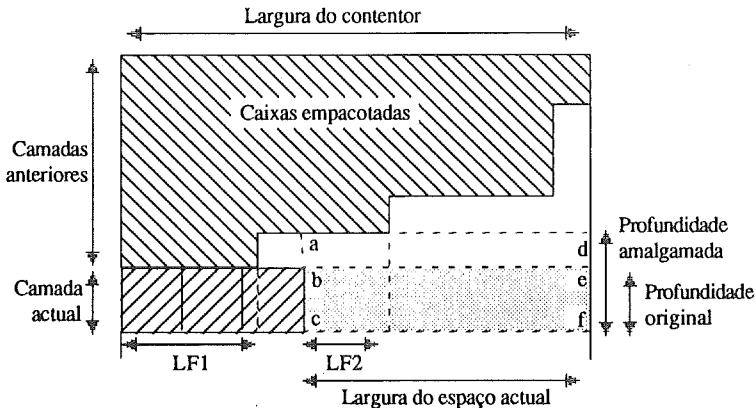


Fig. 3 - Amalgamação de Espaços

A amalgamação de espaços exige que a superfície frontal do empacotamento tenha um perfil em escada. Um perfil deste tipo é conseguido graças aos métodos de construção de camadas e de criação e preenchimento de espaços usados por esta heurística.

Teste e avaliação da heurística de George & Robinson

Para avaliarmos a influência do critério de ordenação na qualidade dos resultados, implementamos três versões da heurística de George & Robinson, utilizando em cada uma delas um critério diferente para a ordenação dos tipos de caixas:

- ordem decrescente da menor dimensão das caixas;
- ordem decrescente do número de caixas por empacotar;
- ordem decrescente do volume das caixas.

Testamos as três versões implementadas com um conjunto de cargas com determinadas características (número de tipos de caixas, número total de caixas e caixas com os seguintes formatos: cúbicas, achatadas, alongadas ou com uma das faces quadradas) e mantendo o volume total constante. A análise dos resultados mostra que o primeiro critério de ordenação conduz geralmente aos melhores resultados [2].

A descrição da heurística feita por George & Robinson [5] não refere nenhum critério para a escolha do parâmetro K . Para avaliar a influência deste parâmetro na qualidade da solução obtida utilizamos o conjunto de cargas atrás descrito. Para cada carga, foram atribuídos vários valores ao parâmetro K , relacionados com as dimensões da carga a empacotar: uma vez que a profundidade de uma nova camada é determinada pela maior dimensão, menor ou igual a K , da caixa escolhida, atribuiu-se a este parâmetro, em testes sucessivos, cada uma das maiores dimensões das caixas que constituem a carga a empacotar. Os resultados obtidos [2] realçam a influência do parâmetro K na qualidade das soluções geradas não permitindo, no entanto, encontrar qualquer relação causal entre este parâmetro e as características da carga a empacotar, isto é, perante uma dada carga a empacotar, não é possível determinar à partida qual o valor ideal para K .

Extensão da heurística de George & Robinson

A heurística proposta por George & Robinson, apresenta uma série de características específicas, não só devido ao facto de ser uma heurística do tipo *first fit*, mas também por recorrer a uma estratégia de criação e gestão de espaços livres bastante particular. Entre estas características são de realçar o escasso tempo de resolução de qualquer um dos problemas testados, a pouca exigência de memória, a influência do critério de ordenação dos tipos de caixas e uma grande dependência em relação ao valor do parâmetro K .

A análise do comportamento da heurística e a qualidade dos resultados obtidos justificam a validade das técnicas utilizadas para o tratamento dos espaços livres ao nível da criação, gestão, amalgamação e também preenchimento dos mesmos. Esta afirmação baseia-se nos níveis de

aproveitamento superiores a 90% que surgem durante praticamente todo o empacotamento. A única excepção ocorre durante o empacotamento das últimas camadas que apresentam um índice de aproveitamento nitidamente inferior, não só por se tratar duma heurística tipo *first fit*, mas também devido à necessidade de se recorrer à criação de uma nova camada para empacotar o(s) último(s) conjunto(s) de caixas, quando este(s) poderia(m) ser empacotado(s) num dos espaços livres existentes numa camada anterior, ou mesmo na actual, utilizando algum do espaço livre existente na parte frontal da mesma, quebrando o perfil em escada.

Em face destas observações, achamos que de momento não vale a pena modificar ou mesmo desenvolver a estratégia usada para o tratamento dos espaços porque, além de estar assegurado um bom aproveitamento volumétrico, consegue-se aliar a esta característica uma simplicidade que possibilita a inclusão das restrições atrás referidas. No entanto, os baixos níveis de aproveitamento das camadas finais justificaram a melhoria desta fase do empacotamento pela relaxação da exigência em manter a superfície frontal com um perfil em escada.

As características atrás mencionadas, podem ser aproveitadas para melhorar o desempenho da heurística. O escasso tempo necessário para a resolução de qualquer um dos problemas testados possibilita a utilização de mais algum tempo na procura de uma melhor solução. Este tempo deve ser preferencialmente usado na escolha da melhor profundidade para uma camada e, conseqüentemente, na selecção da primeira caixa a colocar nessa camada melhorando o procedimento da heurística original. Nesta, o parâmetro K condiciona a escolha da dimensão para profundidade da nova camada, e a escolha da caixa a colocar pode ser imposta pelo critério de ordenação dos tipos de caixas. A selecção da caixa para preencher um espaço livre, que não seja uma nova camada, é independente do critério de ordenação usado, sendo experimentadas todas as caixas e escolhida aquela que conduz a um melhor aproveitamento.

A eliminação da influência do parâmetro K e da dependência do empacotamento relativamente ao critério de ordenação das caixas, baseou-se num procedimento combinatório em que, na fase de selecção da caixa para iniciar uma nova camada, são construídas três camadas fictícias com cada um dos tipos de caixas disponíveis. Escolhe-se então para iniciar uma nova camada o tipo de caixa e a dimensão para profundidade que proporciona um melhor aproveitamento. Trata-se de uma optimização local que não conduzirá necessariamente a uma optimização global.

