

INVESTIGAÇÃO OPERACIONAL

Volume 27 — nº 1 — Junho 2007

Publicação Semestral

Editor Principal: José F. Oliveira
Universidade do Porto

Comissão Editorial

M. Teresa Almeida Inst. Sup. Econ. e Gestão	J. Rodrigues Dias Univ. de Évora	N. Maculan Univ. Fed., Rio Janeiro
C. Henggeler Antunes Univ. de Coimbra	Laureano Escudero IBM, Espanha	Rui Oliveira Inst. Superior Técnico
Marcos Arenales Univ. de São Paulo	Edite Fernandes Univ. do Minho	J. Pinto Paixão Univ. de Lisboa
Jaime Barceló Univ. de Barcelona	J. Soeiro Ferreira Univ. do Porto	M. Vaz Pato Inst. Sup. Econ. e Gestão
Eberhard E. Bischoff Univ. of Wales, Swansea	J. Fernando Gonçalves Univ. do Porto	Mauricio G. Resende AT&T Labs Research
C. Bana e Costa Inst. Superior Técnico	Luís Gouveia Univ. de Lisboa	A. Guimarães Rodrigues Univ. do Minho
M. Eugénia Captivo Univ. de Lisboa	Rui C. Guimarães Univ. do Porto	António J. L. Rodrigues Univ. de Lisboa
Domingos M. Cardoso Univ. de Aveiro	Joaquim J. Júdice Univ. de Coimbra	J. Pinho de Sousa Univ. do Porto
João Clímaco Univ. de Coimbra	J. Assis Lopes Inst. Superior Técnico	Reinaldo Sousa Univ. Católica, Rio Janeiro
J. Dias Coelho Univ. Nova de Lisboa	Carlos J. Luz Inst. Polit. Setúbal	L. Valadares Tavares Inst. Superior Técnico
João P. Costa Univ. de Coimbra	Virgílio P. Machado Univ. Nova de Lisboa	B. Calafate Vasconcelos Univ. do Porto
Ruy Costa Univ. Nova de Lisboa	Manuel Matos Univ. do Porto	L. Nunes Vicente Univ. de Coimbra
		Victor V. Vidal Techn. Univ. of Denmark

Modelo não-linear de longo prazo para a potência requerida do sistema brasileiro de eletricidade

Miguel Afonso Sellitto [†]
José Luis Duarte Ribeiro [‡]
Fabiano Petrillo ^{*}

[†] Programa de Pós-Graduação em Engenharia de Produção e Sistemas.
Universidade do Vale do Rio dos Sinos, São Leopoldo
Brasil
sellitto@unisinos.br

[‡] Programa de Pós-Graduação em Engenharia de Produção.
Universidade Federal do Rio Grande do Sul, Porto Alegre
Brasil.
ribeiro@producao.ufrgs.br

^{*} Programa de Pós-graduação em Engenharia Elétrica
Universidade Federal do Rio Grande do Sul, Porto Alegre
Brasil
fabiano@petrillo.com

Abstract

The formerly closed Brazilian electrical energy market has been disclosed for investors and domestic consumers, which now have choices for their decisions concerning energy sales and purchasing, in long-term contracts. To establish minimal planning condition, suppliers need to forecast the required load in order to project and operate power plants. With the aid of a method that integrates non-linear regression and soft-system methodology, we depict a non-linear model for the gross total electric power requirement, counting on domestic activity, population and human developing index. A goodness-of-fit better than 96% was achieved. The final analysis focused on the strong contingency period occurred in Brazil in 2001.

Resumo

O mercado brasileiro de energia elétrica abriu-se recentemente a investidores e consumidores, que agora têm opções de longo prazo para compra e venda de energia. Para que haja condições mínimas de planejamento neste mercado, é necessário um modelo de longo prazo capaz de prever as necessidades futuras de suprimento de energia elétrica. Por meio de um método que integra regressão não-linear múltipla com SSM (*soft system methodology*), foi proposto um modelo para o requisito de potência, calculado a partir do nível de atividade comercial brasileira, do total de população e de seu nível de desenvolvimento humano. O ajuste obtido pelo modelo foi superior a 96%. Uma análise

final focou no período de contingenciamento de oferta de energia ocorrido no Brasil em 2001, conhecido como apagão.

Keywords: electrical power, multiple nonlinear regression, multivariate data analysis, SSM, long-term load forecasting.

Title: Long-term non-linear model for the power requirement of the brazilian electricity system.

1 Introdução

A indústria brasileira de energia elétrica se caracteriza hoje pela livre concorrência na compra e venda de energia. No fim dos anos 1990, foi introduzida a competição entre produtores pelo mercado consumidor e a contratação de longo prazo para fornecimento de energia [Silva, 2001].

O processo de reforma do setor teve como objetivos finais assegurar os investimentos necessários para a expansão da oferta de energia e a viabilidade econômica do setor. Para tanto, era necessário otimizar o uso dos recursos disponíveis e garantir um fornecimento confiável de energia elétrica ao menor custo possível, satisfazendo clientes e prevenindo operações deficitárias, antes subsidiadas pelo estado brasileiro. Para atingir estes objetivos, foram adotados como princípios a competição em geração e comercialização para clientes livres e monopólios regulados em transmissão, distribuição e comercialização para clientes cativos, semelhantes aos princípios que orientaram o processo de reforma em outros países [Comitê de revitalização do modelo do setor elétrico, 2002]. A energia elétrica torna-se uma *commodity*, comercializada segundo leis de mercado e previsões de preço e consumo.

A taxa por unidade de tempo com que é requisitada a energia é o requisito de potência que o sistema deve oferecer. Como ainda não existe tecnologia viável de armazenagem, a energia requerida é produzida no ato do consumo, impossibilitando a produção antecipada quando os custos de produção estão mais baixos ou quando a demanda for menor. Diferentemente de outros produtos, a energia elétrica tem características que devem ser consideradas na previsão de demanda.

Um modo de fazer previsões de demanda de energia elétrica é reconhecer padrões de comportamento em séries históricas e prever o comportamento futuro. Outro modo é identificar fatores causais que afetam o comportamento e extrapolá-lo, segundo um modelo matemático. As previsões podem ser de curto e de longo prazo. No primeiro caso, prevê-se a demanda dos próximos dias e semanas, assegurando economicidade e segurança na operação cotidiana. No segundo caso, a previsão é para meses e anos, subsidiando o planejamento da expansão do sistema para os próximos períodos [Figueiredo et al., 2005].

O objetivo deste artigo é apresentar um modelo matemático que possa descrever o quanto o mercado brasileiro de energia elétrica requisiou de potência no período 1996-2000 e projetar o requisito para períodos futuros. Trata-se de um modelo de longo prazo. Assume-se que o requisito de potência seja afetado por muitos e complexos fatores, de comportamento não-linear e causalidade recursiva.

Tal assunção justifica uma abordagem do tipo SSM (*soft-system methodology*) e o uso de árvores de realidade atual (ARA) para a estruturação da situação. A SSM não objetiva encontrar uma expressão única (*hard*) para um problema, mas busca um procedimento de investigação que possa ser replicado no espaço e no tempo e que possa chegar a um

modelo válido para um fenômeno específico. O que é sistêmico é o procedimento investigatório e não o modelo obtido [Checkland e Scholes, 1998]. Árvores de realidade, por sua vez, constroem e comunicam visões compartilhadas sobre uma situação não-linear e recursiva e são um passo intermediário em direção ao modelo [Dettmer, 1997]. Formou-se um grupo focado com especialistas e uma regressão não-linear múltipla foi proposta por observação da ARA.

Não é objetivo revisar modelos de previsão de demanda por energia elétrica. Citam-se apenas algumas incidências na literatura, surgidas mais notadamente em veículos de engenharia elétrica. Figueiredo *et al.* [2005] sugerem redes neurais para abordar o problema. Bunn [2000], Park *et al.* [1991] e Peng *et al.* [1992] usam técnicas de inteligência computacional, tais como redes neurais e neuro-fuzzy. Papalexopoulos e Hesterberg [1989], Hagan e Behr [1987] e Rahman e Hazim [1993] usam séries temporais e regressões, incluindo variáveis categóricas, para a previsão de curto prazo (palavra-chave: *short-term load forecasting*). Parlos *et al.*, [1996], Al-Saba e El-Amin [1999], Da, [2000] e Kandil *et al.* [2002] usam técnicas diversas de inteligência computacional, séries temporais e regressões multivariadas para previsão de médio e longo prazo (palavra-chave: *long-term load forecasting*). Caio e Berman [1998] apresentam modelos específicos, não-sistêmicos, usados por companhias de distribuição de energia brasileiras para previsão de demanda.

A contribuição essencial deste artigo é propor um procedimento recursivo para se chegar a uma expressão para o requisito de potência de longo prazo de um sistema elétrico. A expressão não-linear alcançada é válida para o caso do Brasil de 1996 a 2000, mas o procedimento pode ser usado em outros casos e produzir outras expressões. O restante do artigo está organizado em: considerações sobre modelos não-lineares, SSM e estruturação de problemas; metodologia de pesquisa; discussão e considerações finais.

2 Considerações sobre modelos não-lineares

Em um problema não-linear, a relação teórica entre as variáveis de interesse assume a forma da equação 1, na qual ζ_i são preditores, θ_i são parâmetros e f é não-linear.

$$Y = f(\zeta_1, \zeta_2, \dots, \zeta_k; \theta_1, \theta_2, \dots, \theta_p) + \varepsilon \quad (1)$$

Sejam n observações $[Y_i, \zeta_{1i}, \zeta_{2i}, \dots, \zeta_{ki}]$, para $i = [1, 2, \dots, n]$. A soma quadrática de erros para o modelo não-linear é dada pela equação 2, na qual os vetores ζ e θ são respectivamente as observações das variáveis preditoras e as estimativas dos parâmetros. Como Y_i e ζ_i são observações fixas, a soma quadrática resulta uma função exclusiva de θ . O vetor que minimiza esta soma, se os erros forem normais e independentes, é também a estimativa de máxima verossimilhança de θ . Chega-se a esta estimativa derivando parcialmente a equação em relação a todos os θ_i e igualando a zero, o que gera um sistema de equações normais. A solução destas equações fornece o vetor $\theta = [\theta_1, \theta_2, \dots, \theta_n]$ que minimiza $S(\theta)$ [Draper e Smith, 1981].

$$S(\theta) = \sum_{i=1}^n [Y_i - f(\zeta_i, \theta)]^2 \quad (2)$$

Exceção feita a alguns casos, não é possível resolver diretamente as equações normais. Podem ser necessários métodos iterativos de procura numérica direta. Os principais métodos têm formatos semelhantes: (i) estabelecem-se valores iniciais para as variáveis θ ; (ii) gera-se a curva definida pelas variáveis iniciais e calcula-se a soma quadrática de resíduos; (iii) segundo um dado algoritmo capaz de diminuir a soma quadrática, modificam-se as variáveis; e (iv) aplica-se recursivamente o algoritmo até que duas sucessivas aplicações não apresentem melhoria significativa. Diferentes estimativas iniciais podem produzir diferentes resultados em diferentes números de iterações, o que torna desejável o uso de várias estimativas iniciais e observar o resultado e a velocidade de convergência [Mutolsky, 2001].

Diversos métodos surgem na literatura. O método de Gauss-Newton ou da linearização expande o modelo não-linear em uma série de Taylor, aproximando-o por um modelo linear. Estimam-se então os parâmetros θ pelo método dos mínimos quadrados em iterações até que um critério de término seja atingido. Há ainda o método da descida mais íngreme e o método de Levenberg-Marquardt (LM), que combina os outros dois. Se está longe da convergência, LM comporta-se como descida mais íngreme; se está próximo, como Gauss-Newton. Já o método GRC (*generalized reduced gradient*) melhora uma solução inicial, explorando-a em várias direções, até que um critério encerre a exploração. Diferentes métodos podem chegar a resultados levemente diferentes [Draper e Smith, 1981; Mutolsky, 2001; Silva, 2004].

Os resultados da regressão não-linear são representativos se alguns pressupostos forem atendidos: (i) o modelo tem alguma justificativa, pois a regressão apenas ajusta parâmetros da equação escolhida, não tentando outras equações; (ii) os erros são normais, com variância constante, independentes em relação a Y e não-correlacionados entre si; (iii) a imprecisão na medição de Y é pequena perante sua variabilidade; (iv) o resultado apresentado faz sentido científico, ou seja, não há proporções maiores do que um, não há taxas de crescimento negativas ou não há limites fora do âmbito original dos dados; e (v) as estatísticas obtidas são aceitáveis. Além dos valores de melhor ajuste, uma análise de regressão não-linear deve oferecer um erro-padrão da estimativa, intervalos de confiança assintóticos dos parâmetros e o coeficiente de determinação R^2 , que informa a fração da variância total explicada pelo modelo: quanto mais próximo a 1, melhor. Outros testes são o F , em formato de *anova*, o t aplicado aos erros-padrão e o valor- p . Vale o princípio da parcimônia: um bom modelo mantém o compromisso entre o ajuste e o número de preditores [Draper e Smith, 1981].

Quanto à escolha do modelo, além de apresentar boas estatísticas de ajuste, este deve fazer sentido em uma base teórica de conhecimento. Neste artigo, foi usado um método de trabalho integrando duas bases de conhecimento, a regressão não-linear e a SSM (*soft-system methodology*), que produziu um modelo que faz sentido em relação a uma observação empírica da realidade.

3 SSM e estruturação de situações complexas de interesse

A SSM inclui procedimentos cujos objetivos são obter conhecimento sobre uma situação de interesse e estruturá-la, segundo modelos mentais, de modo a gerar um modelo que possa ser útil em interpretações e intervenções modificadoras da realidade. A SSM pode estruturar situações de interesse e relações com o contexto, principalmente quando os agentes têm visões e interesses diversos sobre o tema. Pode ocorrer, também, que objetivos de agentes sejam incertos, múltiplos ou até mesmo conflitantes entre si.

Situações tratadas por SSM têm algumas características em comum. São intrincadas, não existe um modelo unificado e a situação é complexa, com interações e interferências

mútuas entre objetos do cenário e evolução temporal. A SSM admite múltiplas percepções sobre uma dada situação, e que estas se originam da diversidade de modelos mentais que analistas usam para interpretar a realidade. Pode ser usada para explorar a situação, identificar as partes interessadas e seus pontos de vista, modelar critérios de avaliação de desempenho e oferecer contrapontos a modelos quantitativos [Jackson, 1993; Wilson, 2001; Neves *et al.*, 2004].

Abordagens para situações complexas de interesse podem residir em um *continuum* que vai de *hard* a *soft*. Um critério que pode auxiliar a posicionar uma situação de interesse é o grau de concordância entre interessados sobre o que é a situação. Se não houver divergência alguma, como por exemplo, em equações, o problema é totalmente *hard*. Se toda parte interessada tiver uma interpretação diferente para o problema, este é totalmente *soft*.

Na extremidade *hard* do *continuum* pode ser aplicada a engenharia de sistemas (SE), que consiste em: (i) definir a situação de interesse; (ii) formular uma metodologia, composta de técnicas apropriadas; (iii) usar as técnicas para encontrar soluções possíveis iterativamente com (iii); (iv) selecionar uma das soluções possíveis; e (v) implementá-la e medir os resultados [Wilson, 2001]. Na extremidade *soft* pode ser aplicada a SSM: (i) definir a situação de interesse; (ii) expressar a situação por um modo comunicável, tal como diagramas; (iii) selecionar conceitos e objetos que possam ser relevantes para o diagrama; (iv) montar os conceitos e objetos em uma estrutura intelectual coerente; (v) usar a estrutura para explorar a situação iterativamente com (iii) e (iv); (vi) definir mudanças; e (vii) implementá-las e verificar os resultados [Checkland e Scholes, 1998]. No caso da SE, as técnicas utilizadas no passo iii contém tanto os conceitos quanto a estrutura usados no problema, e são unicamente definidos. Na SSM, conceitos (passo iii) e estrutura (passo v) são separados e podem requerer iteratividade durante o aprendizado ou até que se chegue a uma definição.

A SSM é ilustrada na Figura 1. Nas etapas 1 e 2, há esforço para construir uma representação o mais rica possível, não do problema, mas da situação na qual este é percebido como um problema. A etapa 3 inclui expressar conceitos subjacentes ao problema, as definições de raiz, que poderão ser influentes na evolução de objetos que afetam o resultado do mesmo. Na etapa 4, constrói-se em linguagem humana o modelo conceitual, que pode usar lógica formal ou outra qualquer base conceitual de pensamento sistêmico, tal como a dinâmica dos sistemas ou as árvores de realidade. Na etapa 5, compara-se o modelo conceitual com a realidade. O modelo serve como guia de debate e poderá haver iterações, caso haja aprendizado significativo na comparação com o real. Na etapa 6, escolhem-se ações desejáveis e factíveis transformadoras da realidade, para, na etapa final, implementar estas ações. Os resultados devem ser medidos, realimentando o processo [Checkland e Scholes, 1998].

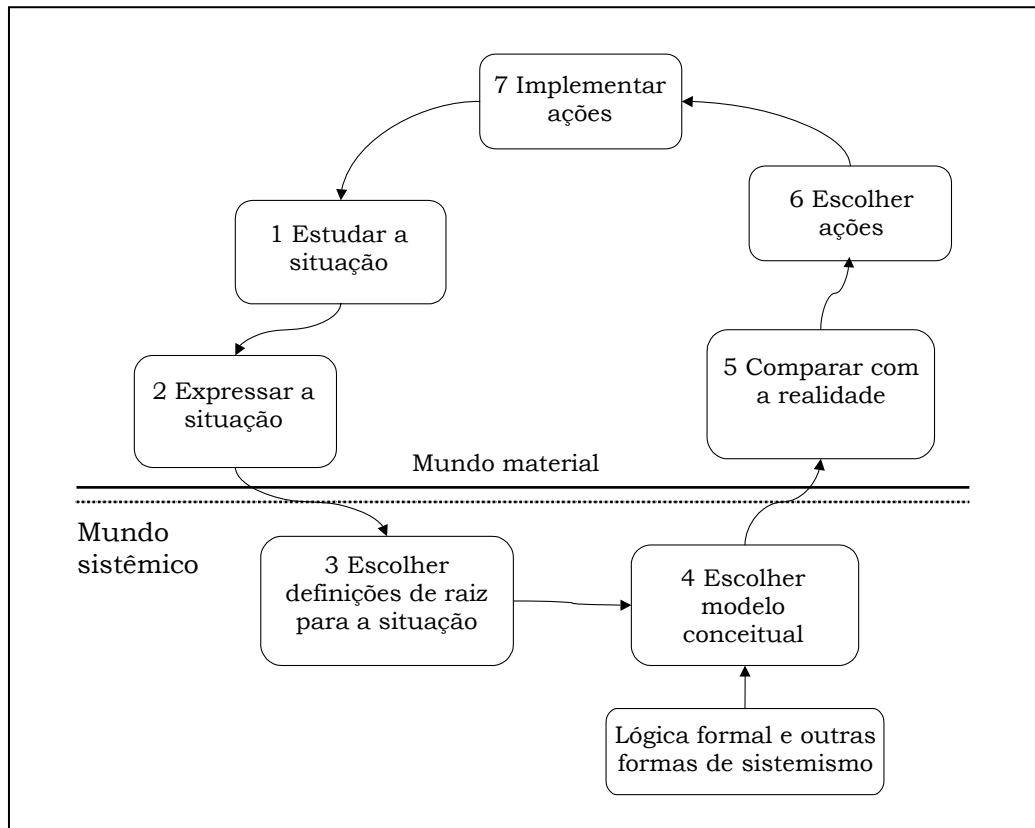


Figura 1: Etapas da SSM [Checkland e Scholes, 1998]

No esquema metodológico da figura, é admitida na etapa 4 a presença de outras formas de estruturação sistêmica do pensamento. Várias formas são apresentadas na literatura.

Uma delas são os mapas cognitivos [Eden, 1988; Ensslin *et al.*, 2001]. Mapas cognitivos são construções gráficas formadas por nós e setas, nas quais os nós representam pontos de vista expressos por frases afirmativas e, se necessário, contrastada pelo seu contraditório, para reforçar a multiplicidade de visões. Nós são ligados por setas, que indicam como pontos de vista são influenciados ou têm implicações em outros pontos de vista. Os mapas cognitivos são um passo intermediário em direção aos pontos de vista fundamentais, os critérios de sucesso contra os quais a situação problemática pode ser julgada [Belton *et al.*, 1997, Bana e Costa *et al.*, 1999].

Outra forma de estruturação sistêmica são as ARA, ou árvores da realidade atual (no original *CRT – Current Reality Tree*). A ARA é uma das ferramentas de análise sistêmica que compõem os processos de raciocínio da TOC, os *TP (thinking processes)*, cujo objetivo é oferecer uma abordagem sistêmica que explicita nexos causais e explique as relações suficientes e necessárias entre fenômenos observados em situações complexas de interesse, explorando a capacidade de aprendizado de analistas. As outras são: árvore da realidade futura (ARF); diagrama de dispersão de nuvem (DDN); árvore de pré-requisitos (APR); e árvore de transição (AT). A ARA define o que mudar, a ARF define para o que mudar, o DDN explicita o pressuposto errôneo que precisa ser mudado, a AT define os passos para mudar e a APR descreve o que superar para mudar [Kendall, 1998].

O objetivo da ARA é representar lógica e claramente as relações efeito-causa-efeito válidas em uma situação complexa de interesse, geralmente multifacetada e sem

consenso. Definida a situação e listados efeitos observados, estes são conectados em relações de precedência-consequência. Aplicam-se ressalvas legítimas, que refinam e depuram a ARA, conduzindo a uma forma final realimentada, que pode, inclusive, não ter entes sem precedências, ou seja, uma situação que é causa de si mesma [Sellitto, 2005]. Para validar relações de causalidade, é preciso descobrir qual o pressuposto que a explica, o nexos causal. Havendo um candidato a pressuposto, aplicam-se os seguintes testes: (i) para que exista [efeito] é necessário que exista [causa], porque existe [pressuposto]; e (ii) a não ser que exista [causa] não existirá [efeito], porque existe [pressuposto] [Goldratt, 1990; Sheinkopf, 1999]. Uma diferença entre ARA e diagrama sistêmico é que aquela preocupa-se apenas com aspectos qualitativos da relação de causalidade (existe ou não), não considerando distribuições ou correlações entre variáveis nem se a influência é positiva ou negativa. A forma qualitativa torna a ARA mais fácil de ser usada em abordagens iniciais, exploratórias da situação.

Técnicas para a construção e validação da ARA são apresentadas em Sheinkopf [1999]. Uma aplicação da ARA em gestão de serviços de saúde pública, uma modelagem de sistemas interorganizacionais complexos, com causalidades circulares e recursivas, é apresentada em Sellitto[2005]. Em Morgan [1996], o fenômeno da inflação e suas recursividades também é ilustrado, embora o autor não mencione a expressão ARA.

4 A pesquisa

O objetivo deste artigo foi obter um modelo matemático capaz de descrever a potência requerida pelo mercado consumidor brasileiro de energia elétrica entre 1996 e 2000. Potência requerida é diferente de potência demandada. A potência requerida é a potência realmente usada; potência demandada é aquela que os consumidores acreditam que vão usar e contratam, podendo não ser usada na integralidade. A potência elétrica não é um produto que o produtor entrega ao consumidor. O consumidor vai requerendo à medida que vai conectando equipamentos elétricos na rede de alimentação, até um limite físico.

Como em Neves *et al.* [2004], foi útil o uso de uma metodologia de estruturação de problema, para aprendizado e ganho de visão compartilhada sobre os múltiplos interesses em questão. Como em Bana e Costa *et al.* [1999], a primeira parte da pesquisa foi dedicada a estruturar e enquadrar o problema em um processo construtivista de aprendizado, que ofereceu uma representação compartilhada da situação. Em Bana e Costa *et al.* [1999], o método estruturante foi o mapeamento cognitivo e os pontos de vista fundamentais. Nesta pesquisa, foi a ARA dos processos de pensamento da TOC.

O método de trabalho foi: (1) formulação, em sessões de grupo focado de especialistas com múltiplas visões e interesses, de um modelo qualitativo sistêmico que represente a complexidade subjacente ao fenômeno; (2) identificação no modelo de uma estrutura central que possa ser apreendida por indicadores; (3) obtenção de dados no tempo para os indicadores; (4) buscar um modelo de regressão que descreva os dados; e (5) testar o modelo e discutir a aplicação, inclusive com projeções sobre o que sucedeu ao contingenciamento devido à escassez na geração, o apagão de 2001. A teoria de grupos focados e o método empregado na sessão são apresentados em Ribeiro e Newmann, [2003]. O que foi feito corresponde às etapas de 1 a 5 da Figura 1. A seguir, descrevem-se os resultados.

4.1 O modelo

As sessões de grupo focado foram conduzidas por dois pesquisadores. Compuseram o grupo um especialista de agência reguladora de energia elétrica, cujo interesse é conhecer

a evolução do requisito de potência para gerenciar expansões do sistema; um especialista de empresa fornecedora de serviços para o setor, cujo interesse é que haja crescimento do requisito; e um pesquisador em produção mais limpa, cujo interesse é entender a evolução do requisito para propor medidas de aumento de eficiência energética. Interesses conflitantes e visões multifacetadas estiveram presentes no grupo. Os pesquisadores contribuíram com visões acadêmicas sobre o tema. Seguem os achados do grupo focado.

O estudo do requisito de potência elétrica depende de fenômenos econômicos e sociais que resultam interligados em arranjo realimentado com causalidades recursivas. Este conjunto de relações foi representado graficamente por uma ARA construída pelos pesquisadores e validada pelos participantes do grupo. A ARA é apresentada na Figura 2.

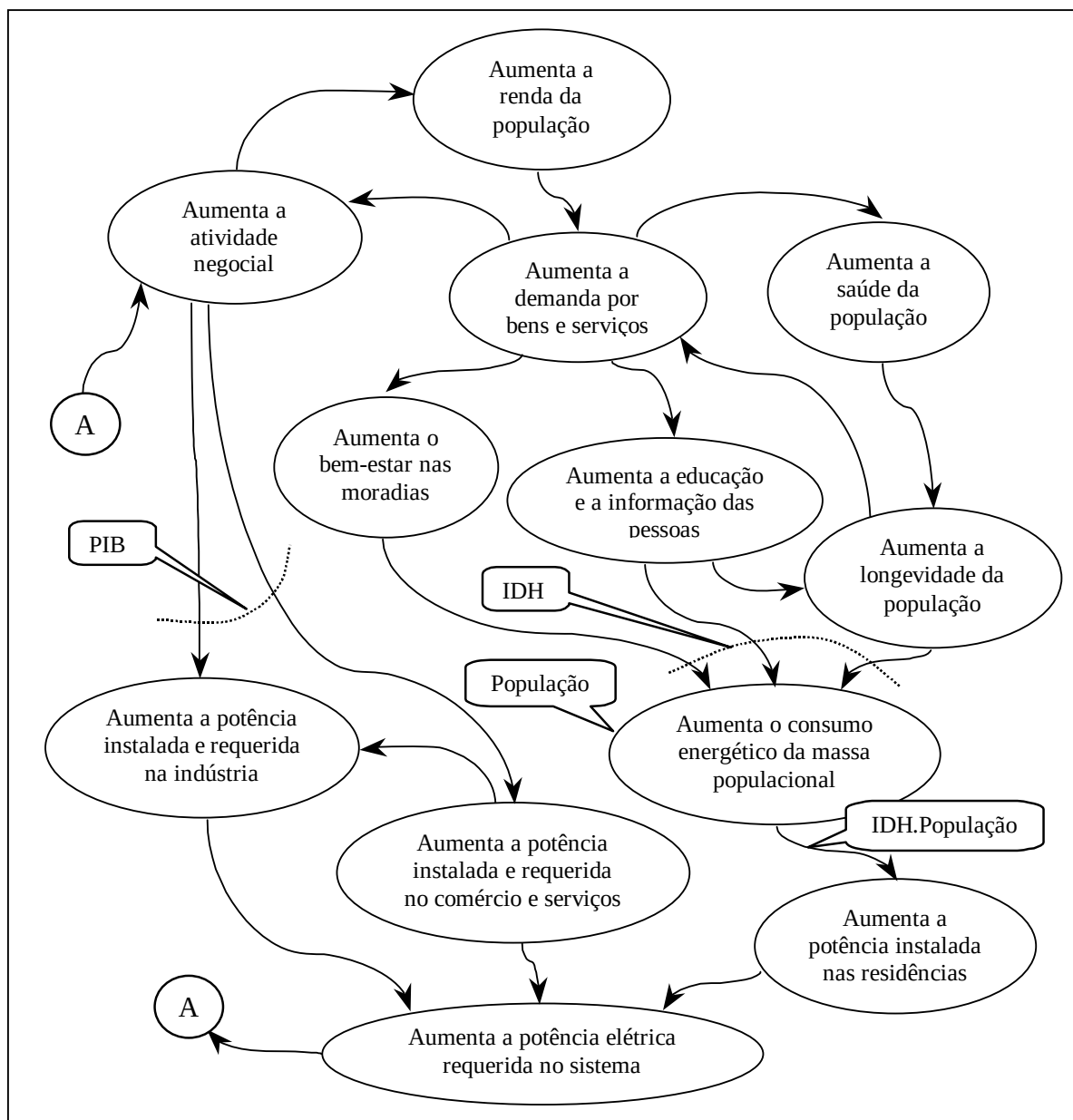


Figura 2: Modelagem sistêmica do requisito de potência pela ARA

A ARA, válida apenas para o caso em estudo, permitiu que se construísse uma visão compartilhada entre as partes interessadas no tema, algumas delas com objetivos conflitantes. Se o mesmo estudo for feito por outros especialistas ou em outro espaço ou tempo, é possível que se chegue a outra ARA. A ARA atende o requerido na revisão: que um modelo de regressão não-linear deve possuir justificativa.

O lado direito da ARA descreve os aspectos sociais do fenômeno, apreendidos por três construtos latentes: a renda, a educação e a longevidade. Estes construtos participam do IDH, o índice de desenvolvimento humano, criado pela ONU para servir como realimentação do processo de desenvolvimento social mundial. O IDH é calculado segundo pesos dados aos construtos e respectivos indicadores, a renda *per capita*, a taxa de analfabetismo, os anos de estudo e a expectativa de vida ao nascer. Para o cálculo do valor dos construtos, subtrai-se do valor observado para o indicador um valor de base, pré-definido e equivalente a zero. O resultado é dividido pela diferença entre os valores máximo e mínimo pré-estabelecidos, resultando um valor entre 0 e 1. Quanto mais próximo de 1, melhor o IDH [Bezerra e Diwan, 2001]. As implicações do uso do IDH podem ser acessadas em [UNDP, 2002].

O lado esquerdo da ARA descreve o fenômeno econômico, apreendido pela atividade geral do mercado, o que remete a indicadores de atividade econômica. Antes de definir o indicador econômico, testaram-se as correlações entre as observações do requisito de potência com os PIB's geral e industrial, brutos e dessazonalizados, e com a produção industrial bruta. Também testaram-se as correlações do requisito de potência com as populações rural, urbana e total. As maiores correlações positivas foram com a população total e o PIB geral dessazonalizado, escolhendo-se estes indicadores para o modelo.

O Quadro 1 resume as informações. Não há informações mensais do requisito de potência antes de 1996, só anuais. Como as frequências dos indicadores são diferentes, optou-se por uma frequência trimestral. As Tabelas 1 e 2 apresentam as vinte observações e suas correlações. Destaque-se a correlação unitária entre IDH e população. Observações posteriores a 2000 foram excluídas no modelo devido à irregularidade que foi introduzida pelo contingenciamento de energia imposto pelo estado brasileiro em 2001, o apagão, que será discutido mais a frente. Mesmo que vinte observações sejam poucas para uma conclusão robusta, reforça-se que esta é uma contribuição metodológica, podendo e devendo ser aplicada em situações com mais dados consolidados.

Quadro 1: Resumo da variável e dos indicadores

resposta e indicadores	fonte	coleta e forma de uso no modelo
potência elétrica	ONS [2005]	medido mensalmente desde 1996/1, apresenta a média de consumo dos meses do trimestre em MW
PIB geral dessazonalizado	IBGE [2002]	medido trimestralmente desde 1990/1, soma os PIB's rural, industrial e de serviços dessazonalizados e referenciados ao ano 2.000
população total	IBGE [2002]	medido por censo de 10 em 10 anos, desde 1970, mais 1996, nos demais anos e trimestres interpolado
índice de desenvolvimento humano IDH	UNDP [2002]	mede a longevidade, a educação e o consumo de cinco em cinco anos, desde 1975 mais 1999, interpolado nos demais anos e trimestres

Tabela 1: Observações para o modelo da regressão

trimestre	potência MW médio	PIB geral R\$ dessaz equivalente a 2000	população total, milhões	IDH de 0 a 1, bom = 1
1996/1	34.370	245,51	155,773	0,731
1996/2	34.060	248,93	156,409	0,733
1996/3	34.590	249,86	157,060	0,734
1996/4	35.330	254,55	157,727	0,735
1997/1	36.336	255,82	158,408	0,736
1997/2	35.917	258,31	159,105	0,737
1997/3	36.960	259,62	159,817	0,738
1997/4	37.640	257,78	160,545	0,739
1998/1	38.523	257,38	161,287	0,740
1998/2	37.872	261,16	162,045	0,741
1998/3	38.413	259,85	162,818	0,742
1998/4	37.819	254,50	163,606	0,744
1999/1	39.137	256,00	164,409	0,745
1999/2	38.722	259,80	165,228	0,746
1999/3	39.300	260,69	166,061	0,747
1999/4	39.573	264,78	166,910	0,748
2000/1	40.650	268,91	167,774	0,749
2000/2	40.639	269,95	168,654	0,750
2000/3	40.719	272,44	169,548	0,751
2000/4	42.003	275,40	170,458	0,752

Tabela 2: Correlações entre os indicadores do modelo

	potência	PIB	população
PIB	0,91		
população	0,97	0,90	
IDH	0,98	0,90	1,00

Na ARA, PIB e população surgem em ramos paralelos, sugerindo adição. IDH e população surgem em ramo serial, sugerindo multiplicação na qual o IDH pondera a população. Exigiu-se que a soma de θ_1 e θ_2 fosse igual a um para evidenciar a contribuição de cada parcela. Sugere-se então, especulativamente, a equação 3, composta por uma parcela econômica, ligada ao PIB e uma parcela social, ligada à população e ponderada pelo IDH. Um fator de escala θ_1 foi incluído.

$$Y = \theta_0 \cdot [\theta_1 \cdot (PIB)^{\theta_{11}} + \theta_2 \cdot (POPULAÇÃO)^{\theta_{22}} \cdot (IDH)] \quad (3)$$

4.2 Ajuste ao modelo

Antes de ajustar-se o modelo da equação 3 aos dados da Tabela 1, foram feitas adequações que se mostraram úteis. Preferiu-se usar para o PIB e para a população valores normalizados em relação às observações de 2000/1 ($POP/POP\ 2000$). O IDH não foi modificado. Como existe um elo entre os aumentos da renda e da atividade comercial,

que aumentam o requisito de potência, suspeitou-se de uma defasagem entre as séries temporais do PIB e da observação. Testaram-se diversas defasagens e a que apresentou maior correlação foi (PIB_{-2}/PIB 2000), ou seja, correlacionando a potência atual com o PIB de dois trimestres atrás.

Para o ajuste, foram usados dois recursos computacionais, a rotina *solver* do *software* Excel® e a rotina *non-linear regression* do *software* SPSS®. A rotina *solver* usa o algoritmo GRC e o SPSS® usa o método L-M. O *solver* foi configurado para minimizar o erro-padrão da estimativa. Testaram-se alguns conjuntos de soluções iniciais. A Tabela 3 apresenta os melhores resultados com cada recurso. Foram calculados R^2 , intervalos de confiança, o erro-padrão da estimativa e o viés. Como o R^2 do modelo do *solver* foi um pouco melhor, segue-se este modelo. Como antecipado na revisão, os métodos chegaram a pequenas diferenças de resultados. Foi tentado um expoente para IDH. O valor ótimo foi 1,001 e não houve melhoria no R^2 . Por parcimônia, excluiu-se o expoente do modelo.

Tabela 3: Modelo não-linear para o requisito de potência do mercado brasileiro [1996-2000]

Modelo	Potência (MW médio) =
produzido	52000. [0,1. $[(PIB_{-2}/PIB\ 2000)^{2,61}] + 0,9. [(POP/POP\ 2000)^{1,71}. IDH]$]
pelo <i>solver</i>	$R^2 = 0,96470$
Modelo	Potência (MW médio) =
produzido	51849. [0,108. $[(PIB_{-2}/PIB\ 2000)^{2,46}] + 0,892. [(POP/POP\ 2000)^{1,72}. IDH]$]
pelo SPSS	$R^2 = 0,96395$
Intervalos de	IC 95% $\theta_0 = [31.713; 71.984]$;
confiança	IC 95% $\theta_{11} = [-25,4; 30,3]$;
assintóticos	IC 95% $\theta_{22} = [-2,049; 5,502]$;
erro-padrão	454,7 MW médios
da estimativa	
viés da	1,00 (nenhum viés)
estimativa	

A adequação do modelo é verificada pela significância da regressão e por inspeção de resíduos. A regressão é significativa se a variável dependente se relacionar de modo significativo com as variáveis dependentes. Um modo de testar esta significância é calculando a proporção entre a variância dos dados explicada e a não explicada pela regressão. Dados os diversos graus de liberdade, uma proporção entre variâncias segue uma distribuição F . O teste é definido em termos de $F_{teste} = [\text{variância explicada}/\text{variância não explicada}]$ e $F_{critico}$, obtido de uma distribuição F com nível de confiança α e $(k-1)$ e $(n-k)$ graus de liberdade. Se $F_{teste} > F_{critico}$, a variância explicada é maior do que a não explicada, caracterizando, com confiabilidade $(1-\alpha)$, que a regressão é significativa, rejeitando-se $H_0: R^2 = 0$ [Corrar et al., 2004].

O teste é apresentado sob a forma de *anova* na Tabela 4. A regressão se vale de quatro parâmetros independentes ($\theta_0, \theta_1, \theta_{11}$ e θ_{22} , θ_2 não é independente), logo os preditos têm quatro graus de liberdade (gl). Como os graus de liberdade totais são 19 (vinte observações menos uma), tem-se $[gl = 15]$ para os resíduos. As somas e as médias quadráticas das parcelas referentes à estimativa e aos resíduos produzem $F_{teste} > F_{critico}$, o que caracteriza a significância da regressão. O valor- p , próximo a zero, é a probabilidade de que se tenha chegado ao resultado por acaso, a regressão não seja significativa e R^2 não seja diferente de zero.

Tabela 4: Anova para o modelo

Anova	gl	SQ	MQ	F teste	F crítico	Valor-p
Predito	4	9,6 E+07	2,4 E+07	102	3,06	1,09 E-10
Resíduo	15	3,5 E+06	2,3 E+05			
Total	19					

Passa-se à análise dos resíduos. Os fundamentos teóricos e os detalhes de cálculos são encontrados em Hair *et al.*, 1998; Milone e Angelini, 1995 e Werkema e Aguiar, 1996.

A Tabela 5 apresenta uma análise sumária dos valores preditos e dos resíduos, informando que as observações extremas dos resíduos padronizados e de Student distam cerca de dois desvios-padrão dos centros das respectivas distribuições. Também informa que a menor significância de uma distância D^2 de Mahalanobis é maior do que 1% e que a maior distância de Cook é menor do que o valor limite mais conservador $[4/(20-3-1)] = 0,25$. Com isto descarta-se a presença de dados atípicos (que deveriam ser retirados da amostra). Em seguida, verificam-se quatro suposições acerca dos resíduos: (i) a normalidade; (ii) a independência em relação aos preditos; (iii) a autocorrelação; e (iv) a homocedasticidade da amostra. A primeira verificação é feita na Figura 3, o gráfico de dispersão de resíduos e preditos. O R^2 próximo a zero e o formato aleatório evidenciam a independência dos resíduos em relação aos preditos. A segunda verificação é apresentada na Tabela 6, na qual dois testes da qualidade do ajuste não rejeitam a hipótese de que os resíduos seguem uma distribuição normal. A terceira verificação é feita pelo teste de Durbin-Watson, apresentado na Tabela 7. Finalmente, a homocedasticidade é verificada pelo método de Pesaran-Pesaran, que faz uma regressão entre os resíduos quadráticos e os preditos: se a regressão não for significativa, há a homocedasticidade. A Tabela 8 apresenta valor- $p > 5\%$, que indica que a regressão não é significativa, confirmando a homocedasticidade.

Tabela 5: Análise de preditos e resíduos

	mínimo	máximo	média	desvio-padrão
valor predito	34.590	42.074	37.982	2.244
resíduo	-822	796	0,89	430
resíduo padronizado	-1,91 (obs. 11)	1,85 (obs. 10)	0,00	1,00
resíduo de Student	-2,08 (obs. 11)	2,00 (obs. 10)	0,01	1,05
significância de D^2	0,16 (obs. 20)	-	-	-
distância de Cook (D_i)	-	0,201 (obs. 9)	-	-

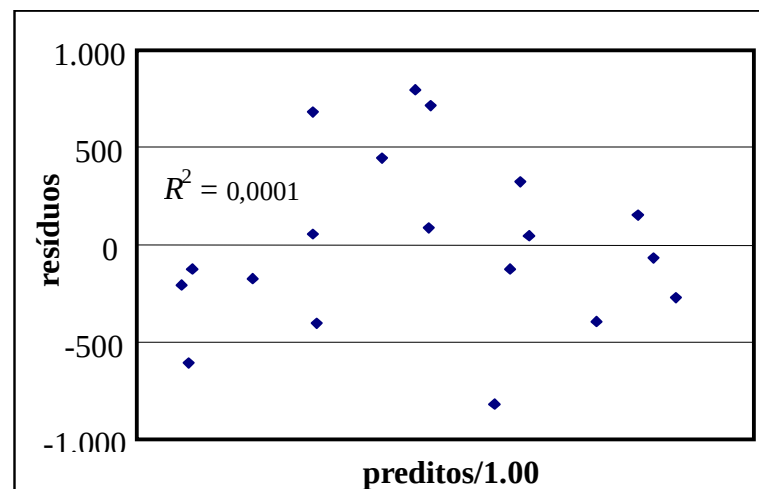
Figura 2: Dispersão (valores preditos X resíduos) e R^2

Tabela 6: Ajuste à normalidade dos resíduos

	teste KS	qui-quadrado	conclusão
Nível de significância para os resíduos	0,2331	0,2295	não rejeitar

Tabela 7: Teste de autocorrelação dos resíduos [20]

estatística DW calculada	limite inferior da área permitida	limite superior da área permitida	conclusão
2,127	1,828	2,172	não há autocorrelação

Tabela 8: Teste da homocedasticidade

Anova	gl	SQ	MQ	F teste	F crítico	Valor-p
Regressão	1	2,89 E+10	2,89 E+10	0,55	4,41	0,47
Resíduo	18	9,45 E+11	5,25 E+10			
Total	19	9,74 E+11				

5. Discussão

Inicia-se a discussão pela consistência do modelo que deu origem à regressão. Apesar do modelo apresentar elos de realimentação, foi possível identificar duas vertentes que contribuem para o requisito de potência elétrica do mercado brasileiro.

Uma vertente parece representar um fenômeno econômico e foi mais bem representado pelo indicador de atividade econômica, o PIB geral dessazonalizado. Outra vertente parece representar um fenômeno social, representado pela multiplicação do IDH

pela população total. A respeito deste fator multiplicativo, argumenta-se que o IDH pode ser considerado como uma densidade de requisito de potência: um cidadão que apresenta um alto IDH requer em média mais potência elétrica em sua residência do que um cidadão que possui IDH baixo. O requisito final de potência surge da multiplicação da densidade do requisito, o IDH, pela massa populacional, a população total. Como a observação inicial é diferente de zero, agregou-se um fator de escala. Retirou-se um grau de liberdade ao manter a soma dos dois fatores igual a um, permitindo que se verifique o percentual de contribuição de cada vertente.

Para cada uma das vinte observações, foi calculado o valor de cada parcela e verificada a relação entre as somas das parcelas. A análise indicou que o fenômeno social influencia 6,6 vezes mais do que o fenômeno econômico nas variações observadas no requisito de potência do mercado brasileiro. É possível que um programa institucional de racionalização do consumo de energia elétrica deva atentar mais ao fenômeno social do que ao econômico, ao menos aquele medido pelo PIB.

Considerando os elementos que formam o IDH, atenção pode ser dada aos entes: *aumenta o bem-estar nas moradias; aumenta a educação e a informação das pessoas; e aumenta a longevidade da população*. Os dois primeiros podem estar correlacionados. O aumento do bem-estar, da educação e da informação pode ser explicado pelo avanço do acesso a bens tecnológicos por que a sociedade brasileira tem passado desde 1994, com o fim do processo inflacionário. Na prática, o bem-estar resulta em maior consumo doméstico de energia, pela aquisição de eletrodomésticos e outras facilidades para as residências. Com mais eletrodomésticos, aumenta a educação e a informação. O terceiro ente também pode estar correlacionado aos dois primeiros, pois bem-estar e informação se relacionam com prevenção de doenças e de hábitos danosos à saúde que, aliados a políticas de saneamento e ao avanço da medicina, resulta em maior longevidade da população. A combinação de famílias que consomem mais energia e pessoas que vivem mais resulta em acréscimo por causas sociais no requisito de potência mais do que proporcional ao por aumento da atividade econômica.

Dado que avanços sociais são desejáveis e devem ser estimulados, uma política para contingenciar o aumento do requisito de potência poderia focar na racionalização do consumo de eletrodomésticos, por estímulo a projetos mais eficientes. Outra ação poderia estimular o uso mais eficiente de energia nas residências, ou por campanhas de conscientização para uso racional, ou por políticas de tarifas diferenciadas por horário ou faixa de consumo, como as que já existem para o setor industrial. Por fim, o desenvolvimento de medidores residenciais de energia mais inteligentes, que denunciasses e coibisses fraudes, também poderiam desestimular consumos irregulares verificados em áreas de difícil controle. Caso tais medidas, ou equivalentes, não sejam cogitadas, sobra, para evitar um novo apagão, o aumento da capacidade instalada, o que, além de demandar recursos, demanda tempo de maturação e não contribui para uma evolução social em direção à eficiência e uso racional da energia.

Quanto à qualidade do modelo, a regressão apresenta um coeficiente de determinação R^2 próximo a 1: menos de 4% da variação nas observações não é explicada pelo modelo. O erro-padrão da estimativa, calculado para dezesseis graus de liberdade, representa menos de 1% do fator de escala. O valor- p do teste F é próximo a zero. Com base nessas evidências, é possível considerar significativa a regressão. Para verificar se as premissas do modelo são cumpridas, inspecionam-se os resíduos. O ajuste dos resíduos à distribuição normal, verificada pelos testes do qui-quadrado e KS, é satisfatória a um nível de confiança superior a 95%. A normalidade dos resíduos é confirmada pelo gráfico de dispersão [valores preditos X resíduos], cujo R^2 é praticamente nulo. Tanto os resíduos padronizados como os resíduos de Student se mantiveram em uma faixa de ± 2 , o teste de Durbin-Watson foi positivo e o teste de Pesaran-Pesaran evidenciou que se trata de uma amostra de variância constante.

Pelos resultados dos testes, é possível concluir que os pressupostos exigidos pelo modelo foram atendidos. A Figura 4 apresenta uma comparação entre os valores observados e os valores preditos ao longo de vinte trimestres, na qual verifica-se que apenas algumas observações centrais se desviaram um pouco mais dos valores preditos.