A heurística implementada com as extensões atrás referidas (quebra do perfil em escada e eliminação do parâmetro K e do critério de ordenação das caixas) foi testada com os mesmos problemas utilizados no teste das versões da heurística de George & Robinson, tornando-se assim possível avaliar a sua qualidade. A heurística modificada obteve os melhores resultados, com níveis de aproveitamento entre 85% e 90% contra os 80% a 85% conseguidos pelas versões da heurística original, o que já seria de esperar, visto basear-se num procedimento mais exaustivo aquando da escolha de caixas que iniciam novas camadas.

Inclusão de restrições

De acordo com os objectivos inicialmente propostos, falta considerar a inclusão de algumas restrições que surgem nos problemas reais. A inclusão destas restrições tornou-se possível graças a uma implementação flexível e modular e ao uso de uma estrutura de dados adequada.

Uma restrição óbvia, é o limite de peso suportado pelo contentor. Além de respeitar este valor máximo, é necessário garantir uma distribuição uniforme do peso. Outra situação usual neste tipo de problemas, é a imposição duma ordem de descarga das caixas em que se deve garantir que as cargas são empacotadas pela ordem inversa à que devem ser descarregadas. Uma última restrição incluída, foi a de garantir que para determinados tipos de caixas certas dimensões não possam ser usadas como altura. Esta restrição é importante na existência de líquidos e na presença de objectos frágeis. Foi ainda sentida a necessidade de considerar o empacotamento em espaços com forma diferente da paralelipédica, passando a ser contemplado o caso do empacotamento em caixas de camiões com rebaixado. Este tipo de camiões, geralmente de grande capacidade (tractor e reboque), não possui um perfil rectangular, apresentando um pequeno degrau na parte dianteira do reboque.

A inclusão da restrição relativa à distribuição do peso no contentor consistiu no cálculo de uma estimativa do peso ideal para uma camada considerando uma distribuição uniforme do peso ao longo do contentor. Porém, pode não ser possível, ou mesmo conveniente, do ponto de vista volumétrico, o empacotamento de camadas cujo peso coincida com este valor. A forma encontrada para garantir uma distribuição aproximadamente uniforme do peso consistiu em compensar o desvio relativamente ao valor ideal para o peso de uma camada por um desvio de sinal contrário na camada seguinte.

A garantia de uma ordem de descarga conveniente foi conseguida dividindo a carga e considerando o empacotamento de cada sub-carga separadamente e pela ordem desejada.

A consideração de restrições na orientação das caixas consistiu em condicionar a utilização de uma dimensão para altura pelo valor da *flag* correspondente a essa dimensão.

Associado a cada dimensão passível de ser usada em altura existe ainda um limite do peso suportado segundo essa dimensão. O número de elementos de um conjunto de caixas a empacotar encontra-se limitado, inicialmente, pelo peso que as caixas colocadas na base desse conjunto podem suportar e, posteriormente, pelo peso limite suportado pela superfície em que essas caixas vão ser colocadas.

Finalmente, a adaptação da implementação ao caso de camiões com rebaixado foi facilmente conseguida simulando a existência de um degrau no espaço paralelipédico envolvente do reboque.

Exemplo empresarial

Um dos principais objectivos que nos propusemos atingir com este trabalho, foi o de desenvolver um sistema capaz de proporcionar planos de empacotamento que possam ser aplicados a problemas reais. Seguidamente é apresentada a resolução de um problema que nos foi proposto por uma empresa de transitários.

A carga que está representada na tabela seguinte constitui a totalidade das caixas, que a referida empresa conseguiu empacotar no camião. Trata-se de uma carga com 11 tipos de caixas, num total de 381 caixas. Existem dois destinos diferentes para a carga, sendo necessário garantir que a carga para o destino A possa ser descarregada em primeiro lugar. Devido às características da carga não foram consideradas restrições nas orientações das caixas.

Estes valores mostram claramente que com a heurística implementada se obtiveram melhores resultados que os conseguidos manualmente pela empresa de transitários, sendo este facto comprovado pela fatia desocupada no fim do contentor, que permitiria o empacotamento de mais algumas caixas se existissem e tal fosse necessário. É ainda de realçar o reduzido tempo necessário para a obtenção do empacotamento.

Este sistema foi desenvolvido num PC 386SX a 16MHz. Para os testes realizados, o tempo de processamento nunca excedeu os 5 minutos e o limite imposto pelo *hardware* (memória) começa a fazer-se sentir a partir dos 50 tipos de caixas. Estas características do sistema são de primordial importância para as empresas de transitários, dado que frequentemente são postas perante alterações de última hora da carga a empacotar (no número e dimensões das caixas) permitindo-lhes, assim, dispôr de uma ferramenta que lhes possibilita alterar rapidamente e no próprio local do carregamento (utilizando um portátil, por exemplo) os planos de carga.

Camião: 6m × 2,4m × 2,6m - 7 ton.

Tipo	Quantidade	Dimensões (cm)	Peso (kg)	Destino
1	11	60 × 40 × 40	142	A
2	6	60 × 40 × 40	110	A
3	22	60 × 30 × 40	222	B
4	217	50 × 36 × 32	3625	B
5	45	60 × 30 × 40	495	B
6	6	60 × 40 × 40	90	B
7	2	60 × 40 × 30	24	B
8	15	71 × 40 × 38	225	B
9	15	73 × 35 × 33	170	B
10	22	60 × 40 × 40	305	B
11	20	64 × 35 × 52,5	368	B

Tabela: Carga a Empacotar

An alternative approach to solve the linear estimation problem is based on the Kalman methodology (see for example Harvey, 1990), that presents the advantages of its dynamical formulation and the facility of implementation.

The main objective of this paper is to apply the Kalman formulation on a DSPP for solving the problem of estimating the intensity process by supposing that $I(t)$ is a stationary random process factorizable as an exponential function and a zero-mean random variable ξ with variance α , that is

$$I(t) = \xi \cdot \exp\{-\beta(t - t_0)\} \tag{1}$$

Then, let us observe that $\mu(t) = E[I(t)] = 0$ and its covariance function is given by

$$R(t, s) = E[I(t) I(s)] = \alpha \cdot \exp\{-\beta |t - s| \} \tag{2}$$

Such a situation is usual in reliability problems where the intensity of a counting process is a random variable that follows an exponential law. Let us think, for example, that in a production process the intensity of failures decreases exponentially with the time in proportion as more controls are established.