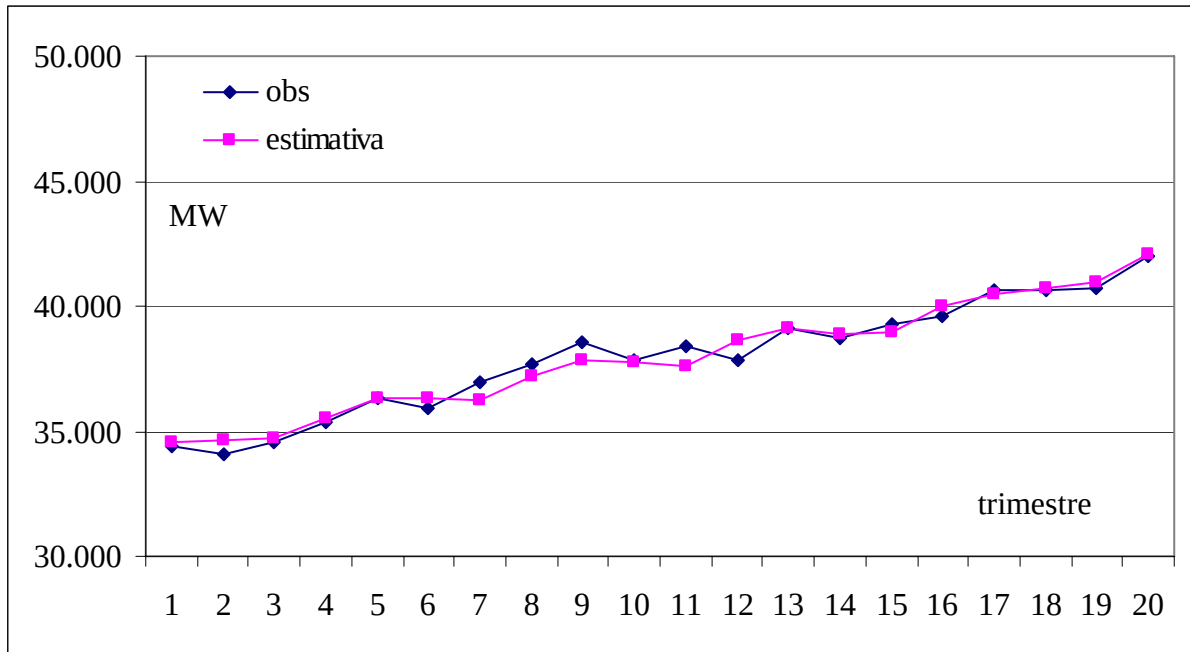


Figura 3: Comparação entre observações e predições

Uma observação é pertinente quanto ao universo de dados. Embora houvesse dados de janeiro de 1996 a novembro de 2005, julgou-se prudente usar apenas as observações entre 1996 e 2000 (não há informações mensais anteriores a 1996). Tomou-se esta decisão em virtude da descontinuidade observada a partir de março de 2001, o apagão. A descontinuidade foi tamanha que um modelo que tentasse explicá-la faria tantas concessões que se tornaria de pouca valia no restante do período (um polinômio de quarto grau explicou apenas 60% das variações). Metodologicamente também se perderia a oportunidade de se chegar a um modelo replicável caso se insistisse em explicar a totalidade dos dados.

Na Figura 5, observam-se séries temporais e retas médias que descrevem o requisito de potência elétrica do mercado brasileiro de janeiro de 1996 ($x = 1$) a março de 2001 e de fevereiro de 2002 a novembro de 2005 ($x = 119$). Na Figura 6, observa-se o comportamento anual complexivo, nos últimos vinte anos. Observa-se que, após quatro anos, o mercado ainda não voltou ao ritmo de consumo anterior, assumindo até agora um comportamento mais econômico. Os coeficientes angulares das retas (132 e 159) são próximos e podem sugerir que o crescimento natural do requisito tenha sido retomado. Se confirmada a mudança, esta pode ser atribuída a um novo comportamento do consumidor e ao lançamento por fabricantes de equipamentos mais eficientes. Integrando-se a diferença entre as retas médias das séries temporais de potência chega-se a uma estimativa de mais de 130.000 GWh de energia poupados entre fevereiro de 2003 e novembro de 2005.

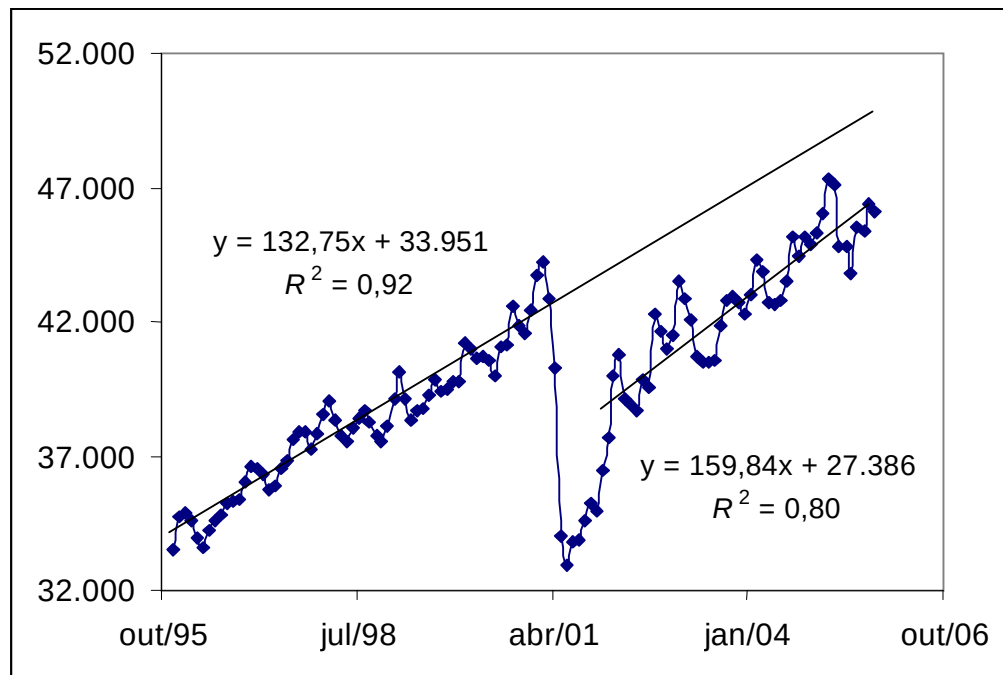


Figura 4: Séries temporais antes e depois do apagão [ONS, 2005]

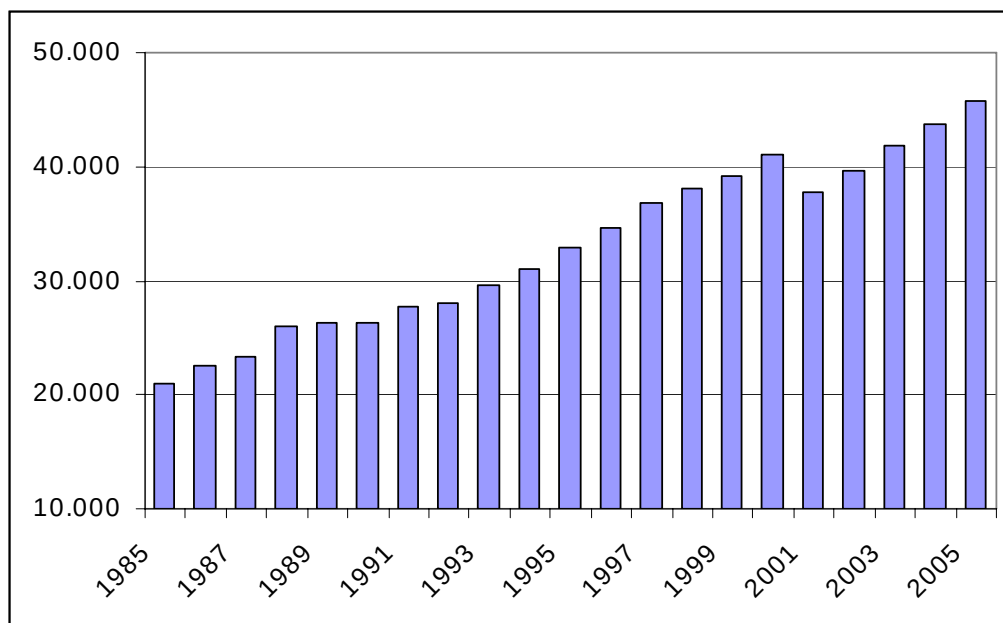


Figura 5: Comportamento anual da potência média, de 1985 a 2005 [ONS, 2005]

A seguir se discutem aspectos do método de pesquisa e do avanço que se pode esperar. A maior contribuição que esta pesquisa talvez possa trazer parece ser metodológica, pois

pode servir como sugestão para pesquisas equivalentes tanto na área da energia elétrica de regiões brasileiras ou de outras nações, como para outras formas de consumo energético.

O método empregado parece ter sido consistente, fácil de implementar e não suscitou dúvidas quando da sua aplicação. Os softwares foram de fácil aplicação, não exigindo investimentos adicionais em softwares especialistas e chegaram a resultados consistentes com as limitações da teoria atual, se bem que não se conseguiu tratar da estocasticidade dos regressores. Quanto ao avanço no assunto, entende-se que o ajuste obtido é satisfatório e o número de variáveis preditoras é baixo, o que atende premissas de precisão e de facilidade operacional. O resultado final parece ser eficaz na previsão de longo prazo (um ano ou mais) do requisito de potência do mercado brasileiro. Os preditores empregados apresentam um movimento lento, perceptível no longo prazo, coerente com o movimento da resposta, sempre crescente. Como não era objetivo de pesquisa a obtenção de um modelo para o curto prazo, não há necessidade de indicadores que apreendam variações imediatas, tais como as ligadas ao clima.

Uma última discussão acerca dos indicadores utilizados é possível, ao se compararem os mesmos com os indicadores usados em Iwamiya e Kermanshahi [2004], que relatam o uso de redes neurais para a previsão de longo prazo para o mercado japonês. Os pesquisadores usaram como fatores de aprendizado da rede: (1) o produto nacional bruto; (2) o produto interno bruto (para o Japão exportador, são muito diferentes); (3) população total; (4) número de domicílios; (5) número de equipamentos de ar condicionados; (6) total de poluição por CO₂; (7) índice de produção industrial; (8) preço de óleo combustível; (9) consumo total de energia, incluindo todas as fontes; e (10) preço da eletricidade. Treinada previamente e alimentada com os dados do ano T , a rede oferece a previsão para os anos $[T + i]$. Obviamente, quanto maior i , menor a confiabilidade da previsão. Observa-se que os fatores de treinamento guardam relação com os indicadores ora empregados, exceção feita a (6), (8) e (9), haja vista a dependência do Brasil da energia hídrica, o que torna o uso de óleo quase irrelevante. Os fatores (1), (2) e (7) se relacionam ao PIB; (3) e (4) se relacionam com a população total; e (5) e (10) se relacionam principalmente com o IDH, havendo multicolinearidade entre fatores. Os pesquisadores apontam um erro de 3% para o método e comentam que até 10% de erro seria aceitável em previsões de longo prazo para o mercado japonês.

O procedimento ora proposto também pode se valer de estimativas para o período T dos indicadores, fornecidas por ONS, IBGE e Banco Central do Brasil, e projetar o requisito de potência para os períodos $[T + i]$. À medida que novas observações vão surgindo, estas são incorporadas ao procedimento que, de modo recursivo, recalcula e robustece o modelo. Recorde-se que o modelo por regressão não-linear por ora atingiu um R^2 de 96,5%.

6. Conclusões

Por meio de uma abordagem que conectou SSM e regressão não-linear, chegou-se a um modelo para o requisito de potência do mercado elétrico brasileiro entre 1995 e 2001. O modelo a que se chegou faz mais do que apresentar boas estatísticas de teste: é calcado em um desenvolvimento qualitativo, que considerou visões e interesses conflitantes e gerou uma figura compartilhada por especialistas no tema.

O desenvolvimento do modelo tomou um tempo mais longo do que o previsto, devido às possibilidades de teste oferecidas pela teoria sobre a regressão múltipla. Para novas aplicações, pode-se agregar variantes no procedimento, tais como o teste do modelo a subconjuntos da amostra original e calcular a variabilidade dos parâmetros da regressão.

O ajuste, a facilidade de obtenção dos dados, a operacionalização dos cálculos e a parcimônia de preditores obtidos sugerem que é possível desenvolver um modelo não-linear de regressão múltipla que descreva o requisito de potência elétrica de um sistema consumidor. Ressalte-se que esta aplicação contou com um número baixo de observações e, por consequência, de graus de liberdade para o cálculo dos preditores, devendo-se destacar os aspectos metodológicos ora propostos.

A regressão múltipla não é a única técnica multivariada cabível ao caso. Outros pesquisadores têm usado redes neurais e lógica neuro-fuzzy. A modelagem por equações estruturadas (SEM) também pode ser aplicada, o que oferece um aspecto que a regressão múltipla não oferece: a possibilidade de a mesma variável ser ora resposta, ora preditor, em circularidade, o que foi observado no modelo deste trabalho. No entanto, a modelagem por equações estruturadas não parece ser indicada para o estágio atual da aplicação, pela pouca quantidade de dados por ora disponíveis.

Sugere-se como continuidade a inserção de mais aspectos sociais na análise. Pode-se inserir no modelo atual mais fatos, tais como a circularidade entre a população e a qualidade de vida, uma vez que se observa a tendência de redução no ritmo de aumento populacional em sociedades que aumentaram sua qualidade de vida. Finalmente sugere-se uma investigação comparada das facilidades oferecidas e da nomenclatura empregada pelos diversos softwares disponíveis para operacionalização da regressão múltipla, e eventual escolha de um deles para replicar o experimento, o que permitirá sobrepujar a limitação ora vigente de não se tratarem os regressores como parâmetros estocásticos.

Cabe uma última consideração. Não era objetivo de pesquisa investigar as consequências do apagão. No entanto chamou a atenção o efeito que este exerceu no mercado consumidor. Após a liberação, mesmo que a taxa de crescimento tenha voltado a um valor próximo do inicial, o nível de consumo ainda não voltou aos níveis que se deveriam esperar se o padrão de consumo anterior fosse mantido. Não foi feita uma investigação específica, mas é possível que um aprendizado tenha se dado quanto à conservação de energia. Trata-se de um assunto com desdobramentos sociais, haja vista as implicações econômicas e ambientais que daí decorrem. A redução do investimento na ampliação do sistema pode ser direcionada a outros campos sociais e a redução do consumo implica redução no uso do manancial hídrico e na agressão ambiental decorrente

7. Referências

- Al-Saba, T.; El-Amin, I. [1999] Artificial neural networks as applied to long-term demand forecasting. *Artificial Intelligence in Engineering*. Vol. 13, No 2, pp. 189-197.
- Bana e Costa, C.; Ensslin, L.; Correa, E.; Vansnick, J. [1999] Decision Support Systems in action: Integrated application in a multicriteria decision aid process. *European Journal of Operational Research*, Vol. 113, No 2, pp. 315-335.
- Belton, V.; Ackerman, F.; Sheperd, I. [1997] Integrated support from problem structuring through to alternative evaluation using COPE and V.I.S.A. *Journal of Multiple Criteria Decision Analysis*. Vol. 6, No 3, pp. 115-130.
- Bezerra, E.; Diwan, J. [2001] Uso de DEA como alternativa ao IDH na mensuração do desenvolvimento humano nos maiores municípios brasileiros, *Anais do XXXIII Simpósio Brasileiro de Pesquisa Operacional, SOBRAPO*, Campos do Jordão.
- Bunn, D. [2000] Forecasting loads and prices in competitive power markets. *Proceedings of the IEEE*, Vol. 88, No 2, pp. 163-169.

- Caio, L.; Berman, C. [1998] Análise das metodologias de previsão de mercado de energia elétrica face ao novo perfil de planejamento no ambiente pós-privatização. III Congresso Brasileiro de Planejamento Energético, Anais, S. Paulo.
- Checkland, P.; Scholes, J. [1998] Systems thinking, systems practice. John Wiley & Sons, Chichester.
- Comitê de Revitalização do Modelo do Setor Elétrico [2002] Relatório de progresso nº 2, fevereiro de 2002, disponível em <http://www.energiabrasil.gov.br/>, acesso em novembro de 2002.
- Corrar, L.; Theóphilo, C.; Bergmann, D. Regressões. [2004] In: Corrar, L.; Theóphilo, C. (org). Pesquisa operacional para decisão em contabilidade e administração, Atlas, S. Paulo.
- Da, X.; Jiangyan Y.; Jilai Y. [2000] The physical series algorithm of mid-long term load forecasting of power systems. Electric Power Systems Research, Vol. 53, No 1, pp. 31-37.
- Dettmer, W. [1997] Goldratt's theory of constraints, ASQ, Milwaukee.
- Draper, N.; Smith, H. [1981] Applied Regression Analysis, John Wiley & Sons, N. York.
- Eden, C. [1988] Cognitive maps: A review. European Journal of Operational Research. Vol. 36, No 1, pp. 1-13.
- Ensslin, L.; Montibeller, S. Noronha, S. [2001] Apoio à decisão, Insular, Florianópolis.
- Figueiredo, K.; Oliveira, F.; Pacheco, M.; Vellasco, M. [2005] Intelligent solutions to energy conservation problems. Núcleo de Inteligência Computacional Aplicada, PUC-Rio, Brazil. Disponível em http://www.ica.ele.puc-rio.br/publicacoes/download/cnf_0002.pdf. Acesso em dezembro de 2005.
- Goldratt, E. [1990] Theory of constraints, North River Press, Great Barrington.
- Hagan, M.; Behr, S. [1987] The time series approach to short term load forecasting. IEEE Transactions on Power Systems, Vol. PWRS-2, No 3, pp. 785-791.
- Hair, J.; Tatham, R.; Anderson, R.; Black, W. [1998] Multivariate Data Analysis, Prentice Hall, New Jersey.
- IBGE [2002] Instituto Brasileiro de Geografia e Estatística, disponível em <http://www.sidra.ibge.gov.br>, acesso em novembro de 2002.
- Iwamiya, H.; Kermanshahi, B. [2004] Long-term load forecasting using neural nets. Working paper. Tokyo University of Agriculture and Technology, disponível em http://www.tuat.ac.jp/~bahman/achivment/hiroshi_ICEE.PDF, acesso em dezembro de 2004.
- Jackson, M. [1993] Systemic methods in management sciences, Plenum Press, New York.
- Kandil, M.; El-Debeiky, S.; Hasanien, N. [2002] Long-term load forecasting for fast developing utility using a knowledge-based expert system. IEEE Transactions on Power Systems. Vol. 17, No 2, pp. 491-496.
- Kendall, G. [1998] Securing the future: strategies for exponential growth using the theory of constraints, St. Lucia Press, Boca Raton.
- Morgan, G. [1996] Imagens da organização, Atlas, S. Paulo.
- Milone, G.; Angelini, F. [1995] Estatística aplicada, Atlas, S. Paulo.
- Mutolsky, H. [2001] The GraphPad guide to nonlinear regression. GraphPad Software Inc., disponível em <http://www.graphpad.com>, acesso em novembro de 2001.
- Neves, L.; Martins, A.; Antunes, C.; Dias, L. [2004] Using SSM to rethink the analysis of energy efficiency initiatives. Journal of the Operational Research Society, Vol. 55. No 9, pp. 968-975.
- ONS [2005] Operador Nacional do Sistema, disponível em <http://www.ons.org.br>, acessos em novembro de 2002 e dezembro de 2005.

- Papalexopoulos, A.; Hesterberg, T. [1989] A regression-based approach to short-term system load forecasting. Conference Papers of Power Industry Computer Application Conference, pp. 414-423.
- Park, D.; El-Sharkawi, M.; Marks, R.; Atlas, L.; Damborg, M. [1991] Electric load forecasting using an artificial neural network. IEEE Transactions on Power Systems, Vol. 6, No 2, pp. 442-449.
- Parlos, A.; Oufi, E.; Muthusami, J.; Patton, A.; Atiya, A. [1996] Development of an intelligent long-term electric load forecasting system. Proceedings of the International Conference on Intelligent Systems Applications to Power Systems, pp. 288-292.
- Peng, T.; Hubele, N.; Karady, G. [1992] Advancement in the application of neural networks for short-term load forecasting. IEEE Transactions on Power Systems. Vol. 7, No 1, pp. 250-257.
- Rahman, S.; Hazim, O. [1993] A generalized knowledge-based short-term load-forecasting technique. IEEE Transactions on Power Systems, Vol. 8, No 2, pp. 508-514.
- Ribeiro, J.; Newmann, C. [2003] Planejamento e condução de grupos focados. In: Ribeiro, J. (org.) Grupos focados: teoria e aplicações.: FEENG-UFRGS-PPGEP, Porto Alegre.
- Sellitto, M. [2005] Processos de pensamento da TOC como alternativa para análise sistêmica em organizações: um estudo de caso em saúde pública. Gestão & Produção, Vol. 12, No 1, pp. 81-96.
- Sheinkopf, L. [1999] Thinking for a change: putting the TOC thinking process to use. St. Lucia Press, Boca Raton.
- Silva, P. [2004] Programação não-linear. In: Corrar, L.; Theóphilo, C. (org). Pesquisa operacional para decisão em contabilidade e administração, Atlas, S. Paulo.
- Silva, E. [2001] Formação de Preços em Mercados de Energia Elétrica, Sagra, Porto Alegre.
- UNDP [2002] United Nations Development Program, disponível em <http://www.undp.org.br>, acesso em novembro de 2002.
- Werkema, M.; Aguiar, S. [1996] Análise de Regressão, Fundação Cristiano Otoni, B. Horizonte.
- Wilson, B. [2001] Soft system methodology: conceptual model building and its contribution, John Wiley & Sons, Chichester.

Seleccção de variáveis em DEA aplicada a uma análise do mercado de energia eléctrica

Lidia Angulo Meza [†]
João Carlos Correia Baptista Soares de Mello [†]
Eliane Gonçalves Gomes [‡]
Artur José Silva Fernandes [†]

[†] Universidade Federal Fluminense
Brasil
lidia@metal.eeimvr.uff.br
jcsmello@producao.uff.br
<http://www.uff.br/decisao>
arth@predial.net.com.br

[‡] Empresa Brasileira de Pesquisa Agropecuária
Brasil
eliane.gomes@embrapa.br

Abstract

In this paper we have two main goals. From a theoretical point of view we intend to develop an improvement to the DEA variable selection models. It is in the variable normalisation step that this improvement takes place. From a practical point of view we intend to measure an energetic efficiency of the 27 Brazilian estates. A state achieves this efficiency when it has high values of some socio-economic indexes and low level of energy consumption. We used the non-parametrical approach of Data Envelopment Analysis (CCR and BCC models). We also used the inverted frontier DEA model.

Resumo

Este artigo tem duplo objectivo: Do ponto de vista teórico pretende-se mostrar um aperfeiçoamento aos métodos de selecção de variáveis em modelos DEA. Este ocorre na etapa de normalização das variáveis. Do ponto de vista da aplicação pretende-se medir uma eficiência energética dos 27 estados brasileiros. Esta eficiência é obtida quando os estados conseguem altos valores nos indicadores socioeconómicos com baixo consumo de energia. São usados os modelos DEA clássicos CCR e BCC e os modelos de fronteira invertida.

Keywords: Data Envelopment Analysis; Variables selection; Electrical Sector

Title: Variables selection in DEA applied to an analysis of the electric energy market

1 Introdução

Uma revisão não extensa da literatura mostra que existem estudos para prever e explicar o consumo de energia eléctrica. Entretanto, as análises quantitativas para explicar o comportamento do mercado nem sempre são adequadas. Por exemplo, Lins et al. (2002) propuseram o uso da metodologia de análise condicional de demanda para estimar o consumo de energia eléctrica do sector residencial brasileiro, em função dos aparelhos usados. Estudos como este obrigam ao conhecimento sobre os hábitos dos consumidores, como as quantidades e os tipos de aparelhos eléctricos usados nas residências. Essa abordagem necessita de um grande trabalho de campo, como aquele feito pela fonte dos dados do artigo citado, qual seja, uma pesquisa realizada pelo Programa Nacional de Conservação de Energia Eléctrica. Pesquisas deste tipo nem sempre são viáveis.

O objectivo prático deste estudo é obter uma forma não paramétrica de identificar os estados com maior eficiência energética. Isto significa obter bons resultados socioeconómicos com baixo consumo de energia eléctrica. Os indicadores socioeconómicos usados são Índice de Desenvolvimento Humano – IDH, Índice de Potencial de Consumo – IPC, Produto Interno Bruto per capita – PIB per capita. Como estes indicadores não são totalmente independentes, pretende-se usar um método de selecção de variáveis para determinar os mais adequados a entrar no modelo. O método a ser usado deverá ser quantitativo e objectivo, isto é, independente de opiniões do decisor. Uma revisão de métodos de selecção de variáveis pode ser vista em Senra (2004) e Senra et al. (2007). Neste artigo será proposta uma evolução do ‘Método Multicritério Combinatório por Cenários’ (Senra et al., 2007).

Para a construção da fronteira de eficiência, duas ópticas podem ser consideradas. A primeira julga como eficiente um estado onde, mesmo com baixos valores dos índices socioeconómicos, a empresa consegue vender uma grande quantidade de energia eléctrica. Esta é considerada a óptica da empresa (Lins et al., 2007; Senra et al., 2007).

A segunda óptica considera eficiente aquele estado que, mesmo com pequeno consumo de energia eléctrica, consegue altos índices socioeconómicos. Esta é a óptica da eficiência energética adoptada, entre outros, por Angulo Meza et al. (2006), Chauhan et al. (2006), Hu e Kao (2007). No trabalho de Utlu e Hepbasli (2007) é realizada uma análise das variáveis no ponto de vista técnico, em que, mais uma vez, a energia consumida deve ser a menor possível. Além disso, pode-se encontrar também uma ampla revisão bibliográfica de todos os enfoques empregados na avaliação e análise da eficiência no uso da energia.

A óptica da eficiência energética pode ser interpretada do ponto de vista ambiental: deseja-se obter os melhores índices de desenvolvimento com o menor gasto possível de energia e, como consequência, com a menor agressão possível ao meio ambiente.

Devido ao objectivo proposto neste artigo, a óptica aqui usada é a da eficiência energética.

2 Análise Envoltória de Dados

2.1 Generalidades

A abordagem por Análise Envoltória de Dados (*Data Envelopment Analysis* – DEA) foi desenvolvida por Charnes et al. (1978) e usa a programação linear para avaliação de medidas de eficiência comparativas de Unidades de Tomada de Decisão (*Decision Making*

Units – DMUs) que utilizam os mesmos recursos (*inputs*) e geram os mesmos produtos (*outputs*).

Há dois modelos DEA clássicos: CCR e BCC. O modelo CCR (também conhecido por CRS ou *constant returns to scale*), adota como hipótese retornos constantes de escala (Charnes et al., 1978). Em sua formulação matemática considera-se que cada DMU k ($k = 1 \dots s$) é uma unidade de produção que utiliza n *inputs* x_{ik} , $i = 1 \dots n$, para produzir m *outputs* y_{jk} , $j = 1 \dots m$.

O modelo DEA CCR maximiza o quociente entre a combinação linear dos *outputs* e a combinação linear dos *inputs*, com a restrição de que para qualquer DMU esse quociente não pode ser maior que 1. Esse problema de programação fraccionária, mediante alguns artifícios matemáticos, pode ser linearizado e transformado no Problema de Programação Linear (PPL) apresentado em (1), no qual h_o é a eficiência da DMU o em análise; x_{io} e y_{jo} são os *inputs* e *outputs* da DMU o ; v_i e u_j são os pesos calculados pelo modelo para *inputs* e *outputs*, respectivamente.

$$\begin{aligned} \max \quad & h_o = \sum_{j=1}^m u_j y_{jo} \\ \text{sujeito a} \quad & \\ & \sum_{i=1}^n v_i x_{io} = 1 \\ & \sum_{j=1}^m u_j y_{jk} - \sum_{i=1}^n v_i x_{ik} \leq 0, \quad k = 1, \dots, s \\ & u_j, v_i \geq 0 \quad \forall i, j \end{aligned} \tag{1}$$

O modelo BCC (Banker et al., 1984), também chamado de VRS (*variable returns to scale*), considera situações de eficiência de produção com variação de escala e não assume proporcionalidade entre *inputs* e *outputs*. A formulação do modelo BCC usa para cada DMU o problema de programação linear (PPL) apresentado em (2).

$$\begin{aligned} \max \quad & h_o = \sum_{j=1}^s u_j y_{jo} - u_* \\ \text{sujeito a} \quad & \\ & \sum_{i=1}^m v_i x_{io} = 1 \\ & \sum_{j=1}^s u_j y_{jk} - \sum_{i=1}^n v_i x_{ik} - u_* \leq 0, \quad k = 1, \dots, n \\ & u_j, v_i \geq 0 \quad \forall x, y \\ & u_* \in \Re \end{aligned} \tag{2}$$

Neste modelo, para a DMU o em análise, a eficiência é dada por h_o ; x_{ik} representa o *input* i da DMU k ; y_{jk} representa o *output* j da DMU k ; v_i e u_j representam os pesos dados aos *inputs* i e aos *outputs* j , respectivamente; u_* é um factor de escala (quando positivo, indica que a DMU está em região de retornos decrescentes de escala; se negativo,

os retornos de escala são crescentes). Se h_o é igual a 1, a DMU o em análise é considerada eficiente. No PPL (2) as variáveis de decisão são v_i , u_j e h_o . De forma não matemática, no modelo BCC uma DMU é eficiente se, na escala em que opera, é a que melhor aproveita os *inputs* de que dispõe. Já no modelo CCR, uma DMU é eficiente quando apresenta o melhor quociente de *outputs* com relação aos *inputs*, ou seja, aproveita melhor os *inputs* sem considerar a escala de operação da DMU.

2.2 Fronteira invertida

Os modelos DEA, devido a produzirem avaliações benevolentes, geralmente resultam em um grande número de DMUs eficientes. Além disso, devido às características matemáticas do modelo BCC, as DMUs que têm o menor valor de um dos *inputs* ou o maior valor de um dos *outputs* são consideradas eficientes, mesmo que as relações com as demais variáveis não sejam as melhores (Ali, 1993). Essas DMUs são chamadas de falsamente eficientes ou eficientes à partida.

Existem vários métodos para melhorar a discriminação em modelos DEA (Adler et al., 2002; Angulo Meza e Lins, 2002; Leta et al., 2005). Um desses métodos, que tem a vantagem adicional de eliminar as DMUs falsamente eficientes no modelo BCC, é a fronteira invertida. Este tipo de fronteira, baseada na inversão de *inputs* e *outputs*, é apresentada em Yamada et al. (1994), Entani et al. (2002), Lins et al. (2005).

O uso da fronteira invertida como ferramenta para aumentar a discriminação entre as DMUs pode ser visto em Pimenta et al. (2004), Angulo Meza et al. (2005), Leta et al. (2005), Soares de Mello et al. (2005, 2006), Gomes et al. (2006). A fronteira invertida com este objectivo é utilizada na construção de um índice chamado de índice de eficiência composta. Ele é definido como sendo a média aritmética entre a eficiência em relação à fronteira DEA convencional (padrão) e o complemento da eficiência em relação à fronteira invertida, tal como visto em (3). O uso do complemento justifica-se pelo fato de a fronteira invertida gerar uma medida de ineficiência.

$$eficiência\ composta = \frac{[eficiência\ padrão + (1 - eficiência\ invertida)]}{2} \quad (3)$$

Para se obter um índice em que as unidades eficientes têm o valor de 1, é feita a normalização da eficiência composta, ao dividirem-se seus valores pela maior de todas as medidas de eficiência compostas.

3 Selecção de Variáveis em DEA

Na literatura DEA, pouco tem sido discutido sobre o problema da selecção das variáveis usadas na modelação. A grande maioria dos trabalhos publicados traz a abordagem da selecção de variáveis segundo a opinião de especialistas ou até mesmo da disponibilidade de dados. Thanassoulis (1996) ressaltou que a modificação do conjunto de variáveis seleccionadas poderá ter grande impacto no resultado da avaliação. Sendo assim, torna-se extremamente relevante discutir e comparar métodos de selecção de variáveis em DEA. No entanto, antes de apresentar qualquer método deve-se atentar para o que significa realmente a selecção de variáveis.

Deve-se ter em conta que o facto de uma diferente escolha de variáveis conduzir a resultados diferentes não deve ser interpretado como uma fraqueza de DEA. Na verdade, escolher variáveis diferentes significa que se pretende levar em conta uma dimensão diferente do problema, ou seja, olhar para as DMUs segundo outro ponto de vista. Assim, por exemplo, em Soares de Mello et al. (2003) diferentes escolhas dos conjuntos de variáveis levaram a 3 diferentes avaliações de empresas aéreas: operacional, de vendas e total. A análise prévia de possíveis conjuntos de variáveis é uma etapa fundamental a ser feita em conjunto por decisores, especialistas e analistas. Só após esta escolha prévia é que faz sentido pensar-se em métodos de selecção.

Caso não seja desejável usar modelos avançados, nem seja possível aumentar o número de DMUs, uma das opções é restringir as variáveis que vão entrar no modelo. Neste contexto, os métodos de selecção de variáveis devem ser vistos como instrumentos de auxílio à decisão, que orientarão a escolha final. Esta não deve ficar presa ao resultado de um modelo matemático, por mais sofisticado que seja. Sempre deve ser feita em conjunto pelos agentes de decisão, especialistas e analistas, que poderão (ou não) usar um método de selecção como ferramenta.

Os próprios métodos de selecção já existentes e o que será aqui proposto permitem diferentes graus de intervenção do agente de decisão. Aqueles que exigem pouca ou nenhuma informação subjectiva são mais adequados a decisores que estejam em grande dúvida e não consigam emitir opiniões. Outros, que exigem alguma escolha, devem ser usados por decisores que queiram impor alguma direcção ao método, sem no entanto serem capazes sozinhos de realizar uma escolha de variáveis. Em qualquer método usado, o decisor deve ser confrontado com o resultado e verificar a sua coerência. Por exemplo, deve verificar se não foi omitida nenhuma variável que considere imprescindível, e se há relação causal entre todos os pares *input-output* seleccionados. Pode ainda ser recomendável fazer a selecção por mais de um método para comparar resultados e tomar a decisão com mais conhecimento do assunto.

Um método de selecção de variáveis foi proposto por Norman e Stoker (1991), no qual aliaram a análise de correlação simples à análise envoltória de dados. Os autores propuseram um procedimento iterativo que simultaneamente identifica os *inputs* e *outputs* relevantes e calcula medidas de ineficiência dos planos de operação observados.

Já Golany e Roll (1989) enunciaram-se três estágios do processo de selecção de variáveis: (1) '*Judgmental screening*', (2) '*Non-DEA quantitative analysis*' e (3) '*DEA based analysis*'. O primeiro refere-se à distinção entre variáveis determinantes de eficiência e variáveis explicativas de ineficiência, que pode ser resolvida através de análise de causalidade. O segundo estágio sugere o uso de análise de regressão para determinar se uma variável deve ser *input* ou *output*. O terceiro estágio, sugerido por Charnes et al. (1978), pode ser usado para identificar as variáveis que deverão ser excluídas, ou seja, são aquelas que não agregam eficiência significativa ao modelo.

Em Lins e Moreira (1999) e Lins e Angulo Meza (2000) é proposto um método baseado apenas na relação causal entre *inputs* e *outputs*. Este método, chamado de 'I-O Stepwise', não é eficaz para melhorar o poder de discriminação dos modelos DEA, apesar de conduzir a modelos com forte relação causal. Em Soares de Mello et al. (2004) é proposto um modelo que leva em conta tanto a relação causal (medida pela eficiência média) quanto o poder de discriminação (medido pelo número de DMUs na fronteira eficiente), denominado 'Método Multicritério de Selecção de Variáveis em DEA'. Em Senra (2004) e Senra et al. (2007) foram apresentados modelos baseados no de Soares de Mello et al. (2004), na qual é retirada parte da subjectividade anterior.

Todos estes métodos são baseados na inserção gradativa de variáveis no modelo, sendo o ponto de paragem inteiramente arbitrário. Senra (2004) e Senra et al. (2007)

consideram que os métodos de inserção gradativa são heurísticas de um método óptimo. Este método deve considerar todas as possíveis combinações de *outputs* e *inputs*, e escolher o conjunto de variáveis com o qual seja obtido o melhor compromisso entre relação causal e poder de discriminação. No entanto, este método é, em muitos casos, extremamente moroso.

Em Senra et al. (2007) foi proposta uma forma de escolha de variáveis que reduz a morosidade do processo. O método chamado 'Multicritério Combinatório por Cenários' é um método em duas fases. Na primeira fase são construídos cenários que serão analisados na segunda. Chama-se 'cenário da primeira fase' ao melhor modelo com duas, três, quatro etc., variáveis. A segunda fase limita-se à escolha do melhor cenário e termina quando todas as variáveis tiverem sido incluídas no modelo.

O método pode ser descrito pelas Figuras 1 e 2, onde S_{EF} é a normalização das eficiências médias para cada conjunto de variáveis, e S_{DIS} é a normalização da quantidade de DMUs eficientes para cada conjunto de variáveis. As formas de normalização utilizadas podem ser vistas em Soares de Mello et al. (2002), Soares de Mello et al. (2004), Senra et al. (2007). O indicador S é a média aritmética entre S_{EF} e S_{DIS} .

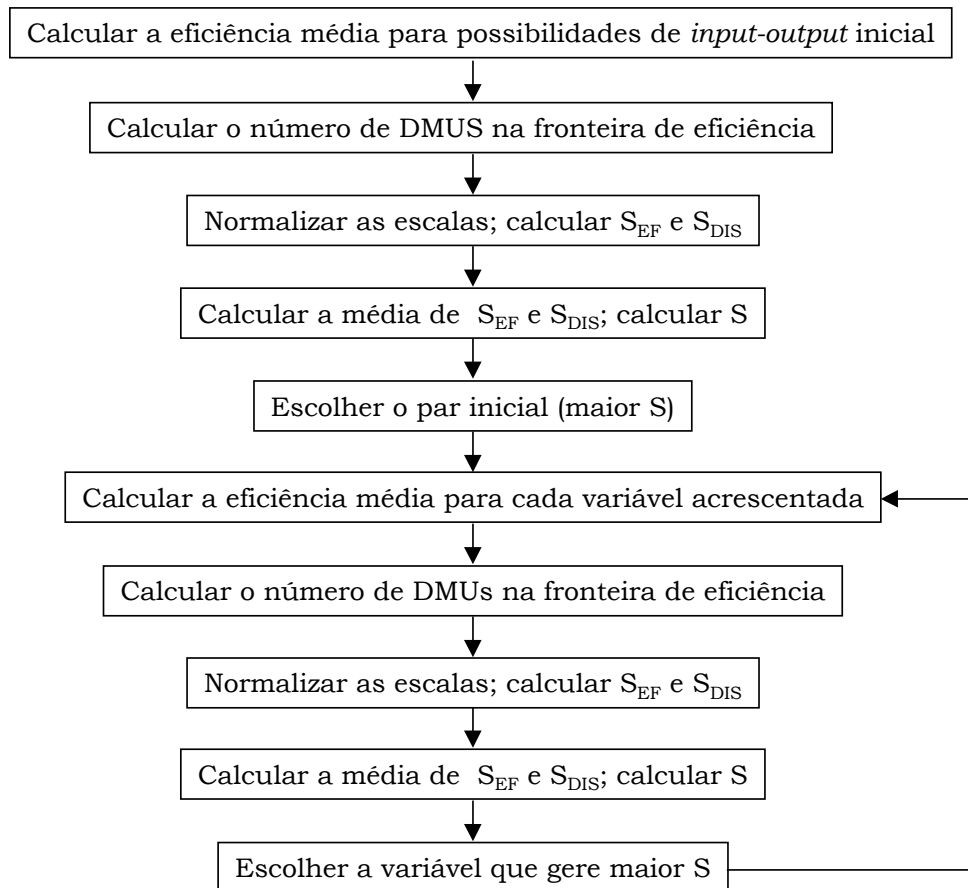


Figura 1: 1ª fase do 'Método Multicritério Combinatório por Cenários' de selecção de variáveis

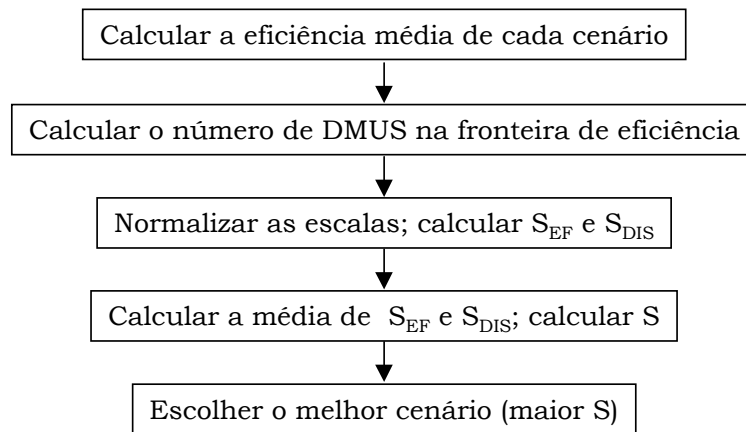


Figura 2: 2ª fase do 'Método Multicritério Combinatório por Cenários' de selecção de variáveis

Este modelo, ao contrário de outros modelos da literatura, não impõe um ponto de paragem arbitrário. Os modelos que impõem este ponto de paragem são, por sua natureza, úteis apenas quando o número de variáveis é grande comparado ao número de DMUs. Já este modelo e o modelo que dele será derivado permitem escolher não só as variáveis como o número delas a entrar no modelo. Assim, eles podem ser úteis mesmo quando há poucas variáveis e muitas DMUs.

Por outro lado, este método tem o grande inconveniente de fazer várias normalizações sucessivas com diferentes valores de referência. Esta característica pode fazer com que a combinação ideal de variáveis seja examinada, mas não escolhida.

Cabe aqui ainda mencionar o modelo de selecção de variáveis descrito por Wagner e Shimshak (2007). Os autores propuseram um procedimento tipo *stepwise* inverso, que envolve a maximização (ou minimização) sequencial da mudança média na eficiência conforme variáveis são retiradas do modelo. O algoritmo proposto pelos autores inicia com um modelo composto por todas as possíveis variáveis de *input* e *output*. O ponto de paragem é atingido quando o modelo é reduzido a uma variável de *input* e uma de *output*. Este é um modelo que só leva em conta o ajuste à fronteira. Ignora, dessa forma, o aumento de discriminação.

3.1 Método de selecção de variáveis proposto

Para evitar o problema das várias normalizações do método de selecção de variáveis 'Multicritério Combinatório por Cenários', propõe-se neste artigo um método alternativo que faz uso de normalização única. O método de normalização usado anteriormente considera uma eficiência média normalizada (S_{EF}) que atinge o valor 1 na eficiência máxima e zero na eficiência mínima. Já o poder de discriminação normalizado (S_{DIS}) apresenta o valor 1 quando há o menor número de DMUs na fronteira e zero para o maior número, para cada conjunto de variáveis usado.

Ao contrário do método anterior, a normalização aqui proposta não considera os valores máximos e mínimos efectivamente atingidos para a eficiência e o número de DMUs na fronteira para cada quantidade de variáveis. Ele leva em conta os valores extremos teoricamente alcançáveis. O número de DMUs na fronteira está no intervalo $[1, n]$ (sendo n o número total de DMUs) e a eficiência média está no intervalo $[0, 1]$. Para obter S_{EF} deve-se levar em conta que quanto maior a eficiência média, maior deve ser o

valor dessa variável. Assim, para obter a normalização faz-se $S_{EF} = \text{eficiência média}/10$. Para obter o S_{DIS} deve-se considerar que quanto maior o número de DMUs na fronteira, menor é o valor da variável. Desta forma, para a sua normalização faz-se $S_{DIS} = 10 \frac{n - N}{n - 1}$, onde n é o número total de DMUs e N é o número de DMUs na fronteira.

Esta normalização pode ser usada em cada etapa do método para escolher a próxima variável a ser inserida, ou, caso o número de variáveis não seja grande, calculada para todas as combinações possíveis de *inputs* e *outputs*. O conjunto final de variáveis a ser escolhido é aquele que apresentar o maior índice $S = S_{EF} + S_{DIS}$, que é uma medida de compromisso entre o bom ajuste à fronteira e a boa discriminação do modelo.

4 Modelação e Resultados

Para avaliação do uso da energia eléctrica no Brasil foram seleccionados como DMUs os 27 estados brasileiros. Como possíveis variáveis a serem escolhidas no modelo final tem-se: Consumo Residencial per capita (medido em GWh/habitante); Produto Interno Bruto (PIB) estadual (medido em US\$/habitante, ponderado pelo poder de compra); Índice de Potencial de Consumo (IPC – índice adimensional); Índice de Desenvolvimento Humano (IDH – índice adimensional). Dado o enfoque de eficiência energética já mencionado previamente, o consumo é um *input* e as demais variáveis são candidatas a *outputs*, em uma abordagem semelhante às de Angulo Meza et al. (2006), Chauhan et al., (2006), Hu e Kao (2007).

Os dados relacionados ao consumo de energia eléctrica foram obtidos do Ministério de Minas e Energia do Brasil. O IDH e o PIB foram obtidos do Instituto de Pesquisas Económicas e Aplicadas do país e o IPC da publicação nacional *Gazeta Mercantil*. Todos os dados referem-se ao ano de 1996.

O IPC, desenvolvido pela Federação do Comércio do Estado de São Paulo em parceria com a Fundação Getúlio Vargas, mostra o perfil de consumo mensal das famílias brasileiras (Borges, 2004). Esse índice considera faixas de renda, gastos (que incluem impostos e tarifas públicas, além dos itens habituais de consumo) e variáveis socioeconómicas.

O conceito de desenvolvimento humano é a base do IDH e parte do pressuposto de que para aferir o avanço de uma população não se deve considerar apenas a dimensão económica, mas também outras características sociais, culturais e políticas que influenciam a qualidade da vida humana (Programa das Nações Unidas para o Desenvolvimento, 2004). Além de computar o PIB per capita, o IDH também considera os componentes longevidade e educação. O primeiro utiliza números de expectativa de vida ao nascer; o segundo é avaliado pelo índice de analfabetismo e pela taxa de matrícula em todos os níveis de ensino. Essas três dimensões têm a mesma importância na composição do índice, que varia de zero a um.

O PIB para cada estado é calculado a partir de estatísticas sobre o valor anual da produção, consumo intermediário e valor adicionado de cada actividade económica. A metodologia de cálculo pode ser encontrada em Instituto Brasileiro de Geografia e Estatística (2002). É importante destacar que tanto o Consumo Residencial quanto o PIB foram divididos pela população para normalizar as variáveis.

Foi usado o método proposto neste artigo e descrito na secção anterior para selecção de variáveis, chamado de 'Método Compensatório de Normalização Única'. Devido ao pequeno número de variáveis iniciais foram testadas todas as combinações possíveis, o

que caracteriza o Método Total (Senra, 2004; Senra et al., 2007). Os dados usados encontram-se na Tabela 1.

Tabela 1: Dados socioeconômicos e de consumo residencial de energia eléctrica por estado brasileiro

DMU	<i>Input</i>		<i>Outputs</i>	
	Consumo residencial per capita (GWh/hab x 10 ⁻⁴)	PIB (U\$/hab.)	IPC	IDH
Acre	2,404	5.741	0,200	0,754
Alagoas	2,090	2.496	0,950	0,538
Amapá	3,131	5.370	0,170	0,786
Amazonas	2,739	5.718	1,170	0,775
Bahia	1,940	3.677	4,770	0,655
Ceará	1,998	2.667	2,400	0,590
Distrito Federal	5,692	14.854	2,230	0,869
Espírito Santo	4,222	6.251	1,450	0,836
Goiás	3,694	5.238	2,600	0,786
Maranhão	1,359	2.158	1,350	0,547
Mato Grosso	3,747	5.003	1,200	0,767
Mato Grosso do Sul	4,269	6.410	1,230	0,848
Minas Gerais	3,778	5.968	8,510	0,823
Pará	1,850	4.268	2,240	0,703
Paraíba	1,984	2.438	1,100	0,557
Paraná	4,074	6.485	5,620	0,847
Pernambuco	2,681	3.213	3,450	0,615
Piauí	1,714	2.004	0,750	0,534
Rio de Janeiro	6,196	8.653	11,630	0,844
Rio Grande do Norte	2,375	4.083	1,030	0,668
Rio Grande do Sul	4,520	7.395	7,080	0,869
Rondônia	2,816	6.448	0,580	0,820
Roraima	3,023	6.231	0,130	0,818
Santa Catarina	4,541	6.519	3,060	0,863
São Paulo	6,415	10.536	34,070	0,868
Sergipe	2,383	5.122	0,660	0,731
Tocantins	1,818	1.575	0,390	0,587

Foi usado o modelo CCR e, posteriormente, o modelo BCC, ambos orientados a *input* (modelos (1) e (2), respectivamente), o que caracteriza a eficiência energética do ponto de vista ambiental mencionado anteriormente. Isto significa que o objectivo é diminuir o consumo de energia eléctrica per capita, enquanto são mantidos os outros índices.

Os resultados para a selecção de variáveis com o uso do modelo CCR orientado a *input* são apresentados na Tabela 2. Verifica-se que o maior índice S, 16,8050, é obtido com o uso de três variáveis, Consumo per capita, PIB e IDH. No entanto, é interessante verificar que a substituição da variável IDH pela variável IPC fornece um índice S, 16,5090, muito próximo ao do modelo anterior. É importante notar que existe uma certa redundância no uso dos índices IDH e PIB, já que este é um componente do primeiro. Assim, pode ser vantajoso considerar um modelo com um índice S mais baixo, mas que evite a redundância das duas variáveis no modelo DEA.