An intensity process like (1) is also frequent in transmission and modulation theory, where a detector counts the arrival of particles of a random flow generated by a certain source. An application in this way with on-off modulated light is included in section 3.

2. Kalman Filtering for the Intensity Process

In order to develop the filtering equations, let us formulate a state-space model for the intensity process. From (1) that state equation will be:

$$\dot{I}(t) = -\beta I(t) + w(t)$$

and the measure equation will consist on the own observation of the intensity process along the time

$$y(t) = I(t) + v(t)$$

where $w(t)$ and $v(t)$ are two uncorrelated white noises. Then the differential equation for the linear estimate of $I(t)$ is

$$\dot{\hat{I}} = -\beta \hat{I}(t) + K(t) [y(t) - \hat{I}(t)] \tag{3}$$

where the Kalman gain $K(t)$ verifies

$$K(t) = 1 + \sum (t) \tag{4}$$

and the quadratic-mean error of estimation

$$\sum (t) = E\{ [I(t) - \hat{I}(t)]^2 \}$$

is given by

$$\dot{\sum} (t) = 1 - k^2(t) - 2\beta \sum (t) \tag{5}$$

By replacing (4) in (5) we obtain the following differential equation:

$$\dot{\sum} (t) = -\sum^2(t) - (2 + 2\beta) \sum (t) \tag{6}$$

with the initial condition $\sum (t_0) = \alpha$, whose solution is

$$\sum (t) = \frac{(2+2\beta)\alpha}{(\alpha+2+2\beta) \exp\{(2+2\beta) (t-t_0)\} - \alpha} \tag{7}$$

and the Kalman gain

$$K(t) = 1 + \frac{(2+2\beta)\alpha}{(\alpha+2+2\beta) \exp\{(2+2\beta) (t-t_0)\} - \alpha} \tag{8}$$

Obviously, at the initial time we have

$$\begin{aligned} \sum (t_0) &= \alpha \\ K(t_0) &= 1 + \alpha \end{aligned}$$

Therefore, equation (3) has as solution

$$\hat{I}(t) = \exp\{-\beta t - \int K(t) dt\} \int \exp\{\beta t + \int K(t) dt\} .K(t) y(t)dt \tag{9}$$

after looking for the integrant factor $\exp\{\beta t + \int K(t) dt\}$, where

$$\begin{aligned} \text{i.f} &= \exp\{\beta t + \int K(t) dt\} = \\ &= \exp\{(\beta+1)t\} [\alpha+2+2\beta - \alpha \exp\{(2+2\beta) (-t+t_0)\}] \end{aligned}$$

Expression (9) must be considered as a Lebesgue's integral of each observed sample function of the measured process $\{y(t)\}$. In fact, these continuous trajectories can be approximated or interpolated if are evaluated in discrete times.

If we impose the condition $\hat{I}(t_0) = I(t_0) = a$, the value of the constant that appears in integral (9) can be calculated. Expressions (7), (8) and (9) can be simplified for certain particular situations. Some of them are below.

a) When $t_0 = 0$

$$\begin{aligned} \sum (t) &= \frac{\alpha(2+2\beta)}{(\alpha+2+2\beta) \exp\{(2+2\beta)t\} - \alpha} \\ K(t) &= 1+ \frac{\alpha(2+2\beta)}{(\alpha+2+2\beta) \exp\{(2+2\beta)t\} - \alpha} \end{aligned}$$

b) If, for example, $\alpha = 3$, $\beta = 0.5$ and $t_0 = 0$

$$\begin{aligned} \sum (t) &= \frac{3}{2e^{3t} - 1} \\ K(t) &= 1+ \frac{3}{2e^{3t} - 1} \\ \text{i.f.} &= 3e^{1.5t} (2 - e^{-3t}) \end{aligned}$$

Let us observe that in all these situations we can derive an explicit formula for the estimated intensity from assumptions imposed to the Poisson process.

3. Application to Filtering on a Photoelectric Binary Source

Let us suppose that a source generates an electronic stream whose intensity is distributed as a random telegraphic wave for $t \geq t_0$, so that in the initial time t_0 it has the same probability of being in state 0 or A, and for $t > t_0$ the intensity alternates between such values in a sequence of

times that may be considered as the occurrence times of an homogeneous Poisson process $\{P(t), t \geq t_0\}$ with intensity λ . Then we can write

$$I(t) = \frac{A}{2} [1 + (-1)^{P(t)} z], \quad t \geq t_0$$

where z is a discrete random variable independent of the process $\{P(t), t \geq t_0\}$ taking values $+1$ and -1 with the same probability (see Papoulis, 1985 or Snyder & Miller, 1991).

The suitable stochastic model for the number of electrons received by a photo-counter is then a DSPP with intensity like (1), being now $\beta = 2\lambda$. On the other hand such an intensity does not have zero-mean but $\mu(t) = A/2$, and its covariance function is then

$$R(t, s) = (A/2)^2 \exp\{-2\lambda |t-s|\}$$

being $\text{Var}[\xi] = (A/2)^2$

For this particular situation expressions (7) and (8) can be written as follows:

$$\sum (t) = \frac{(2+4\lambda)A^2/4}{\left(\frac{A^2}{4} + 2+4\lambda\right) \exp\{(2+4\lambda)(t-t_0)\} - \frac{A^2}{4}}$$

$$K(t) = 1 + \frac{(2+4\lambda)A^2/4}{\left(\frac{A^2}{4} + 2+4\lambda\right) \exp\{(2+4\lambda)(t-t_0)\} - \frac{A^2}{4}}$$

Acknowledgment

The suggestions of an anonymous referee have been very useful for correcting and preparing the final version of this paper.

4. References

- [1] Harvey, A.C., Forecasting, Structural Time Series Model and the Kalman Filter, Cambridge University Press, 1990.
- [2] Jimenez, F., Contribución al Estudio de Procesos Poissonianos de Recuento con Intensidad Aleatoria, PhD Thesis, Universidad de Granada, 1992.
- [3] Papoulis, A., Probabilidad, Variables Aleatorias y Procesos Estocásticos, EUNIBAR, Barcelona, 1985.
- [4] Snyder, D.L. & Miller, M.I., Random Point Processes in Time and Space, Springer-Verlag, New York, 1991.
- [5] Van Trees, H.L., Detection, Estimation and Modulation Theory, Vol.I, Wiley, New York, 1986.