Tabela 2: Selecção de variáveis com o 'Método Compensatório de Normalização Única' e modelo DEA CCR

DMU	Consumo e PIB	Consumo e IPC	Consumo e IDH	Consumo, PIB e IPC	Consumo, PIB e IDH	Consumo, IPC e IDH	Consumo, PIB, IDH e IPC
Acre	0,9150	0,0157	0,7789	0,9150	0,9974	0,7789	0,9974
Alagoas	0,4576	0,0856	0,6393	0,4777	0,6517	0,6393	0,6517
Amapá	0,6572	0,0102	0,6234	0,6572	0,7285	0,6234	0,7285
Amazonas	0,8001	0,0804	0,7029	0,8084	0,8763	0,7029	0,8763
Bahia	0,7263	0,4630	0,8385	0,8854	0,8778	1,0000	1,0000
Ceará	0,5115	0,2262	0,7334	0,5848	0,7452	0,7803	0,7806
Distrito Federal	1,0000	0,0738	0,3792	1,0000	1,0000	0,3807	1,0000
Espírito Santo	0,5673	0,0647	0,4918	0,5762	0,6201	0,4918	0,6201
Goiás	0,5434	0,1325	0,5285	0,5793	0,6049	0,5463	0,6072
Maranhão	0,6087	0,1871	1,0000	0,6640	1,0000	1,0000	1,0000
Mato Grosso	0,5117	0,0603	0,5084	0,5205	0,5716	0,5084	0,5716
Mato Grosso do Sul	0,5754	0,0543	0,4933	0,5799	0,6278	0,4933	0,6278
Minas Gerais	0,6054	0,4241	0,5411	0,7529	0,6648	0,7375	0,7781
Pará	0,8840	0,2280	0,9437	0,9473	1,0000	0,9708	1,0000
Paraíba	0,4708	0,1044	0,6971	0,4979	0,7056	0,6971	0,7056
Paraná	0,6100	0,2597	0,5164	0,6935	0,6643	0,6024	0,7169
Pernambuco	0,4592	0,2423	0,5697	0,5402	0,5902	0,6413	0,6431
Piauí	0,4480	0,0824	0,7737	0,4672	0,7737	0,7737	0,7737
Rio de Janeiro	0,5352	0,3534	0,3383	0,6572	0,5610	0,5261	0,6664
Rio Grande do Norte	0,6588	0,0817	0,6986	0,6717	0,7444	0,6986	0,7444
Rio Grande do Sul	0,6270	0,2950	0,4775	0,7237	0,6727	0,5925	0,7428
Rondônia	0,8774	0,0388	0,7232	0,8774	0,9519	0,7232	0,9519
Roraima	0,7898	0,0081	0,6720	0,7898	0,8608	0,6720	0,8608
Santa Catarina	0,5501	0,1269	0,4720	0,5836	0,6003	0,4923	0,6063
São Paulo	0,6294	1,0000	0,3361	1,0000	0,6480	1,0000	1,0000
Sergipe	0,8238	0,0522	0,7620	0,8238	0,9095	0,7620	0,9095
Tocantins	0,3320	0,0404	0,8020	0,3382	0,8020	0,8020	0,8020
Eficiência média	0,6361	0,1774	0,6311	0,6894	0,7574	0,6903	0,7912
N.º DMUs eficientes	1	1	1	2	3	3	5
SEFF	6,3611	1,7744	6,3114	6,8936	7,5742	6,9025	7,9120
SDIS	10,0000	10,0000	10,0000	9,6154	9,2308	9,2308	8,4615
S	16,3611	11,7744	16,3114	16,5090	16,8050	16,1333	16,3736

Apesar de não existirem mais efeitos de escala devido ao uso de variáveis per capita, foi testado o modelo BCC pelo facto de propiciar um envelopamento mais completo, sem a exigência de proporcionalidade entre *inputs* e *outputs*. Os resultados do modelo BCC orientado a *input* são apresentados na Tabela 3. Pode-se notar que, para o modelo BCC, o melhor resultado foi obtido para um modelo com duas variáveis, Consumo Residencial per capita e IDH, com um índice S de 17,0250.

Este modelo de apenas de um *input* e um *output* apresenta a vantagem de evitar a redundância mencionada no modelo CCR, ou seja, considerar apenas PIB quando incluído no IDH sem o uso explícito. Além disso, a escolha de variáveis do modelo BCC coincidiu com a opinião de especialistas do sector eléctrico que foram ouvidos. Por estes motivos, conclui-se então que, neste caso, o modelo BCC é preferível ao modelo CCR.

Maranhão é eficiente por ser o estado que apresenta menor consumo de energia residencial per capita. No entanto, por ser eficiente também no modelo CCR (tanto a que considera as mesmas duas variáveis, quanto no que foi escolhido pelo método de selecção de variáveis) ela é verdadeiramente eficiente.

O estado de Rio Grande do Sul apresenta o maior IDH. Observa-se na Tabela 3 que no modelo DEA em que seja desconsiderada essa variável, Rio Grande do Sul passa a ser ineficiente. Portanto, esta DMU deve ser considerada como falsa eficiente.

Tabela 3: Selecção de variáveis com o 'Método Compensatório de Normalização Única' e modelo DEA BCC

DMU	Consumo e PIB	Consumo e IPC	Consumo e IDH	Consumo, PIB e IPC	Consumo, PIB e IDH	Consumo, IPC e IDH	Consumo, PIB, IDH e IPC
Acre	0,9919	0,5651	0,9447	0,9919	1,0000	0,9447	1,0000
Alagoas	0,6876	0,6500	0,6500	0,6876	0,6876	0,6500	0,6876
Amapá	0,7186	0,4339	0,8097	0,7186	0,8097	0,8097	0,8097
Amazonas	0,8678	0,4961	0,8927	0,8678	0,9027	0,8927	0,9049
Bahia	0,8827	0,9727	0,8757	1,0000	0,8827	1,0000	1,0000
Ceará	0,7393	0,7612	0,7478	0,7735	0,7478	0,7824	0,7824
Distrito Federal	1,0000	0,2626	0,7940	1,0000	1,0000	0,7940	1,0000
Espirito Santo	0,6087	0,3254	0,7987	0,6087	0,7987	0,7987	0,7987
Goiás	0,5962	0,4201	0,6864	0,6032	0,6864	0,7254	0,7254
Maranhão	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Mato Grosso	0,5650	0,3626	0,6348	0,5650	0,6348	0,6348	0,6348
Mato Grosso do Sul	0,6155	0,3182	0,8876	0,6155	0,8876	0,8876	0,8876
Minas Gerais	0,6531	0,6525	0,7730	0,7725	0,7730	0,9498	0,9498
Pará	1,0000	0,8086	1,0000	1,0000	1,0000	1,0000	1,0000
Paraíba	0,7175	0,6846	0,7005	0,7175	0,7175	0,7005	0,7175
Paraná	0,6516	0,4954	0,9217	0,7114	0,9217	0,9471	0,9471
Pernambuco	0,5984	0,6278	0,5866	0,6489	0,5984	0,6416	0,6489
Piauí	0,7925	0,7925	0,7925	0,7925	0,7925	0,7925	0,7925
Rio de Janeiro	0,5555	0,4757	0,5892	0,6645	0,6333	0,6795	0,7047
Rio Grande do Norte	0,7609	0,5721	0,7326	0,7609	0,7609	0,7326	0,7609
Rio Grande do Sul	0,6605	0,4965	1,0000	0,7375	1,0000	1,0000	1,0000
Rondônia	0,9379	0,4824	1,0000	0,9379	1,0000	1,0000	1,0000
Roraima	0,8476	0,4494	0,9260	0,8476	0,9260	0,9260	0,9260
Santa Catarina	0,5873	0,3574	0,9493	0,6003	0,9493	0,9493	0,9493
São Paulo	0,6430	1,0000	0,6991	1,0000	0,7766	1,0000	1,0000
Sergipe	0,9066	0,5702	0,8735	0,9066	0,9109	0,8735	0,9109
Tocantins	0,7474	0,7474	0,8167	0,7474	0,8167	0,8167	0,8167
Eficiência média	0,7531	0,5845	0,8179	0,7881	0,8376	0,8492	0,8650
N.º DMUs eficientes	3	2	4	5	6	6	8
SEFF	7,5307	5,8445	8,1788	7,8805	8,3759	8,4923	8,6502
S _{DIS}	9,2308	9,6154	8,8462	8,4615	8,0769	8,0769	7,3077
S	16,7615	15,4598	17,0250	16,3420	16,4528	16,5692	15,9579

Na Tabela 4 apresenta-se o índice de eficiência com a aplicação da fronteira invertida para o modelo composto das variáveis Consumo e IDH, obtida pelo *software* SIAD (Angulo Meza et al., 2005). É interessante ressaltar que esta fronteira corresponde, no presente caso, à óptica da empresa fornecedora de energia eléctrica, tal como mencionado na secção 1. Este índice permite ordenar as DMUs e evitar alguns dos problemas clássicos de DEA, como DMUs eficientes ao considerar apenas um conjunto restrito de variáveis e DMUs eficientes à partida.

Tal como pode-se observar, os estados com melhor índice estão situados na região amazónica. Isso pode ser explicado pelo facto de existir um grande número de comunidades isoladas, as quais ou não usam energia eléctrica ou geram a sua própria energia sem recorrer às empresas distribuidoras.

Tabela 4: Resultados que consideram a fronteira invertida

DMU	Eficiência BCC padrão	Eficiência BCC invertida	Índice composto normalizado
Acre	0,9447	0,7446	0,9391
Alagoas	0,6500	1,0000	0,5086
Amapá	0,8097	0,7832	0,8033
Amazonas	0,8927	0,7565	0,8890
Bahia	0,8757	0,8189	0,8269
Ceará	0,7478	0,9102	0,6554
Distrito Federal	0,7940	0,9280	0,6777
Espírito Santo	0,7987	0,8336	0,7552
Goiás	0,6864	0,8365	0,6650
Maranhão	1,0000	0,9762	0,8011
Mato Grosso	0,6348	0,8624	0,6044
Mato Grosso do Sul	0,8876	0,8260	0,8308
Minas Gerais	0,7730	0,8065	0,7563
Pará	1,0000	0,7617	0,9690
Paraíba	0,7005	0,9639	0,5764
Paraná	0,9217	0,8097	0,8701
Pernambuco	0,5866	0,9464	0,5010
Piauí	0,7925	1,0000	0,6202
Rio de Janeiro	0,5892	1,0000	0,4610
Rio Grande do Norte	0,7326	0,8371	0,7007
Rio Grande do Sul	1,0000	0,8275	0,9175
Rondônia	1,0000	0,7221	1,0000
Roraima	0,9260	0,7427	0,9260
Santa Catarina	0,9493	0,8351	0,8719
São Paulo	0,6991	1,0000	0,5471
Sergipe	0,8735	0,7658	0,8668
Tocantins	0,8167	0,9116	0,7083

Na Figura 3 pode-se observar a representação gráfica das duas fronteiras. A região compreendida entre as duas forma a região onde é possível uma DMU estar situada e serve para estimar intervalos de variação do consumo residencial dado o IDH, ou vice-versa.

A análise visual da Figura 3 mostra que as DMUs estão bem próximas à fronteira eficiente, sendo esta constituída por um número pequeno de DMUs, o que caracteriza que o método cumpriu com seus dois objectivos (máxima discriminação e mínimo número de DMUs eficientes na fronteira). No entanto, existe pouca aderência à fronteira invertida, já que o método selecção de variáveis considera apenas a fronteira padrão.

Caso deseje-se usar uma óptica inversa (óptica da empresa), o consumo passa a ser considerado um *output* e os três índices considerados como *inputs*, tal como ser observado na Figura 4. Ao aplicar-se o mesmo método de selecção de variáveis para este caso com o modelo BCC, foram seleccionadas as variáveis PIB e Consumo per capita.

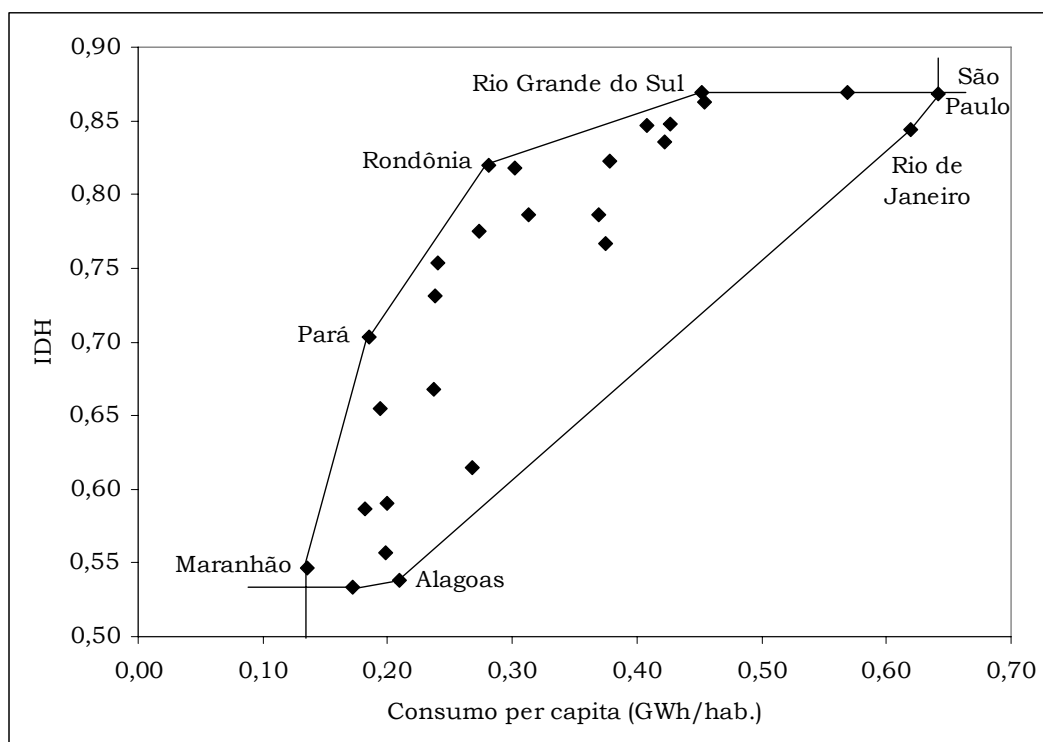


Figura 3: Fronteira DEA BCC e fronteira DEA BCC invertida para o caso investigado

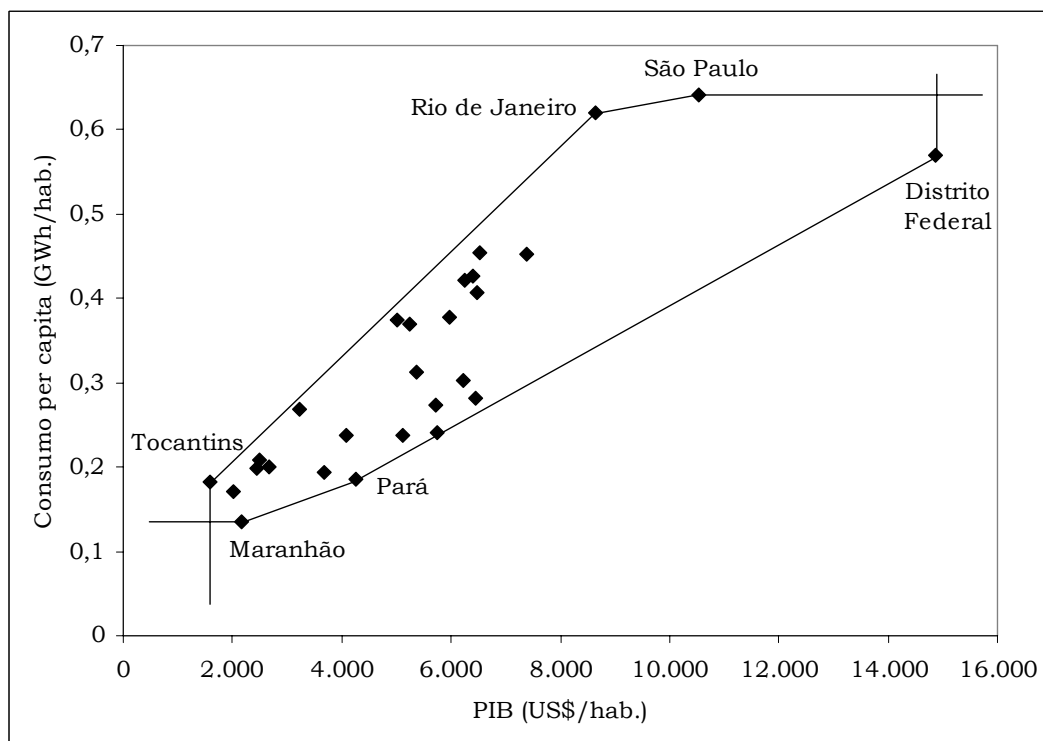


Figura 4: Fronteira DEA BCC e fronteira DEA BCC invertida para a óptica do empresário

4 Conclusões

Uma análise geral do método de selecção de variáveis, com o uso dos modelos CCR e BCC, mostra que em nenhum deles a variável IPC foi seleccionada. Ou seja, os outros índices usados são mais úteis para explicar o consumo do que o índice criado com esse fim.

Salienta-se o facto de os estados de região amazónica terem um bom desempenho segundo o índice composto. Uma possível explicação é o uso de geradores individuais que não entram no cálculo do consumo residencial de energia eléctrica, além da deficiente rede de energia eléctrica da região, que obriga o uso de outras fontes de energia.

Quanto ao objectivo do método de selecção de variáveis, a análise visual da Figura 3 mostra que as DMUs estão bem próximas à fronteira eficiente, sendo esta constituída por um número pequeno de DMUs, o que caracteriza que o método cumpriu com seus dois objectivos (máxima discriminação e mínimo número de DMUs eficientes na fronteira). A Figura 4 mostra a óptica do empresário, onde pode ser observado um melhor ajuste à fronteira.

5 Agradecimentos

Este trabalho teve o apoio do Fundo Setorial de Energia (CT-Energ), por intermédio do CNPq (CT-Energ/CNPq 01/2003), processo 400646/2003-0.

6 Referências

- Adler, N., Friedman, L. and Sinuany-Stern, Z. (2002) Review of ranking methods in the data envelopment analysis context, *European Journal of Operational Research*, Vol 140, pp. 249-265.
- Ali, A.I. (1993) Streamlined computation for Data Envelopment Analysis, *European Journal of Operational Research*, Vol 64, pp. 61-67.
- Angulo Meza, L., Biondi Neto, L., Soares de Mello, J.C.C.B. and Gomes, E.G. (2005) ISYDS – Integrated System for Decision Support (SIAD – Sistema Integrado de Apoio à Decisão): A Software Package for Data Envelopment Analysis Model, *Pesquisa Operacional*, Vol 25, No 3, pp. 493-503.
- Angulo Meza, L. and Lins, M.P.E. (2002) Review of methods for increasing discrimination in Data Envelopment Analysis, *Annals of Operations Research*, Vol 116, pp. 225-242.
- Angulo Meza, L., Soares de Mello, J.C.C.B., Gomes, E.G., Fernandes, A.J.S. and Biondi Neto, L. (2006) Análise da eficiência municipal no uso da energia elétrica considerando temperaturas. In: *Anais do XXXVIII Simpósio Brasileiro de Pesquisa Operacional*, Goiânia, Brasil.
- Banker, R.D., Charnes, A. and Cooper, W.W. (1984) Some models for estimating technical scale inefficiencies in Data Envelopment Analysis, *Management Science*, Vol 30, No 9, pp. 1078-1092.
- Borges, A.C. (2004) Fecomércio lança índice de potencial de consumo, Agência Estado Financeiro. Disponível em: <<http://www.aefinanceiro.com.br/artigos/2004/out/07/295.htm>>. Acesso em: 03 fev. 2005.
- Charnes, A., Cooper, W.W. and Rhodes, E. (1978) Measuring the efficiency of decision-making units, *European Journal of Operational Research*, Vol 2, pp. 429-444.
- Chauhan, N.S., Mohapatra, P.K.J. and Pandey, K.P. (2006) Improving energy productivity in

paddy production through benchmarking - An application of data envelopment analysis, *Energy Conversion and Management*, Vol 47, No 9-10, pp. 1063-1085.

Entani, T., Maeda, Y. and Tanaka, H. (2002) Dual models of interval DEA and its extensions to interval data, *European Journal of Operational Research*, Vol 136, pp. 32-45.

Golany, B. and Roll, Y. (1989) An application procedure for DEA, *Omega International Journal of Management Science*, Vol 17, No 3, pp. 237-250.

Gomes, E.G., Soares de Mello, J.C.C.B. and Mangabeira, J.A.C. (2006) Fronteira DEA difusa na avaliação de eficiência em agricultura, *Investigação Operacional*, Vol 26, No 1, pp. 65-88.

Hu, J.-L. and Kao, C.-H. (2007) Efficient energy-saving targets for APEC economies, *Energy Policy*, Vol 35, No 1, pp. 373-382.

Instituto Brasileiro de Geografia e Estatística (2002) Contas Regionais do Brasil 2002, Contas Regionais, 2a edição, No. 13.

Leta, F.R., Soares de Mello, J.C.C.B., Gomes, E.G. and Angulo-Meza, L. (2005) Métodos de melhora de ordenação em DEA aplicados à avaliação estática de tornos mecânicos, *Investigação Operacional*, Vol 25, No 2, pp. 229-242.

Lins, M.P.E. and Angulo-Meza, L. (2000) *Análise Envoltória de Dados e Perspectivas de Integração no Ambiente do Apoio à Decisão*, Editora da COOPE/UFRJ, Rio de Janeiro, Brasil.

Lins, M.P.E. and Moreira, M.C.B. (1999) Método I-O Stepwise para seleção de variáveis em modelos de Análise Envoltória Dados, *Pesquisa Operacional*, Vol 19, No 1, pp. 39-50.

Lins, M.P.E., Novaes, L.F.L. and Legey, L.F.L. (2005) Real estate value assessment: a double perspective data envelopment analysis, *Annals of Operations Research*, Vol 138, No 1, pp. 79-96.

Lins, M.P.E., Silva, A.C.M. and Rosa, L.P. (2002) Regional variations in energy consumption of appliances: conditional demand analysis applied to Brazilian households, *Annals of Operations Research*, Vol 117, No 1-4, pp. 235-246.

Lins, M.P.E., Sollero, M.K.V., Calôba, G.M. and Silva, A.C.M. (2007) Integrating the regulatory and utility firm perspectives, when measuring the efficiency of electricity distribution, *European Journal of Operational Research*, Vol 181, No 3, pp. 1413-1424.

Norman, M. and Stoker, B. (1991) *Data Envelopment Analysis: The Assessment of Performance*, John Wiley and Sons, New York, USA.

Pimenta, H.L.N, Macedo, M.A. and Soares de Mello, J.C.C.B. (2004) Decisão de realização de investimentos em tecnologia da informação com análise envoltória de dados, *Produção Online*, Vol 4, No 2.

Programa das Nações Unidas para o Desenvolvimento (2004) *Desenvolvimento Humano e IDH*. Disponível em: <<http://www.pnud.org.br/idh/>>. Acesso em: 03 fev. 2005.

Senra, L.F.A.C. (2004) Métodos de seleção de variáveis em DEA: estudo de caso no setor elétrico brasileiro. Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal Fluminense, Niterói, Brasil.

Senra, L.F.A.C., Nanci, L.C., Soares de Mello, J.C.C.B. and Angulo-Meza, L. (2007) Estudo sobre métodos de seleção de variáveis em DEA, *Pesquisa Operacional*, Vol 27, No 2, pp. 191-207.

Soares de Mello, J.C.C. B., Angulo-Meza, L., Gomes, E.G. and Biondi Neto, L. (2006) Evaluación de la concentración en una ruta aérea brasileira con modelo DEA y frontera invertida, *Revista de la Facultad de Ingeniería - Universidad de Tarapacá*, Vol 14, No 1, pp. 64-71.

Soares de Mello, J.C.C.B., Angulo-Meza, L., Gomes, E.G., Serapião, B.P. and Lins, M.P.E. (2003) Análise de envoltória de dados no estudo da eficiência e dos benchmarks para companhias aéreas brasileiras, *Pesquisa Operacional*, Vol 23, No 2, pp. 325-345.

Soares de Mello, J.C.C.B., Gomes, E.G., Angulo-Meza, L., Biondi Neto, L. and Sant'Anna, A.P. (2005) Fronteiras DEA difusas, *Investigação Operacional*, Vol 25, No 1, pp. 85-103.