FUZZY FLOWS IN LINEAR NETWORKS WITH CORRECTION FOR SOME DEPENDENCIES IN NODAL UNCERTAINTIES

Vladimiro Miranda

Manuel A.C.C. Matos

J. Tomé Saraiva

INESC - Instituto de Engenharia de Sistemas e Computadores
and FEUP/DEEC - Faculdade de Engenharia da Universidade do Porto

Abstract

This paper proposes a new steady state fuzzy network flow modelling, in an unified linearized approach, to take in account qualitative aspects or vagueness and uncertainty in nodal data that have not probabilistic or random nature. Based on concepts from fuzzy set theory, approximate possibility distributions for branch flows and node potentials are then evaluated, starting from a previously calculated feasible operating point. The equations derived are identified with linear models for gas and electric networks. The formulation presented includes a corrective phase in order to calculate "exact" distributions taking in account additional dependences in the system equations, resulting from extra knowledge on the network operation.

Resumo

Neste artigo apresenta-se uma nova modelização do problema dos fluxos imprecisos em redes lineares, numa abordagem unificada que tem em conta os aspectos vagos ou qualitativos na especificação dos dados nodais resultantes de incertezas que não têm carácter aleatório ou probabilístico. Assente em conceitos da teoria dos conjuntos imprecisos, podem-se determinar distribuições de possibilidade aproximadas para fluxos em ramos ou potenciais em nós, a partir de um ponto de operação viável previamente calculado. As equações deduzidas são identificáveis com modelos lineares de redes de gás ou electricidade. A formulação apresentada inclui uma fase correctiva, a fim de permitir o cálculo de distribuições "exactas" sob condições suplementares de dependência entre grandezas, resultando de conhecimento adicional sobre o sistema físico ou seu modo de operação.

Keywords

Fuzzy sets, Network analysis, Linear flows, Fuzzy flows, Power systems, Gaz distribution.

Introduction

Computer simulations of steady state power flows in electrical systems or gaz flows in gaz networks aim at estimating the values of node voltages or pressures and branch flows for known consumptions at load nodes and, eventually, known source node voltages or pressures. The presentation of algorithms in the literature to solve this problem usually deals with it from a deterministic point of view. However, it is recognized that nodal data may have other nature,

When solving the DC model in the deterministic case, one usually adopts as the slack bus the same node for which the reference $Q = 0$ was fixed. This means that, in fact, one is actually solving a system of $(n-1)$ independent equations, and the injected power of the slack bus is calculated (based on the conservation principle) from the algebraic sum of all the other nodal injections.

In the fuzzy case, equation (6) means that the uncertainty of nodal injection in the slack bus depends on the uncertainties of the other nodal injections. But in terms of the definition of a fuzzy problem, this is not the most interesting case. The practical interesting case arises when one makes statements about uncertainties at all the nodes. This has an advantage: one can define the membership function of the injected powers so that no one will violate generation constraints.

The limiting case can be taken from the observation of the physical behavior of the system, in which the generation is driven by the consumption: this leads to the principle that **the total uncertainty in generation must not be inferior to the total uncertainty in consumption** (it means that for every possible scenario of loads there must be at least one scenario of generation that meet the demand), for uncorrelated loads.

One can therefore solve the fuzzy power flow problem in several ways:

- a) declare uncertainties at all nodes but the slack bus, and solve with (5) and (6).
- b) declare uncertainties at all nodes but the slack bus and admit that total uncertainty in generation equals total uncertainty in consumption: the fuzzy representation of the generation at the slack bus will be determined by the deconvolution of the fuzzy equation

$$\sum_{\text{gen.-slack}} \underline{\Delta L}_s + \underline{\Delta L}_{\text{slack}} = \sum_{\text{loads}} \underline{\Delta L}_1$$

- c) declare uncertainties at all nodes but the slack bus and declare for this node the max and min admissible generation values. This can be thought of as equivalent to describing by a rectangular membership function the feasible power injections at this node. Then, the fuzzy representation of the generation at the slack bus will be obtained by the intersection (with the min operator) of this fuzzy set with the one obtained with (6).
- d) declare uncertainties at all nodes including the slack bus.

Cases b, c and d cannot be solved directly by (5) and (6). The specification of the uncertainties at the slack bus acts as a new restriction in the space defined by the several $\underline{\Delta L}_i$ and $\underline{\Delta L}_{sj}$, eliminating from the set of feasible solutions some vertices defined (at each level α) by the extreme values of the $\underline{\Delta L}$. The formulation (4) and (5) does not take in account this fact: it therefore leads to larger branch flow and angle distributions than the real ones.

In order to evaluate the "exact" branch flow distributions in cases b, c and d, which can be considered as conceptually the same, one can admit that calculations are performed at each level

α , for the support intervals so defined. Equation (5), at each level α , would then be solved with the rules of interval arithmetic.

This equation also gives the coefficients of an objective function to determine the max and min values of the flow in any line, given the max and min limits of all the the injected powers (the limits on the slack bus will be affecting the algebraic sum of the injected powers at all the other nodes, because for every scenario total generation must meet total demand).

One faces therefore an optimization maximizing (minimizing) problem where a solution given by (5) may be feasible (and that's ok, it's the final result) or unfeasible, meaning that all constraints but one are satisfied, and this one is exactly the extra one due to the condition imposed to the slack bus.

A procedure must then be adopted to correct the extreme values of branch flows given by (5). It can be based on the technique of the dual simplex, but executed by inspection of the problem features:

- 1 - Verify if the power values calculated in (5) with the fuzzy arithmetic rules (to give the branch flow distributions) satisfy the additional constraint in the limits of injected power at the slack bus.
- 2 - If not, by inspection of the sign of the coefficients in $C T \{ Y_1 \}$ and their module, choose which value (or values) of ΔL to increase or decrease in order to satisfy the additional constraint. This means that the extreme value of a branch flow can be obtained at a solution where one of the generators or loads is not at any of its extreme possible values.

In real world problems, where the specified nodal uncertainties are likely to be of much smaller range than the feasible generation ranges, this procedure will usually be very efficient.

Fuzzy Gaz Flow

In general, the pipe flow equation in line i-j is described by

$$(q_{ij})^a = K_{ij}^{-1} (P_i - P_j), \text{ for } P_i > P_j,$$

where "a" depends on operating pressure and P_i, P_j are, in general, the square of actual node pressure values, usually in bar (Oziadacz, 1988). A linearization through the origin and q_0 is

$$q_{ij} = q_0 + (q_0)^{1-a} K_{ij}^{-1} [(P_i - P_j) - (P_{i0} - P_{j0})]$$

So, in this case, when building a system similar to (2), one will have $C_{ij} = (q_0)^{1-a} K_{ij}^{-1}$.

Fuzzy gaz flow equations are then (5) and (6), where flows q are identified with vector f and measured, for instance, in m^3/h . In this case, one has a number n_1 of source nodes, with fixed pressure values, and equation (6) becomes important, as it regulates the amount of gaz injection at the source nodes.

A special important case appears in gaz networks: the presence of a compressor station. A compressor type node must be split into two nodes, one of load type and one of source type such as in fig. 2.

The following conditions are imposed:

$$P_i = P \text{ specified; } L_j + L_k = 0$$

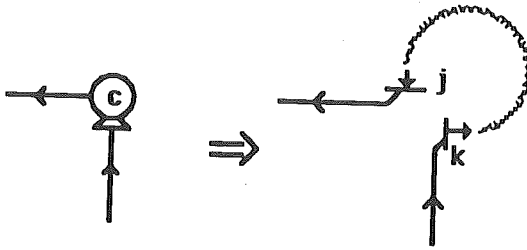


Fig. 2 – Representation of split compressor node

Equation (3) is partitioned into

$$\begin{bmatrix} Y_{ss} & Y_{s1} & Y_{s1a} \\ Y_{sas} & Y_{sa1} & Y_{sa1a} \\ Y_{1s} & Y_{11} & Y_{11a} \\ Y_{1as} & Y_{1a1} & Y_{1a1a} \end{bmatrix} \begin{bmatrix} 0 \\ \Delta P_1 \\ \Delta P_{1a} \end{bmatrix} = \begin{bmatrix} \Delta L_s \\ \Delta L_{sa} \\ \Delta L_1 \\ \Delta L_{1a} \end{bmatrix}$$

where matrices indexed "sa" correspond to compressor source type nodes and "1a" compressor load type nodes. Taking in account that $\Delta L_{sa} + \Delta L_{1a} = 0$, by adding lines we can get

$$\begin{bmatrix} Y_{ss} & Y_{s1} & Y_{s1a} \\ Y_{sas} & Y_{sa1} & Y_{sa1a} \\ Y_{1s} & Y_{11} & Y_{11a} \\ Y_{1as} + Y_{sas} & Y_{1a1} + Y_{sa1} & Y_{1a1a} + Y_{sa1a} \end{bmatrix} \begin{bmatrix} 0 \\ \Delta P_1 \\ \Delta P_{1a} \end{bmatrix} = \begin{bmatrix} \Delta L_s \\ \Delta L_{sa} \\ \Delta L_1 \\ 0 \end{bmatrix}$$

From this system one can derive expressions similar to (4), (5) and (6) that give

$$[\Delta P_1 | \Delta P_{1a}]^t, [f] \text{ (or } [q]) \text{ and } [\Delta L_s | \Delta L_{sa}]^t.$$

The need for a corrective phase in gaz networks is in practice maybe not as important as in electrical networks, because the pressure at the supply nodes is in general fixed. However, if at one of these nodes (k, for instance) one would not specify its pressure, but instead declare some uncertainty on the injected flow, a case similar to the one described above for electrical systems would occur.

The total uncertainty in the system, as a consequence of (6), may be described by the fuzzy equation

$$\sum_{\text{sources-k}} \underline{\Delta L}_s + \underline{\Delta L}_{\text{src.k}} = \sum_{\text{cons}} \underline{\Delta L}_1$$

When the uncertainty of source k is declared, the uncertainty at the other sources becomes determined by the deconvolution of the above equation; this limits the degrees of freedom of the possible occurring scenarios. One can then follow a similar kind of "dual simplex" strategy as

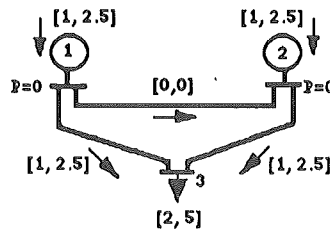
above, departing from the solutions given by (5) and (6) and correcting, at each α level, the membership function values.

4. Example

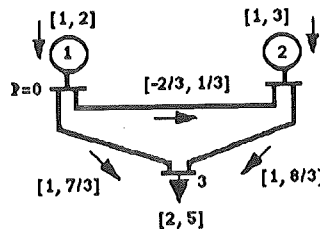
Numerical examples for gas and electricity networks can be found in (Miranda, 1991). In this section a small general illustrating examples is presented, based on a network with only three branches, two source nodes and one consumption node. Results are presented under graphic form, where:

- nodes 1 and 2 represent sources or injections, either of gas or electric power or any other product;
- node 3 represents a load or a consumption node, described uncertain in the interval $[2,5]$;
- branches between nodes correspond to pipes or electric power lines, and their coefficients have all numerical value 1.

Uncertainties will be described as rectangular membership functions, which allows also the reading of the results as an exercise on interval arithmetics. Arrows indicate the positive direction of flows. In the first exercise, the potential is fixed at nodes 1 and 2 (at 0, taken as a reference value). This is equivalent to the gas system problem. The results are as follows:

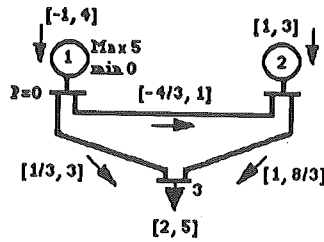


In the second exercise, the potential is only fixed at node 1 (at zero value), an uncertainty in supply is declared at node 2 as $[1,3]$ and the uncertainty in supply is assumed equal to the uncertainty in demand. This fixes the uncertainty of supply in node 1 as $[1,2]$. The calculations of branch flows need now the corrective phase and give as a result

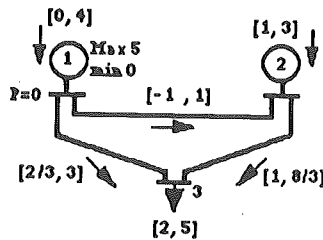


In the third exercise, the potential is fixed as zero at node 1, an uncertainty in supply is declared at node 2 as $[1,3]$ and the limits of supply capacity at node 1 are fixed at $[0,5]$, without

further assumptions. The calculation of supply uncertainty at this node without taking in consideration its capacity limits would give



As one can see, a negative supply in bus 1 would have to be considered, which is not admissible; therefore, one must take the supply uncertainty at this node as $[0, 4]$, from the intersection of the result obtained and the capacity constraint. The calculations are now resumed and include the need for the corrective phase.