Soares de Mello, J.C.C.B., Gomes, E.G., Angulo-Meza, L. and Lins, M.P.E. (2004) Selección de

variables para el incremento del poder de discriminación de los modelos DEA, *Revista de La Escuela de Perfeccionamiento en Investigación Operativa*, Vol 24, pp. 40-52.

Soares de Mello, J.C.C.B., Gomes, E.G., Soares de Mello, M.H.C. and Lins, M.P.E. (2002) Método multicritério para seleção de variáveis em modelos DEA, *Pesquisa Naval*, Vol 15, pp. 55-66.

Thanassoulis, E. (1996) Assessing the efficiency of schools with pupils of different ability using Data Envelopment Analysis, *Journal of the Operational Research Society*, Vol 47, No 1, pp. 84-97.

Utlü, Z. and Hepbasli, A. (2007) A review on analyzing and evaluating the energy utilization efficiency of countries, *Renewable & Sustainable Energy Reviews*, Vol 11, pp. 1-29.

Wagner, J.M. and Shimshak, D.G. (2007) Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives, *European Journal of Operational Research*, Vol 180, pp. 57-67.

Yamada, Y., Matui, T. and Sugiyama, M. (1994) New analysis of efficiency based on DEA, *Journal of the Operations Research Society of Japan*, Vol 37, No 2, pp. 158-167.

Optimização da Exploração de Recursos Hídricos considerando o Efeito de Queda

João Catalão [†]
Silvio Mariano [†]
Victor Mendes [‡]
Luís Ferreira [§]

[†] Departamento de Engenharia Electromecânica
UBI – Universidade da Beira Interior
catalao@ubi.pt
sm@ubi.pt

[‡] Departamento de Engenharia Electrotécnica e Automação
ISEL – Instituto Superior de Engenharia de Lisboa
vfmendes@isel.pt

[§] Departamento de Engenharia Electrotécnica e de Computadores
IST – Instituto Superior Técnico
lmf@ist.utl.pt

Abstract

This paper focuses on the optimal scheduling of hydro power systems. A new methodology based on nonlinear optimisation is proposed for solving the problem considering the head change effect. The numerical results obtained for a case study illustrate the behaviour of the new methodology proposed.

Resumo

Neste artigo é enunciado e descrito o problema de optimização da exploração de recursos hídricos. Uma nova metodologia baseada em optimização não linear é proposta para a resolução deste problema tendo em consideração o efeito de queda. Apresentam-se os resultados numéricos obtidos para um caso de estudo e conclui-se sobre o desempenho da nova metodologia proposta.

Keywords: Optimal scheduling, Nonlinear optimisation, Head change effect

Title: Optimal scheduling of hydro power systems considering the head change effect.

1 Introdução

O sistema electroprodutor nacional é caracterizado actualmente por uma acentuada dependência externa dos combustíveis fósseis e por uma elevada intensidade energética do Produto Interno Bruto, que traduz a quantidade de energia necessária para produzir riqueza. Esta intensidade para Portugal, sendo uma das mais elevadas na União Europeia, resulta não só de um problema estrutural da economia nacional, mas também da menor eficiência de exploração do sistema electroprodutor nacional no que respeita aos recursos endógenos e renováveis, como por exemplo, os recursos hídricos. Assim, a promoção da eficiência energética e das energias renováveis constitui uma importante contribuição para a descarbonização da economia nacional (Catalão, 2006).

A componente hídrica assume relevo na produção anual de energia eléctrica, quer pela potência já instalada, quer pelo potencial ainda por aproveitar, embora a sua contribuição varie de anos húmidos para anos secos de acordo com a maior ou menor pluviosidade. Atendendo à evolução típica do diagrama de cargas ao longo do dia, os recursos hídricos têm a vantagem de poder responder prontamente a solicitações de potência, visto que, o armazenamento de energia sob a forma potencial constitui uma reserva estratégica de utilização praticamente imediata. Pelo que, os recursos hídricos desempenham um papel fundamental para o sistema electroprodutor nacional, aliando segurança, fiabilidade e economia de produção, apresentando ainda uma elevada flexibilidade de operação.

Os aproveitamentos hidroeléctricos podem ser classificados de acordo com a potência instalada, atendendo ao Decreto-Lei n.º 189/88, em:

- Mini-hídricas, de potência instalada inferior a 10 MW;
- Grandes aproveitamentos hidroeléctricos, de potência instalada superior a 10 MW.

Ainda, os grandes aproveitamentos hidroeléctricos podem ser classificados de acordo com a capacidade de armazenamento em:

- Aproveitamentos a fio de água — apresentam uma capacidade de armazenamento reduzida, aproveitando a energia do caudal afluente;
- Aproveitamentos de albufeira — armazenam o caudal afluente para utilização em condições mais vantajosas. Alguns destes aproveitamentos possuem capacidade de bombagem, podendo funcionar como turbina gerando energia eléctrica, ou podendo funcionar como bomba voltando a repor água na albufeira.

Nos aproveitamentos a fio de água a capacidade de armazenamento e a altura de queda são reduzidas. O tempo de vazamento, i.e., o tempo que levaria um reservatório a atingir o nível mínimo de água partindo do nível máximo de água e sem afluência hídrica, é de apenas algumas horas. Devido ao reduzido volume de água nestes reservatórios, a altura de queda pode variar significativamente, tornando a eficiência de exploração sensível à altura de queda: efeito de queda. Em Portugal, essa é a situação por exemplo no rio Douro, cujos aproveitamentos contribuem de forma relevante para a produção de energia eléctrica, dispondo no seu conjunto de uma potência instalada que é cerca de metade da potência total do sistema hidroeléctrico nacional.

O efeito de queda é um problema complexo e relevante na optimização da exploração de recursos hídricos. Tipicamente, considera-se que os recursos hídricos operam em condições estacionárias com altura de queda constante, sendo assumindo que a potência entregue é representada por uma função afim do caudal de água turbinado. Contudo, nos aproveitamentos a fio de água a potência entregue tem que ser considerada como uma função não só do caudal de água turbinado, mas também da altura de queda.

Assim, para aproveitamentos a fio-de-água é necessário considerar matematicamente o efeito de queda ao nível da função objectivo, e não só a nível contabilístico com actualizações de potência para a altura de queda realmente obtida. O problema do efeito de queda na optimização é ainda mais agudizado pelo aumento da potência instalada, quer de produção, quer de bombagem. O volume de água nos reservatórios mantém-se, as capacidades de turbinar e de bombear aumentam, logo o efeito de queda pode ser mais acentuado.

A programação dinâmica foi uma das primeiras metodologias utilizadas para a resolução do problema de optimização da exploração de recursos hídricos (Amado and Ribeiro, 1987). A utilização da programação dinâmica para a resolução do problema permite obter a solução óptima. Contudo, a maldição da dimensionalidade implica que a sua utilização só seja possível para problemas de dimensão reduzida.

A programação linear em rede (Hreinsson, 1988; Piekutowski et al., 1994; Wood and Wollenberg, 1996) foi proposta como alternativa à programação dinâmica, tendo despertado interesse nos investigadores pelo facto de uma cascata hídrica ter uma estrutura que pode ser matematicamente modelada por uma rede de fluxos, o que torna esta alternativa apropriada ao problema de optimização da exploração de recursos hídricos. A programação linear em rede acomoda facilmente várias restrições, tais como, o balanço da água para cada reservatório, limites mínimos e máximos do volume de água nos reservatórios, e tempo de trânsito hídrico. A programação linear em rede proporciona códigos eficientes e robustos, comercializados e facilmente à disposição dos utilizadores. Contudo, a programação linear em rede apresenta a desvantagem de não permitir a consideração correcta das curvas características de potência entregue versus caudal de água turbinado para as centrais hídricas, tipicamente curvas não lineares e não convexas.

Mais recentemente, a metodologia baseada em optimização linear inteira-mista foi utilizada para a resolução do problema de optimização da exploração de recursos hídricos (Chang et al., 2001; Conejo et al., 2002; Guan et al., 1999; Nilsson and Sjelvgren, 1999), usufruindo das vantagens da programação linear em rede e, simultaneamente, permitindo considerar a existência de zonas proibidas, i.e., valores de potência para os quais não se pode manter a turbina hídrica em funcionamento. Contudo, a discretização das curvas características de potência entregue versus caudal de água turbinado, utilizada nesta metodologia para considerar as variações da altura de queda, requer um incremento significativo na capacidade de memória e no tempo de computação.

Metodologias baseadas em inteligência artificial, nomeadamente redes neuronais artificiais (Liang and Hsu, 1996; Naresh and Sharma, 2002) e algoritmos genéticos (Cau and Kaye, 2002; Chen and Chang, 1996), foram também utilizadas para a resolução do problema de optimização da exploração de recursos hídricos. Embora estas metodologias se apresentem como prometedoras para a resolução deste problema, apenas podem ser encontradas soluções sub-óptimas devido às heurísticas utilizadas no processo de decisão, para além da desvantagem do elevado tempo de computação.

A evolução tecnológica no final da década de noventa, ao nível da capacidade de cálculo dos computadores e do desempenho dos *solvers*, motivou a utilização de metodologias baseadas em optimização não linear para a resolução do problema de optimização da exploração de recursos hídricos (Ni et al., 1999; Pursimo et al., 1998). As metodologias baseadas em optimização não linear permitem uma representação mais aproximada e fidedigna da realidade, comparativamente à programação linear em rede, apresentando hoje em dia tempos de computação aceitáveis (Catalão et al., 2006; Mariano et al., 2006).

Neste artigo é enunciado e descrito o problema de optimização da exploração de recursos hídricos. Uma nova metodologia baseada em optimização não linear é proposta para a resolução deste problema tendo em consideração o efeito de queda. Apresentam-se os resultados numéricos obtidos para um caso de estudo e conclui-se sobre o desempenho da nova metodologia proposta.

2 Formulação do Problema

A notação usada neste artigo para a formulação do problema é apresentada seguidamente:

- k — índice da hora;
- j — índice do reservatório;
- $f(j)$ — índice do reservatório a montante da central j ;
- $t(j)$ — índice do reservatório a jusante da central j ;
- K — número total de horas;
- J — número total de reservatórios;
- λ_k — preço da energia eléctrica na hora k ;
- p_{jk} — potência entregue pela central j na hora k ;
- v_{jk} — volume de água no reservatório j no final da hora k ;
- $v_j^{\min}$ — volume mínimo de água no reservatório j ;
- $v_j^{\max}$ — volume máximo de água no reservatório j ;
- v_{j0} — volume inicial de água no reservatório j ;
- v_{jK} — volume final de água no reservatório j ;
- a_{jk} — afluência ao reservatório j na hora k ;
- q_{jk} — caudal de água turbinado proveniente do reservatório j na hora k ;
- $q_j^{\min}$ — caudal mínimo de água turbinado proveniente do reservatório j ;
- $q_{jk}^{\max}$ — caudal máximo de água turbinado proveniente do reservatório j na hora k ;
- τ_{yj}^q — tempo que o caudal de água turbinado demora desde o reservatório y até ao reservatório j , em número de horas;
- s_{jk} — caudal de água descarregado pelo reservatório j na hora k ;
- τ_{yj}^s — tempo que o caudal de água descarregado demora desde o reservatório y até ao reservatório j , em número de horas;
- η_{jk} — eficiência de exploração associada à central j na hora k ;
- $\eta_j^{\min}$ — eficiência mínima de exploração associada à central j ;
- $\eta_j^{\max}$ — eficiência máxima de exploração associada à central j ;
- h_{jk} — altura de queda entre o reservatório j e o reservatório imediatamente a jusante, na hora k ;
- $h_j^{\min}$ — altura mínima de queda entre o reservatório j e o reservatório imediatamente a jusante;
- $h_j^{\max}$ — altura máxima de queda entre o reservatório j e o reservatório imediatamente a jusante;
- l_{jk} — nível de água no reservatório j , em relação ao mar, na hora k ;

- $l_j^{\min}$ — nível mínimo de água no reservatório j , em relação ao mar;
- $l_j^{\max}$ — nível máximo de água no reservatório j , em relação ao mar;
- Ψ_j — função que determina o valor económico futuro da água armazenada no reservatório j ;
- $\mathbf{z}$ — vector das variáveis de fluxo;
- $\mathbf{z}^{\min}$ — vector dos limites mínimos associados às variáveis de fluxo;
- $\mathbf{z}^{\max}$ — vector dos limites máximos associados às variáveis de fluxo;
- $\mathbf{b}^{\min}$ — vector dos limites mínimos para as restrições;
- $\mathbf{b}^{\max}$ — vector dos limites máximos para as restrições;
- $\mathbf{H}$ — matriz Hessiana;
- $\mathbf{A}$ — matriz dos coeficientes das restrições;
- $\mathcal{Y}_j$ — Conjunto de reservatórios imediatamente a montante do reservatório j .

O problema de optimização da exploração de recursos hídricos visa determinar os níveis de utilização para cada reservatório da cascata, durante o horizonte temporal considerado, atendendo à disponibilidade de água e às restrições físicas e operacionais de cada recurso. A forma de resolução deste problema consiste na maximização do valor económico da produção hidroeléctrica durante o horizonte temporal de uma semana, subdividido em períodos de uma hora.

O problema de optimização da exploração de recursos hídricos corresponde neste artigo a um problema de optimização não linear. A função objectivo a ser maximizada é dada por:

$$F = \sum_{j=1}^J \sum_{k=1}^K \lambda_k p_{jk} + \sum_{j=1}^J \Psi_j(v_{jK}) \quad (1)$$

A função objectivo em (1) é composta por dois termos. O primeiro termo representa o lucro total auferido com a exploração dos recursos envolvidos no planeamento. O segundo termo representa o valor económico que a água armazenada nos reservatórios tem para a exploração futura, sendo este valor fornecido pelo planeamento de médio prazo.

O problema de optimização da exploração de recursos hídricos é sujeito a restrições de igualdade e restrições de desigualdade ou limites nas variáveis. As seguintes restrições descrevem as limitações de exploração associadas aos recursos:

$$v_{jk} = v_{j,k-1} + a_{jk} + \sum_{y \in \mathcal{Y}_j} (q_{y,k-\tau_{y_j}^q} + s_{y,k-\tau_{y_j}^s}) - q_{jk} - s_{jk} \quad j \in J \quad k \in K \quad (2)$$

$$p_{jk} = q_{jk} \eta_{jk}(h_{jk}) \quad j \in J \quad k \in K \quad (3)$$

$$h_{jk} = l_{f(j)k}(v_{f(j)k}) - l_{t(j)k}(v_{t(j)k}) \quad j \in J \quad k \in K \quad (4)$$

$$v_j^{\min} \leq v_{jk} \leq v_j^{\max} \quad j \in J \quad k \in K \quad (5)$$

$$q_j^{\min} \leq q_{jk} \leq q_{jk}^{\max}(h_{jk}) \quad j \in J \quad k \in K \quad (6)$$

$$s_{jk} \geq 0 \quad j \in J \quad k \in K \quad (7)$$

Em (2) é apresentada a equação do balanço da água para cada reservatório. O volume de água contido num determinado reservatório da cascata depende, quer da afluência a esse reservatório, quer do volume de água utilizado na conversão energética ou descarregado por esse reservatório, e ainda depende do volume de água proveniente dos reservatórios imediatamente a montante. Assim, os volumes de água são interdependentes devido à configuração física dos trânsitos hídricos na cascata, i.e., uma vez tomadas quaisquer decisões envolvendo um determinado reservatório, estas decisões vão influenciar, naquela hora e nas horas seguintes, as decisões posteriores, quer neste reservatório, quer em outros reservatórios a montante e a jusante deste. Esta interdependência entre os volumes de água obriga à utilização de restrições sobre uma estrutura que pode ser matematicamente modelada por uma rede de fluxos (Ferreira et al., 1989). A diferença entre o tempo que o caudal de água turbinado e o caudal de água descarregado demora entre dois reservatórios poderá ser considerável quando existe conduta forçada entre reservatórios para o trânsito do caudal de água turbinado, sendo normalmente o caudal de água descarregado para o leito do rio. O tempo de trânsito hídrico é considerado apenas no caso da duração desse trânsito hídrico ser significativa relativamente ao intervalo de tempo estabelecido para um período do horizonte temporal.

Em (3) a potência entregue p_{jk} é considerada uma função do caudal de água turbinado q_{jk} e da eficiência de exploração η_{jk} , que depende da altura de queda h_{jk} entre o reservatório a montante, donde provém a água para a central, e o reservatório a jusante, para onde é enviada posteriormente a água, da central.

Em (4) a altura de queda é considerada uma função do nível de água nos reservatórios a montante $l_{f(j)k}$ e a jusante $l_{t(j)k}$ da central, que dependem do volume de água nos reservatórios respectivos.

Em (5) impõem-se limites mínimos $v_j^{\min}$ e máximos $v_j^{\max}$ para o volume de água em cada reservatório j .

Em (6) impõem-se também limites mínimos e máximos para o caudal de água turbinado. O caudal mínimo de água turbinado $q_j^{\min}$ é considerado nulo neste artigo, podendo no entanto ser imposto por razões de navegabilidade, recreativas ou ambientais. O caudal máximo de água turbinado $q_{jk}^{\max}$ pode ser diferente em cada hora k , quando existe uma significativa dependência com a altura de queda.

Em (7) impõe-se um limite mínimo nulo para o caudal de água descarregado, não tendo contudo limite máximo, i.e., o seu valor é suficientemente grande. Por exemplo, esta descarga ocorre quando o volume de água está no seu limite máximo, sendo a descarga necessária por razões de segurança.

O volume inicial de água nos reservatórios v_{j0} , bem como a afluência aos reservatórios, são considerados valores conhecidos. Assume-se também que não existem quaisquer restrições de carácter dinâmico associadas aos recursos hídricos, em consequência da sua elevada flexibilidade de operação.

3 Problema de Optimização Não Linear

O problema de optimização não linear considerado neste artigo é formulado por:

$$\text{Max } F(\mathbf{z}) \quad (8)$$

sujeito a:

$$\mathbf{b}^{\min} \leq \mathbf{A} \mathbf{z} \leq \mathbf{b}^{\max} \quad (9)$$

$$\mathbf{z}^{\min} \leq \mathbf{z} \leq \mathbf{z}^{\max} \quad (10)$$

Em (9) é apresentada uma restrição mais genérica que não só descreve (2), com $\mathbf{b}^{\min} = \mathbf{b}^{\max}$, correspondendo à equação do balanço da água para cada reservatório, mas também descreve a dependência que o caudal máximo de água turbinado tem relativamente à altura de queda. Esta dependência é assim representada por uma nova restrição de desigualdade que, após conveniente aproximação por uma função linear e posterior tratamento algébrico, é escrita de acordo com (9).

A restrição (10) corresponde às restrições (5) a (7), i.e., corresponde aos limites mínimos e máximos para o volume de água, para o caudal de água turbinado, e para o caudal de água descarregado.

Em (3) a eficiência de exploração depende da altura de queda entre os reservatórios a montante e a jusante da central. Assumindo uma aproximação por uma função linear, esta dependência é dada por:

$$\eta_{jk} = \eta_j^0 + \alpha_j h_{jk} \quad j \in J \quad k \in K \quad (11)$$

em que os parâmetros η_j^0 e α_j são a ordenada na origem e o declive, respectivamente, dados por:

$$\eta_j^0 = \eta_j^{\max} - \alpha_j h_j^{\max} \quad j \in J \quad (12)$$

$$\alpha_j = (\eta_j^{\max} - \eta_j^{\min}) / (h_j^{\max} - h_j^{\min}) \quad j \in J \quad (13)$$

Em (13) o parâmetro α_j para cada reservatório j depende dos valores mínimos e máximos para a eficiência de exploração e para a altura de queda.

Em (4) o nível de água depende do volume de água no reservatório respectivo. Assumindo uma aproximação por uma função linear, esta dependência é dada por:

$$l_{jk} = l_j^0 + \beta_j v_{jk} \quad j \in J \quad k \in K \quad (14)$$

em que os parâmetros l_j^0 e β_j são a ordenada na origem e o declive, respectivamente, dados por:

$$l_j^0 = l_j^{\max} - \beta_j v_j^{\max} \quad j \in J \quad (15)$$

$$\beta_j = (l_j^{\max} - l_j^{\min}) / (v_j^{\max} - v_j^{\min}) \quad j \in J \quad (16)$$

Em (16) o parâmetro β_j para cada reservatório j depende dos valores mínimos e máximos para o nível de água e para o volume de água.

Substituindo (11) em (3) tem-se que:

$$p_{jk} = q_{jk} (\eta_j^0 + \alpha_j h_{jk}) \quad j \in J \quad k \in K \quad (17)$$

Ainda, substituindo (4) e (14) em (17), a potência entregue é representada por uma função não linear, sendo dada por:

$$p_{jk} = \eta_j^0 q_{jk} + \alpha_j l_{f(j)}^0 q_{jk} - \alpha_j l_{t(j)}^0 q_{jk} + \alpha_j \beta_{f(j)} q_{jk} v_{f(j)k} - \alpha_j \beta_{t(j)} q_{jk} v_{t(j)k} \quad j \in J \quad k \in K \quad (18)$$

Um novo parâmetro χ_j , que corresponde ao produto de α_j por $\beta_{f(j)}$, é definido para ponderar a dependência não linear da potência entregue com o caudal de água turbinado e com o volume de água.

Em (6) o caudal máximo de água turbinado é considerado dependente da altura de queda. Assumindo uma aproximação por uma função linear, esta dependência é dada por:

$$q_{jk}^{\max} = q_j^0 - \delta_j h_{jk} \quad j \in J \quad k \in K \quad (19)$$

em que os parâmetros q_j^0 e δ_j são, respectivamente, dados por:

$$q_j^0 = q_j^{\text{hmin}} + \delta_j h_j^{\text{min}} \quad j \in J \quad (20)$$

$$\delta_j = (q_j^{\text{hmin}} - q_j^{\text{hmax}}) / (h_j^{\text{max}} - h_j^{\text{min}}) \quad j \in J \quad (21)$$

Em (21) o parâmetro δ_j para cada reservatório j depende dos valores mínimos e máximos para a altura de queda e para o correspondente caudal máximo de água turbinado.

Substituindo (4) e (14) em (19) tem-se que:

$$q_{jk}^{\max} = q_j^0 - \delta_j l_{f(j)}^0 + \delta_j l_{t(j)}^0 - \delta_j \beta_{f(j)} v_{f(j)k} + \delta_j \beta_{t(j)} v_{t(j)k} \quad j \in J \quad k \in K \quad (22)$$

Assim, o caudal máximo de água turbinado é representado por uma função linear, sendo dado por:

$$q_{jk}^{\max} = \gamma_j^0 - \gamma_j^1 v_{f(j)k} + \gamma_j^2 v_{t(j)k} \quad j \in J \quad k \in K \quad (23)$$

em que os parâmetros γ_j^0 , γ_j^1 e γ_j^2 são, respectivamente, dados por:

$$\gamma_j^0 = q_j^0 - \delta_j l_{f(j)}^0 + \delta_j l_{t(j)}^0 \quad j \in J \quad (24)$$

$$\gamma_j^1 = \delta_j \beta_{f(j)} \quad j \in J \quad (25)$$

$$\gamma_j^2 = \delta_j \beta_{t(j)} \quad j \in J \quad (26)$$

Substituindo (23) em (6) obtém-se uma nova restrição de desigualdade, escrita de acordo com (9), sendo dada por:

$$q_{jk} + \gamma_j^1 v_{f(j)k} - \gamma_j^2 v_{t(j)k} \leq \gamma_j^0 \quad j \in J \quad (27)$$

Na nova metodologia baseada em optimização não linear a exploração de recursos hídricos é determinada tendo em consideração o efeito de queda.

4 Caso de Estudo

O caso de estudo é constituído por um sistema hidroeléctrico com sete reservatórios em cascata, apresentado na figura 1, sendo o horizonte temporal de 168 horas. Este sistema hidroeléctrico é baseado no Douro Nacional: aproveitamentos a fio de água de 1-Pocinho, 2-Valeira, 4-Rêgua, 5-Carrapatelo e 7-Crestuma-Lever; aproveitamentos de albufeira de 3-Vilar-Tabuaço e 6-Torrão.