One can notice that the flow in some branches is not affected by these successive steps, meaning that the uncertainty in their flows does not depend on the uncertainty of supply in some (more or less) distant nodes.

4. Conclusions

From the analysis of the examples above and the results obtained by the authors for other real size case studies, one may in general conclude that:

- The possibility distributions calculated cover in general a wider range of values than the interval delimited by the values obtained with the two deterministic load flow studies using maxima and minima loads and generated powers - this is a very important conclusion of practical interest for those involved in planning or system assessment and a new tool to readily evaluate flow possibilities is now at hand.
- This methodology reveals the possibility of power flow occurring in either direction at some lines (due to different combinations of loads and generated powers), which could remain undetected if only some deterministic scenarios were studied.
- Performing a fuzzy flow study requires a careful consideration of the case to be modelled, namely if all the recognized uncertainties have the same level of importance or are originated by the same level of ignorance. Clearly, in many problems supply is a consequence of demand (and therefore, one might say that the level of ignorance on the demand proceeds or is previous to the uncertainty on

supply). But in other cases (for instance, a wind generator in an electrical network or a power station whose control, by contract, is not in the hands of the main dispatcher), consumption and some particular generation have uncertainties of the same level of ignorance.

We have therefore presented a way of approximately deducing node potentials and branch flows from load data affected by imprecision, represented by fuzzy numbers. We have included a corrective phase to previous formulations to take in account several ways of defining the fuzzy problem, with interest for the engineering practice.

Furthermore, we have shown that the fuzzy flow formulation derived can easily be identified with linearizations of electrical or gaz systems, and by obvious reasons with other network systems as well.

However, one problem of interest for the electrical networks has not yet been addressed, which is the inclusion in the models of limits to the line flows. The knowledge that generation dispatch is conditioned by these limits will have an influence on the resulting solutions. If uncertainties have a "small range", compared to the crisp values of the previous deterministic study, the distributions obtained will be a good approximation of the actual ones. In any case, they will allow the identification of critical branches where overloads might occur.

There is a high practical interest in exploring the possibilities offered by the fuzzy flow model. One can foresee two areas of application. One is in system growth planning, because of the uncertainties associated to load definition and forecasting - it can have a general application in electric power systems, gaz networks, water supply, etc. In (Miranda, 1989a) one may find a proposal for the definition of fuzzy load curves adequate for planning purposes. The other area is more specific to electricity distribution networks, namely for operation or short range planning because, due to the lack of metering devices, loads are in many companies roughly estimated from data such as monthly energy consumptions or installed equipment capacity.

Acknowledgment

This paper reports works partially sponsored by JNICT under contract PMCT/C/TIT/413/90.

References

- [1] Zadeh, L., Fuzzy sets, *Information and Control*, 8 (1965) 338-353.
- [2] Zadeh, L., Fuzzy sets a basis for a theory of possibility, *Fuzzy Sets and Systems*, 1 (1978) 3-28.
- [3] Freska, C., Linguistic description of human judgements in expert systems and in the soft sciences, in *Approximate Reasoning in Decision Analysis* - M.M.Gupta & E.Sanchez ed., North Holland, 1982.
- [4] Osiadacz, A., Pienkosz, K., Methods of steady state simulation for gaz networks - *Int.J.Systems Sciences*, 19 (1988) 1311-1321.
- [5] Miranda, V. e Matos M., A Holistic Approach in Multicriteria Decision Aid, in *Improving Decision Making in Organizations*, A.G.Lockett Ed., Lecture Notes in Economics and Mathematical Sciences, Springer Verlag, Berlin, 1989.
- [6] Miranda, V., Matos, M. e Saraiva, J.T., Fuzzy Load Flow - New Algorithms Incorporating Uncertain Generation and Load Representation, in *Proceedings of the 10th PSCC*, Butterworths, London, 1990.
- [7] Miranda, V., Using fuzzy reliability indices in a decision aid environment for establishing interconnection and switching location policies, in *Proceedings of CIREN 91 (IEE/AIM) Liège, Belgium, 1991*.

- [8] Miranda, V., Fuzzy flows in gaz and electricity networks, in Proceedings of the European Simulation Multiconference - EMS'91, Copenhagen Denmark, Erik Mosekilde ed., Tech.Univ.Denm., Lyngby, 1991.

Appendix

Fuzzy set theory can be understood as an extension of n-valued logic when the number of the admissible logic values tends to infinity. For example, given an universe X and a subset X_1 of X , the membership value of an element x_1 to X_1 belongs to:

- $\{0, 1\}$ in the Boolean logic.
- $\{0, 0.5, 1\}$ in the trivalued Lukasiewicz's logic.
- $[0, 1]$ if normalized fuzzy sets are considere.

Therefore, a fuzzy set \tilde{A} is characterized by a membership function $\mu_{\tilde{A}}(x)$ associating an element x_1 to its compatibility degree to X_1 . In this sense, the transition between full and no membership is gradual rather than abrupt.