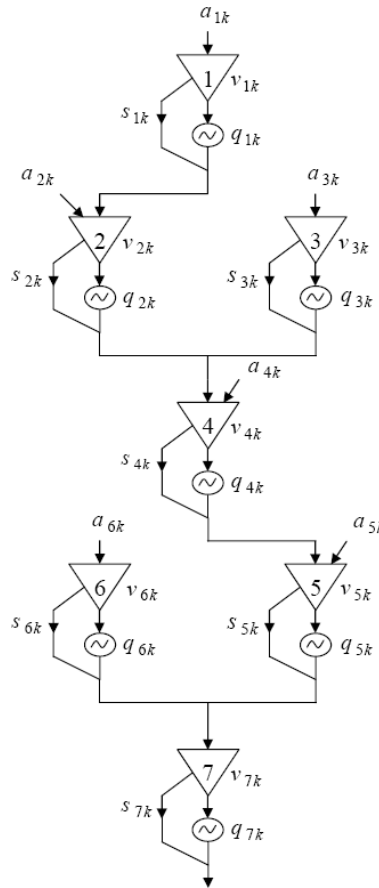


Figura 1: Sistema hidroeléctrico com sete reservatórios em cascata

O tempo de trânsito hídrico não é considerado neste caso de estudo, visto que, a duração desse trânsito hídrico não é significativa relativamente ao intervalo de tempo estabelecido para um período do horizonte temporal. A afluência aos reservatórios 1 a 6 é apresentada na figura 2, sendo que o reservatório situado no fim da cascata recebe apenas a água dos reservatórios a montante. Ainda, não é considerada a possibilidade de bombagem no reservatório 6. O perfil dos preços da energia eléctrica é apresentado na figura 3, sendo \$ uma quantidade monetária simbólica.

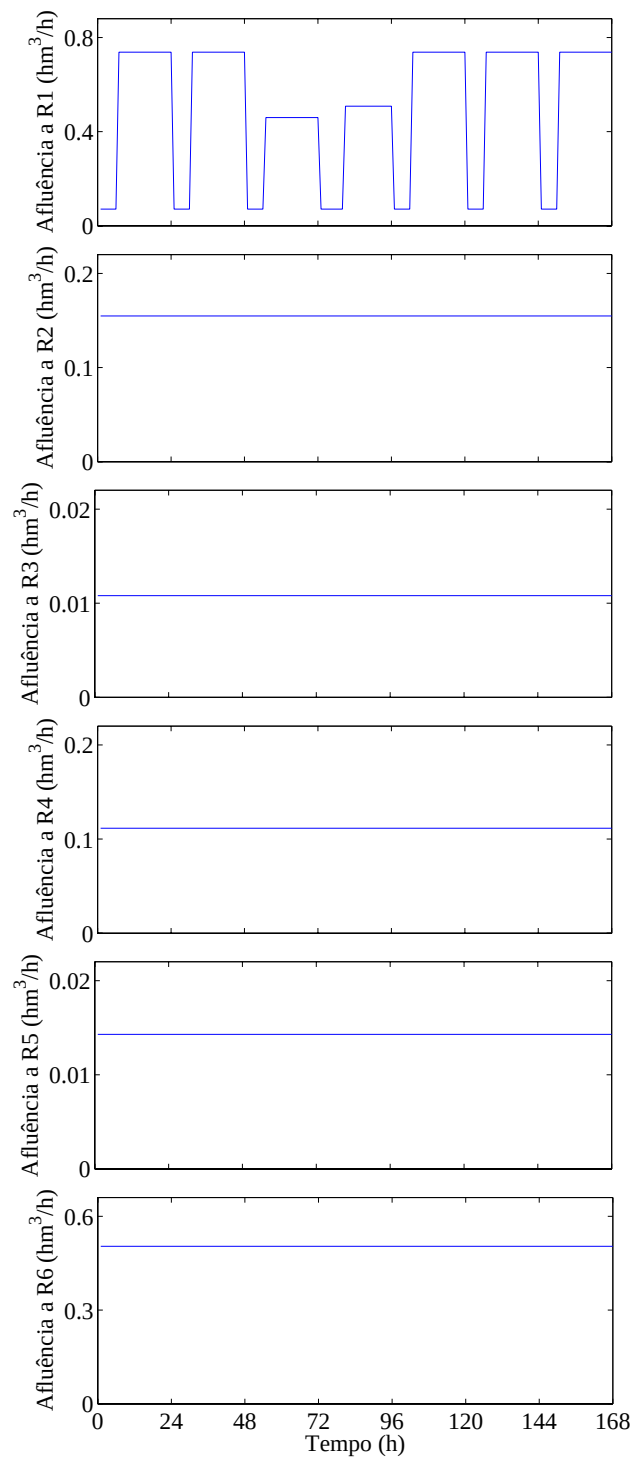


Figura 2: Afluência aos reservatórios 1 a 6

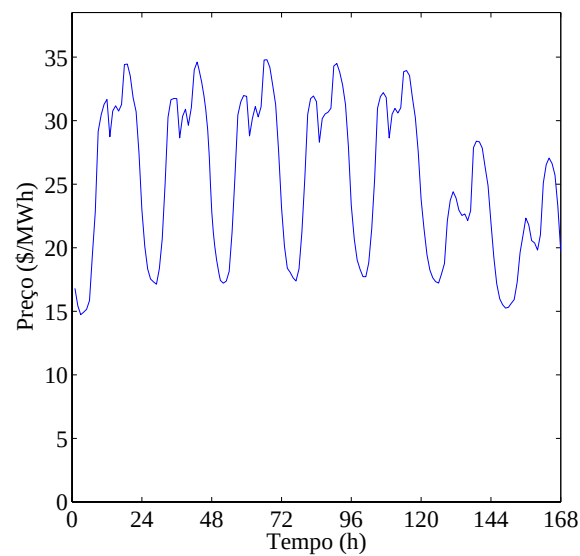


Figura 3: Perfil dos preços da energia eléctrica

A nova metodologia proposta foi implementada num computador com processador a 600-MHz e 256 MB de RAM, usando a aplicação informática MATLAB e o *solver* Xpress-MP.

Nas figuras 4 e 5 são apresentadas as trajetórias para o volume de água nos reservatórios e para o caudal de água turbinado.

O volume final de água nos reservatórios é restringido de forma a ser igual ao volume inicial de água nesses reservatórios, correspondendo a 80% do volume máximo de água. Ainda, o volume mínimo de água nos reservatórios corresponde a 40% do volume máximo de água.

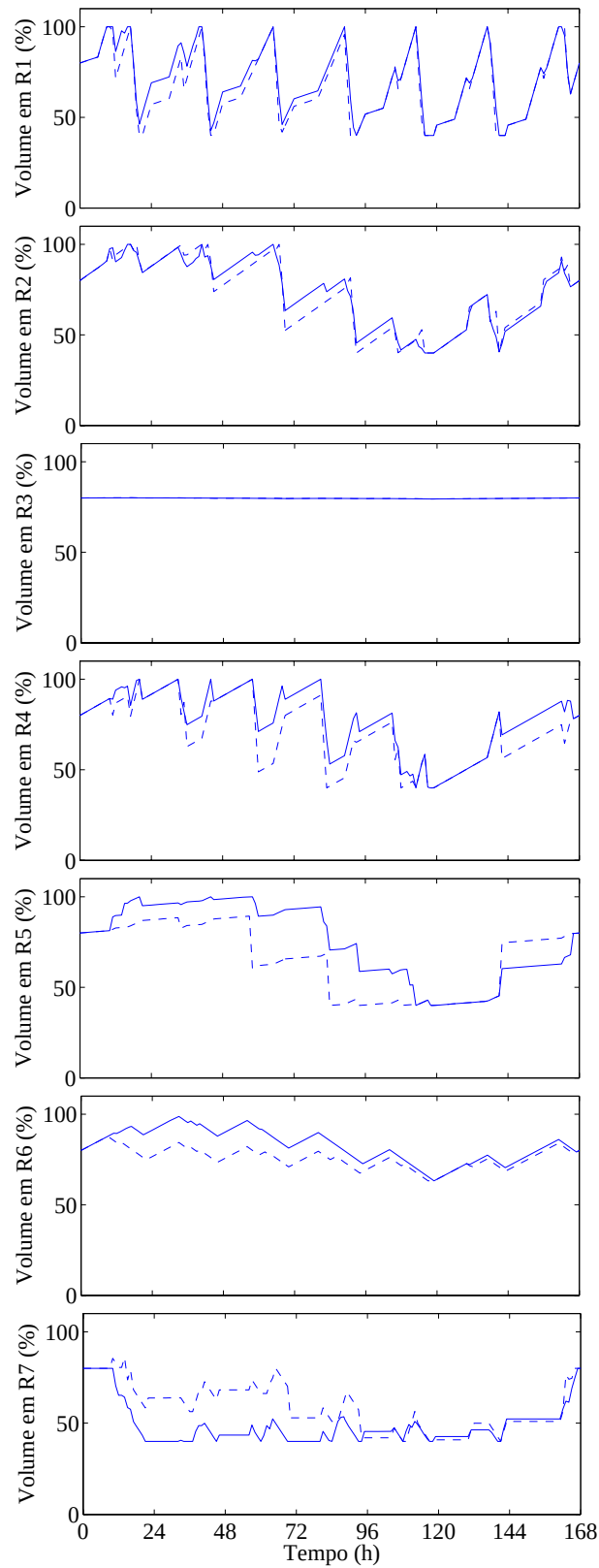


Figura 4: Volume de água nos reservatórios. As linhas a traço contínuo e as linhas a traço interrompido representam, respectivamente, os resultados obtidos com a nova metodologia baseada em optimização não linear e os resultados obtidos com uma metodologia baseada em optimização linear

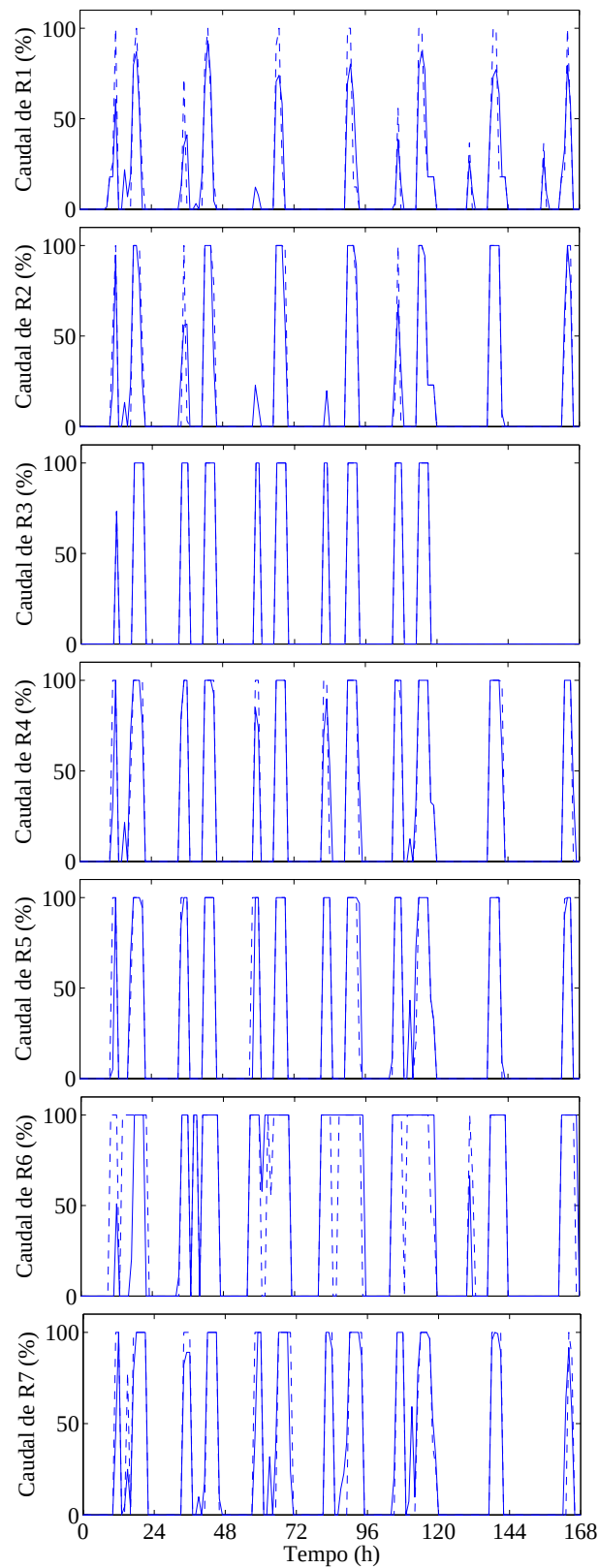


Figura 5: Caudal de água turbinado proveniente dos reservatórios. As linhas a traço contínuo e as linhas a traço interrompido representam, respectivamente, os resultados obtidos com a nova metodologia baseada em otimização não linear e os resultados obtidos com uma metodologia baseada em otimização linear

Os resultados obtidos para o volume de água nos reservatórios, apresentados anteriormente na figura 4, permitem ilustrar a influência do efeito de queda na exploração ótima de recursos hídricos. Nos resultados obtidos com a nova metodologia baseada em otimização não linear, a trajectória para o volume de água nos reservatórios situados no início da cascata é elevada, visando a maximização do lucro total. Contrariamente, a trajectória para o volume de água no reservatório situado no fim da cascata é reduzida, visando aumentar a altura de queda para os reservatórios imediatamente a montante.

Os resultados obtidos para o caudal de água turbinado proveniente dos reservatórios, apresentados anteriormente na Figura 5, mostram que o caudal de água turbinado varia mais acentuadamente, entre o mínimo e o máximo, para uma metodologia baseada em otimização linear comparativamente à nova metodologia baseada em otimização não linear, visto que, a metodologia baseada em otimização linear ignora o efeito de queda.

Na tabela 1 são apresentados os resultados comparativos entre as metodologias baseadas em otimização linear e não linear, para este caso de estudo. Assim, a nova metodologia proposta conduz a um lucro total superior.

Tabela 1: Resultados comparativos entre as metodologias baseadas em otimização linear e não linear

Optimização	Lucro (\$)	% Aumento	Tempo de computação (s)
Linear	1 522 440	-	2.25
Não Linear	1 598 700	5.01	4.14

Ainda, o tempo de computação adicional requerido pela nova metodologia proposta é de apenas 1.89s.

5 Conclusões

Nos aproveitamentos a fio de água, por exemplo no rio Douro, a reduzida capacidade de armazenamento torna a eficiência de exploração sensível à altura de queda. Assim, o problema de optimização da exploração de recursos hídricos deve ter em consideração o efeito de queda.

A resolução deste problema foi obtida através de uma nova metodologia baseada em optimização não linear, que considera não só a potência entregue como uma função do caudal de água turbinado e da altura de queda, mas também que o caudal máximo de água turbinado pode ser função da altura de queda, visando beneficiar a eficiência de exploração. A exploração ótima de recursos hídricos foi ilustrada em consequência do novo parâmetro apresentado, que pondera a dependência não linear da potência entregue com o caudal de água turbinado e com o volume de água.

Os resultados numéricos confirmam o bom desempenho da nova metodologia proposta, visto que, obtém uma melhor exploração comparativamente a uma metodologia baseada em optimização linear, com um tempo de computação reduzido.

6 Referências

- Amado, S. M. and Ribeiro, C. C. (1987) Short-Term Generation Scheduling of Hydraulic Multi-Reservoir Multi-Area Interconnected Systems, *IEEE Transactions on Power Systems*, Vol. PWRS-2, No. 3, pp. 758-763.
- Catalão, J. P. S., Mariano, S. J. P. S., Mendes, V. M. F., and Ferreira, L. A. F. M. (2006) Parameterisation Effect on the Behaviour of a Head-Dependent Hydro Chain using a Nonlinear Model, *Electric Power Systems Research*, Vol. 76, No. 6-7, pp. 404-412.
- Catalão, J. P. S. (2006) *Novas Metodologias de Optimização em Sistemas de Energia Hidrotérmicos*, Dissertação de Doutoramento, Universidade da Beira Interior, Covilhã.
- Cau, T. D. H. and Kaye, R. J. (2002) Evolutionary Optimisation Method for Multistorage Hydrothermal Scheduling, *IEE Proceedings-Generation Transmission and Distribution*, Vol. 149, No. 2, pp. 152-156.
- Chang, G. W., Aganagic, M., Waight, J. G., Medina, J., Burton, T., Reeves, S., and Christoforidis, M. (2001) Experiences With Mixed Integer Linear Programming Based Approaches on Short-Term Hydro Scheduling, *IEEE Transactions on Power Systems*, Vol. 16, No. 4, pp. 743-749.
- Chen, P. -H. and Chang, H. -C. (1996) Genetic Aided Scheduling of Hydraulically Coupled Plants in Hydro-Thermal Coordination, *IEEE Transactions on Power Systems*, Vol. 11, No. 2, pp. 975-981.
- Conejo, A. J., Arroyo, J. M., Contreras, J., and Villamor, F. A. (2002) Self-Scheduling of a Hydro Producer in a Pool-Based Electricity Market, *IEEE Transactions on Power Systems*, Vol. 17, No. 4, pp. 1265-1272.
- Ferreira, L. A. F. M., Anderson, T., Imparato, C. F., Miller, T. E., Pang, C. K., Svoboda, A., and Vojdani, A. F. (1989) Short-Term Resource Scheduling in Multi-Area Hydrothermal Power Systems, *International Journal of Electrical Power & Energy Systems*, Vol. 11, No. 3, pp. 200-212.
- Guan, X., Svoboda, A., and Li, C. -A. (1999) Scheduling Hydro Power Systems with Restricted Operating Zones and Discharge Ramping Constraints, *IEEE Transactions on Power Systems*, Vol. 14, No. 1, pp. 126-131.
- Hreinsson, E. B. (1988) Optimal Short Term Operation of a Purely Hydroelectric System, *IEEE Transactions on Power Systems*, Vol. 3, No. 3, pp. 1072-1077.
- Liang, R. -H. and Hsu, Y. -Y. (1996) Short-Term Hydro-Scheduling using Hopfield Neural Network, *IEE Proceedings-Generation Transmission and Distribution*, Vol. 143, No. 3, pp. 269-275.
- Mariano, S. J. P. S., Catalão, J. P. S., Mendes, V. M. F., and Ferreira, L. A. F. M. (2006) Analysis of the Short-Term Profit-Based Behavior of a Hydroelectric Energy System using a Nonlinear Programming Method, *Proceedings of the 19th Mini-EURO Conference on Operational Research Models and Methods in the Energy Sector*, Coimbra, Portugal.
- Naresh, R. and Sharma, J. (2002) Short Term Hydro Scheduling using Two-Phase Neural Network, *International Journal of Electrical Power & Energy Systems*, Vol. 24, No. 7, pp. 583-590.
- Ni, E., Guan, X., and Li, R. (1999) Scheduling Hydrothermal Power Systems with Cascaded and Head-Dependent Reservoirs, *IEEE Transactions on Power Systems*, Vol. 14, No. 3, pp. 1127-1132.
- Nilsson, O. and Sjelvgren, D. (1996) Mixed-Integer Programming Applied to Short-Term Planning of a Hydro-Thermal System, *IEEE Transactions on Power Systems*, Vol. 11, No. 1, pp. 281-286.
- Piekutowski, M., Litwinowicz, T., and Frowd, R. J. (1994) Optimal Short-Term Scheduling for a Large-Scale Cascaded Hydro System, *IEEE Transactions on Power Systems*, Vol. 9, No. 2, pp. 805-811.

Pursimo, J. M., Antila, H. K., Vilkkko, M. K., and Lautala, P. A. (1998) A Short-Term Scheduling for a Hydropower Power Plant, *International Journal of Electrical Power & Energy Systems*, Vol. 20, No. 8, pp. 525-532.

Wood, A. J. and Wollenberg, B. F. (1996) *Power Generation, Operation and Control*, Wiley, New York, USA.

Experiências computacionais com modelos de fluxo para múltiplos produtos com funções de custo não lineares e não separáveis

Luis Ernesto Torres Guardia [†]
João Carlos C. B. Soares de Mello [†]

[†] Departamento de Engenharia de Produção – Universidade Federal Fluminense
Brasil
tepletg@vm.uff.br
jcsmello@pesquisador.cnpq.br
<http://www.uff.br/decisao>

Abstract

In this paper we present the study and the numerical implementation of the primal-dual interior-point method for the solution of the convex nonlinear multicommodity network flow problem. At each iteration of the interior-point method, we solve the corresponding linear system, expressed by the augmented-indefinite system, using an indefinite preconditioned conjugate gradient algorithm combined with the AINV algorithm. We conduct some numerical experiments for networks of different dimensions and number of products and for some nonlinear costs. The computational results show the effectiveness of the interior-point method for this class of network problem.

Resumo

Neste trabalho são apresentados o estudo e a implementação numérica do método de pontos interiores primal-dual para o problema não linear convexo de fluxo em rede com múltiplos produtos. Em cada iteração do método de pontos interiores, resolve-se o correspondente sistema linear, expresso na forma de aumentado indefinido usando o algoritmo do gradiente conjugado com um pré-condicionador indefinido apropriado combinado com o algoritmo AINV. Foram realizados alguns testes numéricos para redes de várias dimensões e vários produtos, com funções de custos não lineares. Os resultados computacionais mostram a eficiência do método de pontos interiores para o caso de fluxo em rede para múltiplos produtos.

Keywords: Nonlinear programming. Interior-point method. Network flow problem

Title: Computational experiments with flow models for multicommodity using non linear and non separable cost functions.

1 Introdução

O presente trabalho trata da solução numérica de um problema de fluxo em rede, com vários produtos em circulação pelo mesmo arco. Considera-se que, de forma geral, o custo associado a cada arco é uma função não linear e não separável. Há ainda restrições de capacidade para os arcos.

Este problema pode ser representado em um grafo orientado $G = (N, E)$ em que N é o conjunto de nós e E o conjunto de arcos. O grafo representa uma rede no qual K produtos diferentes podem ser enviados a partir de certas origens até determinados destinos. Será usada a seguinte nomenclatura: $\mathbf{b}_k$ é o vetor de suprimento/procura para cada produto k , onde $b_{ki} > 0$ representa o suprimento do produto k no nó i , e $b_{ki} < 0$ representa a procura do produto k no nó i . A matriz $\mathbf{A} \in \mathbb{R}^{m \times n}$ é a matriz de incidência nó-arco do grafo G . O vetor $\mathbf{x}_k = (x_{ka})$, $a \in E$ representa o fluxo do produto k nos arcos da rede, e $\mathbf{x}_v \in \mathbb{R}^n$ é o vetor da variável de folga sendo $\mathbf{x} = (\mathbf{x}_k)_{k=1, \dots, K}$. Considera-se que a matriz $\mathbf{A}$ é de posto completo. Caso isso não aconteça, retira-se qualquer linha dessa matriz de forma que ela fique de posto completo. Supõe-se ainda que a função f é de classe C^2 (possui segunda derivada continua).

O problema em estudo é apresentado em seguida:

$$\text{Minimizar } f(\mathbf{x}) = \sum_{k=1}^K f_k(x) \quad (1)$$

$$\text{sujeito a : } \mathbf{A} \mathbf{x}_k = \mathbf{b}_k, \quad k = 1, \dots, K \quad (2)$$

$$\sum_{k=1}^K \mathbf{x}_k + \mathbf{x}_v = \mathbf{b}_{mc}, \quad (3)$$

$$\mathbf{x}_v \geq \mathbf{0}, \mathbf{x}_k \geq \mathbf{0}, \quad k = 1, \dots, K. \quad (4)$$

A relação (1) é a função objectivo, a minimizar. A igualdade (2) é a restrição de conservação de fluxo na rede enquanto a (3) é conhecida como a restrição de capacidade sendo $\mathbf{b}_{mc} \in \mathbb{R}^n$. A restrição (4) expressa o facto de os fluxos serem não negativos.

Existem poucos trabalhos na literatura relacionados a este tipo de problema. Entre eles destaca-se o trabalho de Lawphongpanich [2000] que usa uma decomposição de Dantzig-Wolfe. A maioria dos trabalhos neste assunto trata com funções não lineares, mas que sejam separáveis. Entre eles pode-se citar os trabalhos de Goffin, *et al.* [1996] e Ouorou, *et al.* [2000], onde a optimização é realizada em função dos fluxos nos caminhos. O presente trabalho difere dos citados por fazer a optimização em função dos fluxos nos arcos.

Outros trabalhos que tratam de problemas relacionados a este foram apresentados por Babonneau e Vial [2005] que apresentam um método que combina plano de corte com a ideia do centro analítico. No que diz respeito às aplicações, Nagurney [2006] apresenta diversos modelos de equilíbrio para redes de transporte, de telecomunicações e de economia e finanças, formulados como um problema de inequações variacionais, em que no caso geral, os custos no arco da rede não dependem exclusivamente do fluxo nesse arco. Outras aplicações podem ser encontradas no trabalho de Nagurney [2006].