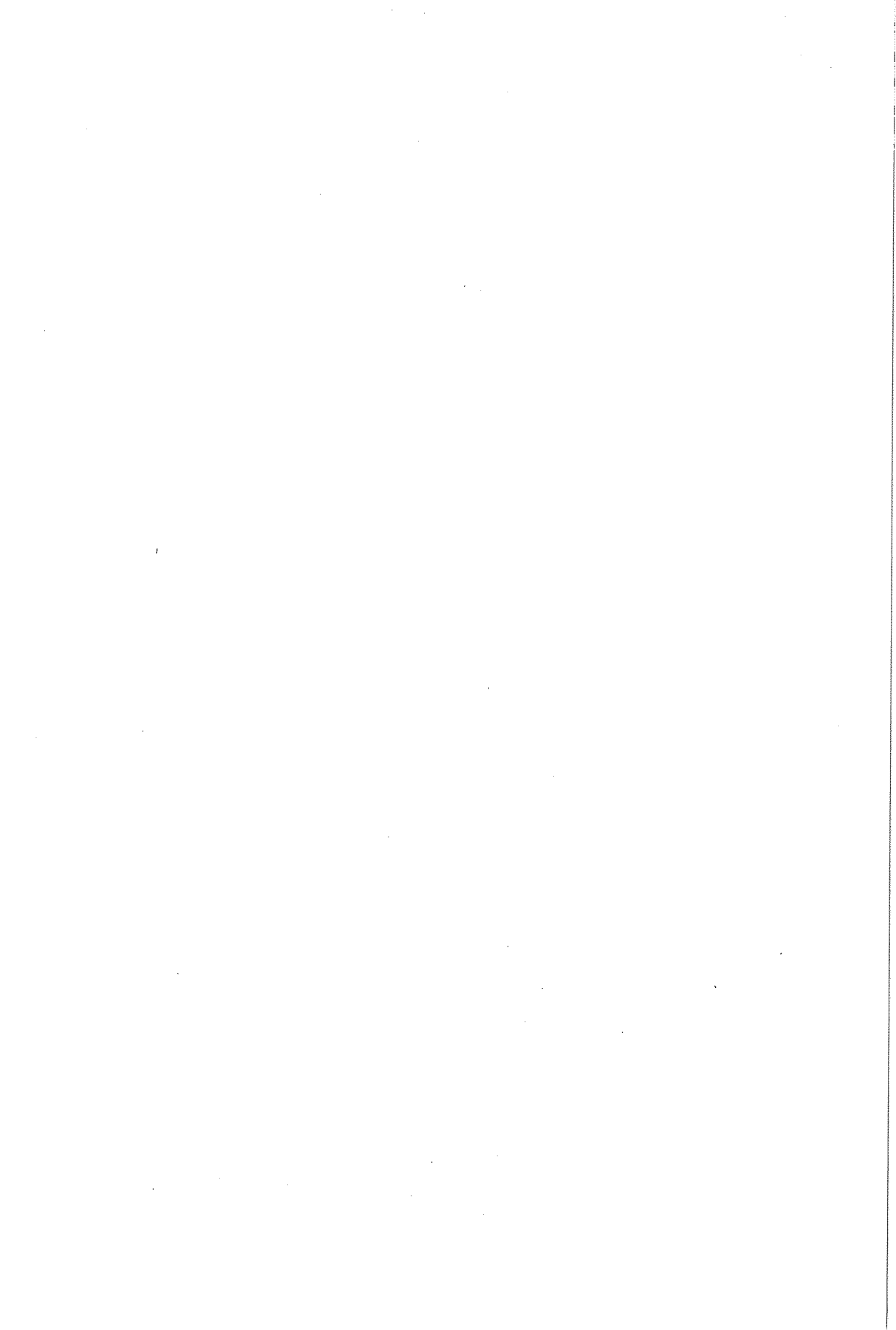
$$\tilde{A} = \{ (x_1, \mu_{\tilde{A}}(x_1)), x_1 \in X_1 \} \quad (A1)$$

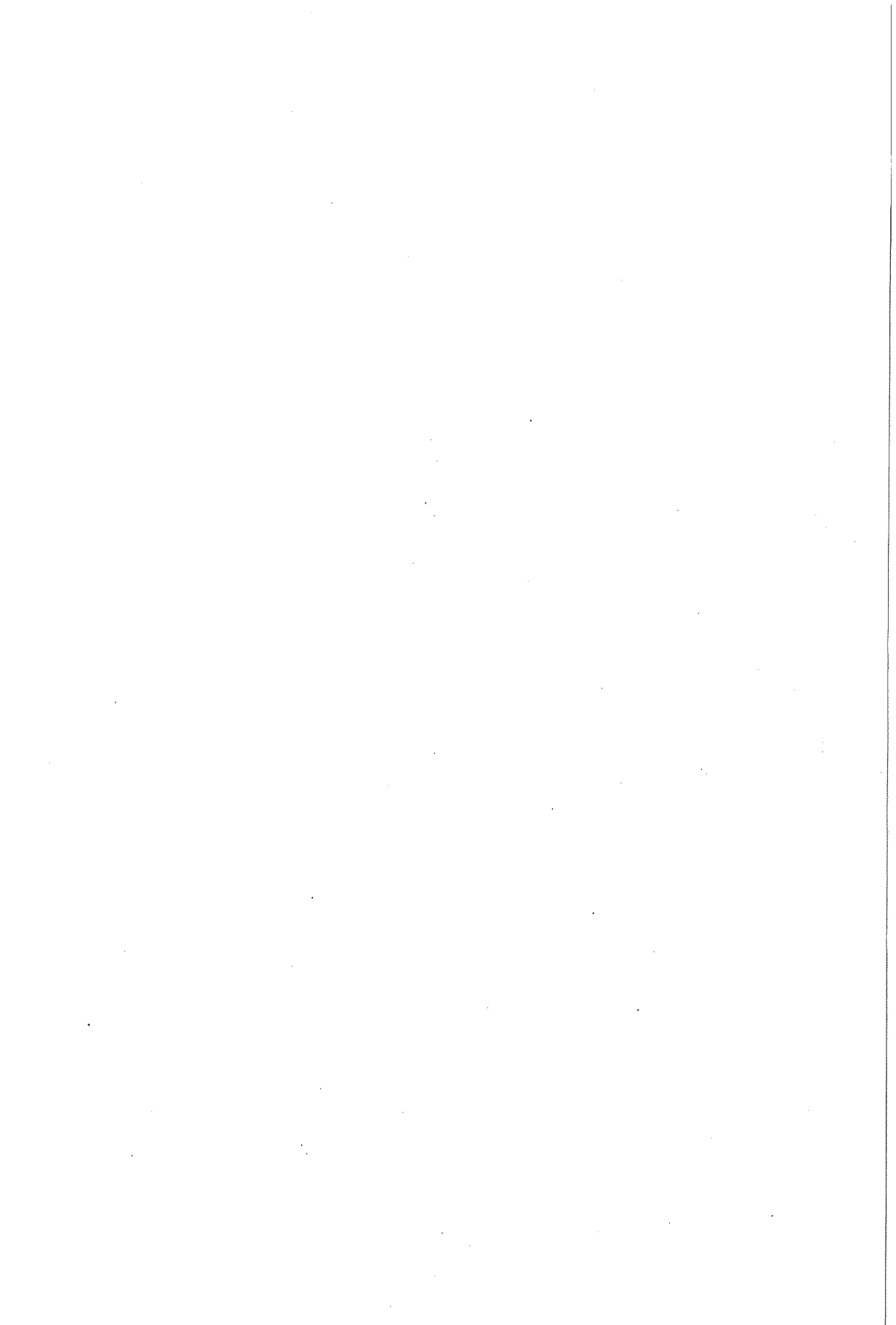
An α -cut of a fuzzy set \tilde{A} can be defined as the crisp set A_α of elements whose membership value to \tilde{A} is not inferior to α .

$$A_\alpha = \{ x_1 \in X_1 : \mu_{\tilde{A}}(x_1) \geq \alpha \} \quad (A2)$$

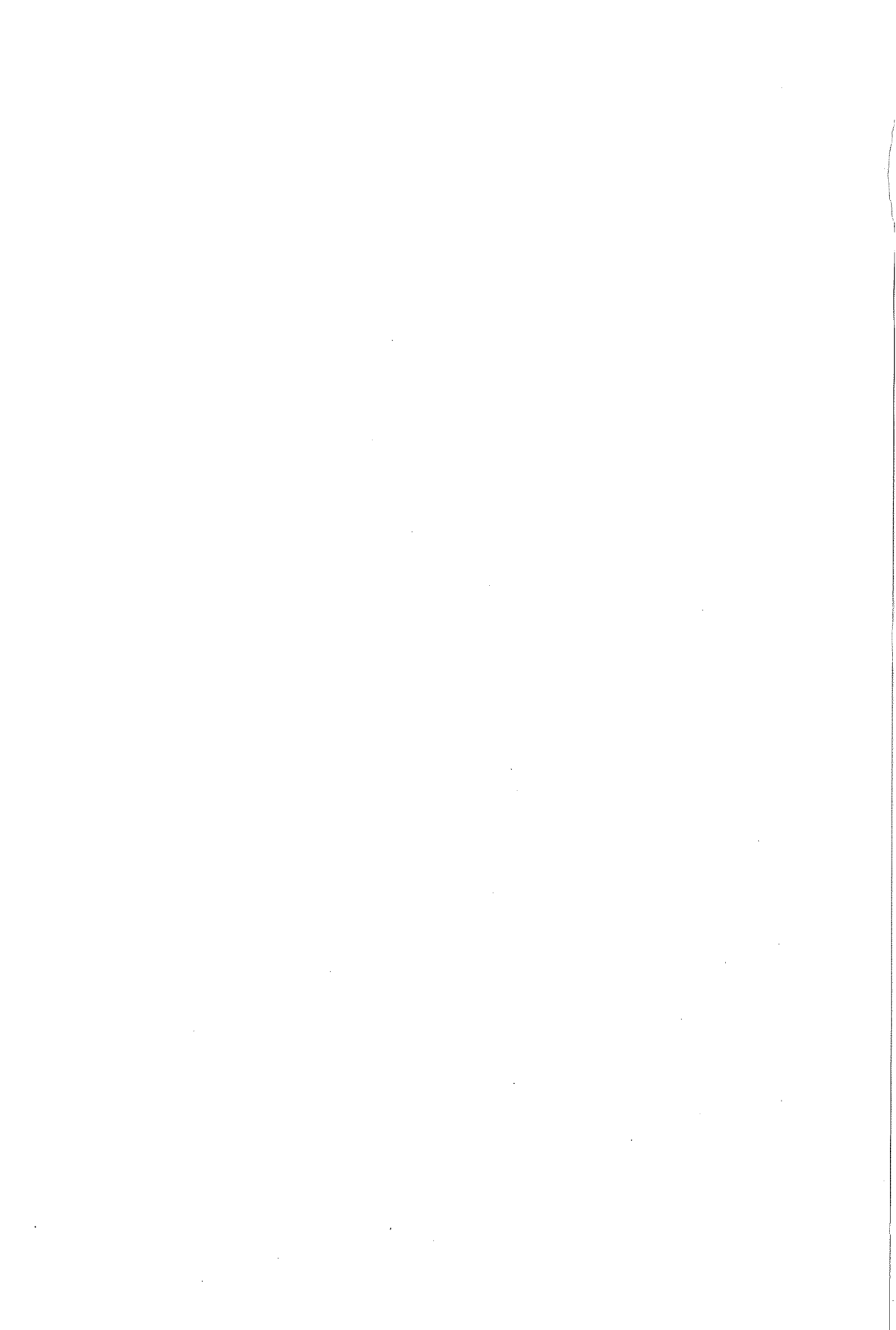
A fuzzy number can be defined as a normal fuzzy set over the real line.

The membership function of a fuzzy number can be interpreted as a possibility distribution (namely when translating linguistic declarations to a fuzzy set representation), because the so called *possibility measures* are defined over a set of axioms that are analog to the definition of fuzzy sets.





Fotografia, Montagem
Impressão e Acabamentos
Tip. Nocamil
COIMBRA



ÍNDICE

<i>J. A. O. Soares, J. A. Assis Lopes e C. M. F. Monteiro</i> , Estudo de eficiência na Bolsa de Valores de Lisboa: ciclicidade e regras de compra	119
<i>J. A. S. Cabral e R. C. Guimarães</i> , Análise de uma regra de compra de mercadorias sujeitas a flutuações de preços	129
<i>M. L. Alves e M. T. Almeida</i> , Simulated annealing for the simple plant location problem: a computational study	145
<i>J. O. Cerdeira</i> , The multi-depot traveling salesman problem	159
<i>M. H. Correia, A. M. Gomes, J. F. Oliveira e J. S. Ferreira</i> , Problemas de empacotamento tridimensional	169
<i>F. Jimenez Gomez e M. J. V. Bonnet</i> , On the Kalman filtering for the random stationary intensity of a doubly-stochastic Poisson process	181
<i>V. Miranda, M. A. C. Matos e J. T. Saraiva</i> , Fuzzy flows in linear networks with correction for some dependencies in nodal uncertainties	185



Associação Portuguesa para o Desenvolvimento
da Investigação Operacional

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