Quando o problema de optimização correspondente é formulado em função de fluxo nos arcos, o método tradicional é usar o algoritmo de Frank-Wolfe, ou seu caso melhorado. Também existe o método alternativo simplicial e de geração de colunas, dentre outros. Para maiores detalhes destes métodos, consultar o trabalho de Migdalas [2006].

Este trabalho é organizado da seguinte forma. Na secção 2 descreve-se brevemente o método de pontos interiores primal-dual. Na secção 3 desenvolve-se uma especialização do método de pontos interiores para o problema não linear de fluxo e a formação do respectivo sistema aumentado e sua solução. Na secção 4 apresentam-se os resultados computacionais para redes de diferentes dimensões e um número variados de produtos para algumas funções de custo não linear e não separáveis. Finalmente na última secção são apresentadas as conclusões.

2 Método de pontos interiores primal-dual

O método de pontos interiores, iniciado pelo algoritmo de programação linear de Karmarkar [1984] teve grande sucesso computacional por causa de sua complexidade polinomial. Este método foi posteriormente aplicado em programação não linear. As vantagens deste método adequam-no, principalmente, à resolução de problemas de grande porte. Para uma melhor compreensão da teoria de fluxo em rede, pode-se consultar o livro de Ahuja, et al. [1993]. Para a teoria de pontos interiores primal-dual, analisar o livro de Wright [1997].

Apresenta-se a seguir uma breve descrição do método de pontos interiores primal-dual para resolver o seguinte problema de programação não linear com restrições lineares:

$$\begin{aligned} &\text{Minimizar } f(\mathbf{x}) \\ &\text{sujeito a: } \mathbf{E}\mathbf{x} = \mathbf{b} \\ &\mathbf{x} \geq \mathbf{0}, \end{aligned} \quad (5)$$

onde $\mathbf{x} \in \mathbb{R}^q$, $\mathbf{b} \in \mathbb{R}^p$ e $\mathbf{E}$ é uma matriz, $\mathbf{E} \in \mathbb{R}^{p \times q}$, $p < q$ e de posto completo. Assume-se que a função objectivo $f: \mathbb{R}^q \rightarrow \mathbb{R}$ possui segunda derivada contínua e conhecem-se a primeira e segunda derivadas da função objectivo.

A função de Lagrange associada ao problema (5) é:

$$L(\mathbf{w}) = L(\mathbf{x}, \mathbf{y}, \mathbf{z}) = f(\mathbf{x}) + (\mathbf{E}\mathbf{x} - \mathbf{b})^t \mathbf{y} - \mathbf{x}^t \mathbf{z}$$

onde o multiplicador de Lagrange $\mathbf{y} \in \mathbb{R}^p$ corresponde às restrições de igualdades e $\mathbf{z} \in \mathbb{R}_+^q$ corresponde às restrições de desigualdades. As variáveis $\mathbf{x}$ e $\mathbf{z}$ são denominadas de variáveis primais, e a variável $\mathbf{y}$ é denominada de variável dual.

As condições necessárias de primeira ordem de optimalidade de Karush-Kuhn-Tucker (KKT), ver El-Bakry *et al.* (1996), para a solução do problema (5) são dadas por:

$$\mathbf{F}(\mathbf{w}) = \begin{pmatrix} \nabla_{\mathbf{x}} L(\mathbf{w}) \\ \mathbf{E}\mathbf{x} - \mathbf{b} \\ \mathbf{X}\mathbf{Z}\mathbf{e} \end{pmatrix} = \mathbf{0} \quad (6a)$$

$$\mathbf{x} \geq \mathbf{0}, \quad \mathbf{z} \geq \mathbf{0}. \quad (6b)$$

sendo:

$$\begin{aligned} \nabla_{\mathbf{x}} L(\mathbf{w}) &= \nabla_{\mathbf{x}} f(\mathbf{x}) + \mathbf{E}^t \mathbf{y} - \mathbf{z}, \\ \mathbf{X} &= \text{diag}(x_1, x_2, \dots, x_q), \\ \mathbf{Z} &= \text{diag}(z_1, z_2, \dots, z_q), \\ \mathbf{e} &= (1, 1, \dots, 1)^t \in \mathbb{R}^q. \end{aligned}$$

Nesta formulação $\nabla_{\mathbf{x}} L(\mathbf{w})$ é o gradiente de $L(\mathbf{w})$, e $\mathbf{X}$ expressa a matriz diagonal tendo como componentes os elementos do vector $\mathbf{x}$. Analogamente, a matriz $\mathbf{Z}$ também é diagonal e os componentes são os elementos do vector $\mathbf{z}$. As equações (6a) são denominadas condições primais, condições duais e condições de complementaridade, respectivamente.

O problema (5) pode ser aproximado introduzindo a função de barreira $\phi_{\mu}(\mathbf{x})$, conforme apresentada no problema de optimização (7):

$$\begin{aligned} &\text{minimizar } \phi_{\mu}(\mathbf{x}) := f(\mathbf{x}) - \mu \sum_{i=1}^q \ln(x_i) \\ &\text{sujeito a } \mathbf{E}\mathbf{x} = \mathbf{b}, \end{aligned} \quad (7)$$

onde $\mu > 0$ é denominado o parâmetro de barreira dado.

Sabe-se que se μ é suficientemente pequeno, o problema (7) é uma aproximação do problema original (5), ver Fiacco e McCormick (1968).

Assim, para $\mu > 0$, as condições de optimalidade de barreira de KKT associadas ao problema (7) são dadas por:

$$F_{\mu}(\mathbf{w}) = \begin{pmatrix} \nabla_x L(\mathbf{w}) \\ E\mathbf{x} - \mathbf{b} \\ XZ\mathbf{e} - \mu\mathbf{e} \end{pmatrix} = \mathbf{0} \quad (8a)$$

$$\mathbf{x} > \mathbf{0}, \quad \mathbf{z} > \mathbf{0}, \quad (8b)$$

Note-se que se $\mu = 0$ e $\mathbf{x} \geq \mathbf{0}$, $\mathbf{z} \geq \mathbf{0}$, então as condições de KKT para o problema (8) coincidem com as condições de KKT do problema original (5). Por isso, a escolha do parâmetro μ tem um papel importante nos métodos de pontos interiores.

O algoritmo primal-dual determina a solução para uma sequência de problemas de barreiras (8) o qual dá lugar a uma sequência de parâmetros de barreira μ de tal modo a convergir para zero. Assim, o problema de barreira é resolvido para um valor fixo dado μ , isto é, resolve-se o sistema (8a) usando o método de Newton. Seja essa direcção de Newton $\mathbf{dw} = (\mathbf{dx}, \mathbf{dy}, \mathbf{dz})^t$, obtida através da linearização do sistema (8a), e definida pela solução do sistema de equações lineares:

$$\begin{pmatrix} \nabla_x^2 L(\mathbf{w}) & E^t & -I \\ E & 0 & 0 \\ Z & 0 & X \end{pmatrix} \cdot \begin{pmatrix} d\mathbf{x} \\ d\mathbf{y} \\ d\mathbf{z} \end{pmatrix} = - \begin{pmatrix} \nabla_x L(\mathbf{w}) \\ E\mathbf{x} - \mathbf{b} \\ XZ\mathbf{e} - \mu\mathbf{e} \end{pmatrix} \quad (9)$$

onde $\nabla_x^2 L(\mathbf{w})$ é a matriz Hessiana de $L(\mathbf{w})$. Neste caso, tem-se $\nabla_x^2 L(\mathbf{w}) = \nabla^2 f(\mathbf{x})$, sendo $\nabla^2 f(\mathbf{x})$ a matriz Hessiana de $f(\mathbf{x})$. Assume-se que a matriz do sistema (9) é não singular. Essa matriz também é conhecida por matriz **KKT**.

A solução de um sistema linear deste tipo é necessária em cada iteração de um método de pontos interiores. Este, em termos computacionais, é um passo crucial do método. Esse sistema linear pode ser representado como um sistema simétrico indefinido, cuja solução pode ser obtida usando o algoritmo do gradiente conjugado com um pré-condicionador indefinido.

Seja na iteração k , $\mathbf{dw}_k = (\mathbf{dx}_k, \mathbf{dy}_k, \mathbf{dz}_k)^t$ solução do sistema (9), obtida por algum processo de resolução. Em uma nova iteração, determina-se um novo ponto $\mathbf{w}_{k+1} = (\mathbf{x}_{k+1}, \mathbf{y}_{k+1}, \mathbf{z}_{k+1})^t$ o qual é dado por:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{x}_k + \alpha_k \mathbf{dx}_k \\ \mathbf{y}_{k+1} &= \mathbf{y}_k + \alpha_k \mathbf{dy}_k \\ \mathbf{z}_{k+1} &= \mathbf{z}_k + \alpha_k \mathbf{dz}_k, \end{aligned} \quad (10)$$

sendo α_k o tamanho de passo determinado por um procedimento apropriado de busca de linha. Devido a que o sistema de restrições do problema (5) é linear, usa-se um esquema similar ao usado em programação linear para determinar α_k :

$$\alpha_{k\max} = \min\{ \min_i\{-(\mathbf{x}_k)_i / (\Delta\mathbf{x}_k)_i \mid (\Delta\mathbf{x}_k)_i < 0\}, \min_i\{-(\mathbf{z}_k)_i / (\Delta\mathbf{z}_k)_i \mid (\Delta\mathbf{z}_k)_i < 0\} \} \quad (11)$$

$$\alpha_k = \beta \alpha'_k, \quad \alpha'_k = \min\{ \alpha_{k\max}, 1 \} \quad (12)$$

para $\beta \in (0,1)$ e próximo de 1.

Com este novo ponto $\mathbf{w}_{k+1}$, actualiza-se o parâmetro de barreira μ segundo certas regras, forma-se um novo sistema (9) o qual é resolvido usando qualquer processo de solução, e o processo continua até que um critério de paragem do algoritmo de pontos interiores seja satisfeito.

O método de pontos interiores primal-dual pode ser descrito da seguinte maneira: Sejam dados os vectores $\mathbf{x} \in \mathbb{R}^q$, $\mathbf{z} \in \mathbb{R}^q$ tal que $(\mathbf{x}, \mathbf{z}) > \mathbf{0}$ e um parâmetro de barreira $\mu > 0$. Determina-se a direcção de Newton $(\mathbf{dx}, \mathbf{dy}, \mathbf{dz})$ resolvendo o sistema linear equivalente ao sistema da relação (9). Escolhe-se um passo α , actualiza-se um novo ponto usando a relação (10). Finalmente um novo parâmetro de barreira $\mu > 0$ é determinado e o processo continua até que um certo critério de paragem do algoritmo primal-dual seja satisfeito.

3 Fluxo em rede para múltiplos produtos e o método primal-dual

Nesta secção aplicar-se-á o método de pontos interiores primal-dual apresentado na secção anterior para o caso de múltiplos produtos definido no problema (1)-(4).

Usando o segundo membro do sistema (9), define-se:

$$\begin{aligned}\xi_c &= \nabla_{\mathbf{x}} L(\mathbf{w}) = \nabla f(\mathbf{x}) + \mathbf{E}^t \mathbf{y} - \mathbf{z}, \\ \xi_b &= \mathbf{E}\mathbf{x} - \mathbf{b}, \\ \xi_\mu &= \mathbf{X}\mathbf{Z}\mathbf{e} - \mu\mathbf{e}.\end{aligned}$$

Uma forma de solução do sistema (9) consiste em eliminar a variável $\mathbf{dz}$ desse sistema tal que:

$$\mathbf{dz} = -\mathbf{X}^{-1} (\xi_\mu + \mathbf{Z} \mathbf{dx}), \quad (13)$$

e assim tem-se o seguinte sistema simétrico indefinido, também denominado de sistema aumentado, ver Wright [1997]:

$$\begin{bmatrix} \nabla^2 f(\mathbf{x}) + \mathbf{X}^{-1}\mathbf{Z} & \mathbf{E}^t \\ \mathbf{E} & \mathbf{0} \end{bmatrix} \begin{pmatrix} d\mathbf{x} \\ d\mathbf{y} \end{pmatrix} = - \begin{pmatrix} \xi_c + \mathbf{X}^{-1}\xi_\mu \\ \xi_b \end{pmatrix} \quad (14)$$

Por facilidade de notação, o sistema (14) pode ser escrito da seguinte forma:

$$\begin{bmatrix} \mathbf{D} & \mathbf{E}^t \\ \mathbf{E} & \mathbf{0} \end{bmatrix} \begin{pmatrix} d\mathbf{x} \\ d\mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{pmatrix} \quad (15)$$

onde

$$\begin{aligned}\mathbf{v}_1 &= -(\xi_c + \mathbf{X}^{-1}\xi_\mu), \\ \mathbf{v}_2 &= -\xi_b \\ \mathbf{D} &= \nabla^2 f(\mathbf{x}) + \mathbf{X}^{-1}\mathbf{Z}.\end{aligned}$$

A matriz (15) é simétrica mas indefinida e algoritmos para realizar a decomposição de tal matriz são mais complicados. Por outro lado, este sistema tem vantagens quanto à estabilidade e flexibilidade. Para uma melhor compreensão da teoria de sistemas indefinidos e algoritmos de solução, consultar os livros de Wright [1997] e Nocedal e Wright [1999].

Um dos maiores esforços computacionais requerido no método primal-dual é determinar a solução do sistema linear (9), agora expresso como o sistema simétrico indefinido (14) ou (15).

Matrizes da forma (15) frequentemente aparecem na solução de problemas em muitas áreas, por exemplo na formulação de elementos finitos em problemas de equações diferenciais, em dinâmica de fluidos, etc. Para uma extensa revisão deste assunto, ver Benzi, *et al.* [2005].

Um método directo para a solução, como o de decomposição triangular esparsa de Bunch e Parlett [1971] pode ter alto custo computacional em termos de tempo de processamento e espaço de armazenamento. Este problema é agravado no caso de problemas de grande porte, mesmo que os coeficientes da matriz sejam esparsos e esta esparsidade seja explorada. Por outro lado, a matriz associada a (15) é indefinida, já que possui autovalores positivos e negativos. Portanto, o sistema (15) não pode ser resolvido usando métodos interactivos, como o do gradiente conjugado, porque pode fracassar, a menos que um bom condicionador seja empregado. Por outro lado, existem diversos métodos interactivos que usam bons condicionadores para determinar a solução do sistema (15). Entre eles, citam-se os trabalhos de Keller *et al.* [2000], de Golub *et al.* [2001], de Bergamaschi *et al.* [2004], Luksan *et al.* [2005], de Bai e Wang [2006], de Dollar e Wathen [2006]. No presente trabalho, usar-se-á o algoritmo do gradiente conjugado pré-condicionado de Durazzi e Ruggiero [2003]. Ele promove a combinação com um pré-condicionador indefinido apropriado de tal forma a acelerar a convergência para a solução (menor tempo computacional) que o problema original (15). O esquema interactivo pode ser dado como segue:

- (1) Sejam $\mathbf{dx}^0$, $\mathbf{dy}^0$ os vectores iniciais dados;
- (2) $\mathbf{r}_1^0 = \mathbf{v}_1 - \mathbf{D} \mathbf{dx}^0 - \mathbf{E}^t \mathbf{dy}^0$; $\mathbf{r}_2^0 = \mathbf{v}_2 - \mathbf{E} \mathbf{dx}^0$;
- (3) $\mathbf{z}_2^0 = \mathbf{R}^{-1} (-\mathbf{r}_2^0 + \mathbf{E} \mathbf{D}'^{-1} \mathbf{r}_1^0)$; $\mathbf{z}_1^0 = \mathbf{D}'^{-1} (\mathbf{r}_1^0 - \mathbf{E}^t \mathbf{z}_2^0)$;
- (4) $\mathbf{p}_1^0 = \mathbf{z}_1^0$; $\mathbf{p}_2^0 = \mathbf{z}_2^0$;
- (5) $\text{num} = (\mathbf{z}_1^0)^T \mathbf{r}_1^0 + (\mathbf{z}_2^0)^T \mathbf{r}_2^0$;
- (6) para $i = 0, 1, \dots$
- (7) $\mathbf{t}_1^i = \mathbf{D} \mathbf{p}_1^i + \mathbf{E}^t \mathbf{p}_2^i$; $\mathbf{t}_2^i = \mathbf{E} \mathbf{p}_1^i$;
- (8) $\text{den} = (\mathbf{p}_1^i)^T \mathbf{t}_1^i + (\mathbf{p}_2^i)^T \mathbf{t}_2^i$;
- (9) $\alpha = \text{num} / \text{den}$;
- (10) $\mathbf{dx}^{i+1} = \mathbf{dx}^i + \alpha \mathbf{p}_1^i$; $\mathbf{dy}^{i+1} = \mathbf{dy}^i + \alpha \mathbf{p}_2^i$;
- (11) $\mathbf{r}_1^{i+1} = \mathbf{r}_1^i - \alpha \mathbf{t}_1^i$; $\mathbf{r}_2^{i+1} = \mathbf{r}_2^i - \alpha \mathbf{t}_2^i$;
- (12) se um critério de paragem é satisfeito, parar;
- (13) $\mathbf{z}_2^{i+1} = \mathbf{R}^{-1} (-\mathbf{r}_2^{i+1} + \mathbf{E} \mathbf{D}'^{-1} \mathbf{r}_1^{i+1})$;
- (14) $\mathbf{z}_1^{i+1} = \mathbf{D}'^{-1} (\mathbf{r}_1^{i+1} - \mathbf{E}^t \mathbf{z}_2^{i+1})$;
- (15) $\text{den} = \text{num}$;
- (16) $\text{num} = (\mathbf{z}_1^{i+1})^T \mathbf{r}_1^{i+1} + (\mathbf{z}_2^{i+1})^T \mathbf{r}_2^{i+1}$;
- (17) $\beta = \text{num} / \text{den}$;
- (18) $\mathbf{p}_1^{i+1} = \mathbf{z}_1^{i+1} + \beta \mathbf{p}_1^i$; $\mathbf{p}_2^{i+1} = \mathbf{z}_2^{i+1} + \beta \mathbf{p}_2^i$;
- (19) final

Para iniciar o processo exposto, apresentam-se as matrizes $\mathbf{X}$ e $\mathbf{Z}$ para o caso de múltiplos produtos, que são as seguintes matrizes diagonais de bloco:

$$\mathbf{X} = \begin{bmatrix} X_1 & & & \\ & \ddots & & \\ & & X_K & \\ & & & X_v \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} Z_1 & & & \\ & \ddots & & \\ & & Z_K & \\ & & & Z_v \end{bmatrix},$$

sendo cada sub-matriz $\mathbf{X}_k$, $k=1, \dots, K$ uma matriz diagonal com componentes x_{ki} , $i=1, \dots, n$ para cada produto k , e $\mathbf{X}_v$ é a matriz diagonal com componentes da variável de folga $\mathbf{x}_v$. Uma estrutura similar é dada para a matriz $\mathbf{Z}$.

Por outro lado, a matriz de restrições desse problema (1)-(4) é visualizada como segue:

$$\mathbf{E} = \begin{bmatrix} A & O & . & O & O \\ O & A & . & . & O \\ . & . & . & . & . \\ O & . & . & A & O \\ I & I & . & I & I \end{bmatrix} \quad (16)$$

onde cada bloco $\mathbf{A}$ corresponde à matriz de incidência nó-arco para cada produto k , $k = 1, \dots, K$. As matrizes identidades $\mathbf{I}$ correspondem às restrições de capacidade (3).

Finalmente, $\mathbf{D}$ é uma matriz diagonal de blocos dada por:

$$\mathbf{D} = \begin{bmatrix} D_1 & & & & \\ & . & & & \\ & & . & & \\ & & & D_K & \\ & & & & D_v \end{bmatrix} \quad (17)$$

sendo cada sub-matriz $\mathbf{D}_k$, $k = 1, \dots, K$ é uma matriz dada por

$\mathbf{D}_k = [\nabla^2 f_k(\mathbf{x}) + \mathbf{X}_k^{-1} \mathbf{Z}_k]$ e sendo $\mathbf{D}_v = \mathbf{X}_v^{-1} \mathbf{Z}_v$ uma matriz diagonal associada ao vector de folga $\mathbf{x}_v$.

Do esquema interactivo anterior, escolhe-se $\mathbf{D}'$ como a matriz diagonal de blocos da matriz $\mathbf{D}$, isto é, $\mathbf{D}' = \text{diag}(\mathbf{D}_k)$, $k=1, \dots, K$. Por outro lado, a escolha da matriz $\mathbf{R}$ deve ser tal que se aproxime à matriz $\mathbf{E} \mathbf{D}'^{-1} \mathbf{E}^t$ e seja também uma matriz não singular. O formato da matriz $\mathbf{E} \mathbf{D}'^{-1} \mathbf{E}^t$ é dado em (18):

$$\mathbf{E} \mathbf{D}'^{-1} \mathbf{E}^t = \begin{bmatrix} A D_1^{-1} A^t & O & . & O & A D_1^{-1} \\ . & . & . & . & . \\ . & . & . & . & . \\ O & . & . & A D_K^{-1} A^t & A D_K^{-1} \\ D_1^{-1} A^t & . & . & D_K^{-1} A^t & D_v^{-1} + \sum_{k=1}^K D_k^{-1} \end{bmatrix} \quad (18)$$

Este esquema interactivo requer a inversa da matriz $\mathbf{R}$, isto é, a inversa da matriz $\mathbf{E} \mathbf{D}'^{-1} \mathbf{E}^t$. Como isto poderia ser um elemento complicador para o problema em estudo, usa-se a diagonal de blocos da matriz $\mathbf{E} \mathbf{D}'^{-1} \mathbf{E}^t$, isto é, $\mathbf{R} = \text{diag}(\mathbf{E} \mathbf{D}'^{-1} \mathbf{E}^t)$, ou seja $\mathbf{R}$ é a matriz dada por:

$$\mathbf{R} = \begin{bmatrix} A D_1^{-1} A^t & O & . & O & O \\ . & . & . & . & . \\ . & . & . & . & . \\ O & . & . & A D_K^{-1} A^t & O \\ O & . & . & O & D_v^{-1} + \sum_{k=1}^K D_k^{-1} \end{bmatrix}$$

Assim, a inversa da matriz $\mathbf{R}$ é determinada calculando as matrizes inversas de $\mathbf{A}\mathbf{D}'_k^{-1}\mathbf{A}^t$, para $k=1,\dots,K$. Deve ser lembrado que a última matriz de bloco de $\mathbf{R}$ é uma matriz diagonal, portanto a sua matriz inversa é fácil de ser calculada.

Para calcular as matrizes inversas de $\mathbf{A}(\mathbf{D}'_k)^{-1}\mathbf{A}^t$, usa-se o algoritmo AINV de Benzi *et al.* [2000] tal que:

$$(\mathbf{A}(\mathbf{D}'_k)^{-1}\mathbf{A}^t)^{-1} \cong \mathbf{Z}_k \mathbf{P}_k^{-1} \mathbf{Z}_k^t, \quad k = 1, \dots, K.$$

Esse algoritmo é baseado em um processo de uma $(\mathbf{A}(\mathbf{D}'_k)^{-1}\mathbf{A}^t)$ – ortogonalização aplicado aos vectores unitários base $\mathbf{e}_i$. O algoritmo AINV de Benzi *et al.* [2000] é apresentado a seguir para obter a matriz inversa da matriz $\mathbf{B}_k = \mathbf{A}(\mathbf{D}'_k)^{-1}\mathbf{A}^t$, de dimensão m , onde $\mathbf{e}_i$ é o i -ésimo vector unitário base, e por facilidade foi eliminado o sub-índice k :

- (1) Seja $\mathbf{z}_i^{(0)} = \mathbf{e}_i \quad (1 \leq i \leq m)$
- (2) Para $i = 1, 2, \dots, m$ fazer
- (3) $\mathbf{v}_i = (\mathbf{A}\mathbf{D}'^{-1}\mathbf{A}^t) \mathbf{z}_i^{(i-1)}$
- (4) Para $j = i, i+1, \dots, m$ fazer
- (5) $\mathbf{p}_j^{(i-1)} = \mathbf{v}_i^T \mathbf{z}_j^{(i-1)}$
- (6) final
- (7) se $i = m$ ir a (12)
- (8) Para $j = i+1, \dots, m$ fazer
- (9) $\mathbf{z}_j^{(i)} = \mathbf{z}_j^{(i-1)} - (\mathbf{p}_j^{(i-1)} / \mathbf{p}_i^{(i-1)}) \mathbf{z}_i^{(i-1)}$
- (10) final
- (11) final
- (12) Seja $\mathbf{z}_i = \mathbf{z}_i^{(i-1)}$ e $\mathbf{p}_i = \mathbf{p}_i^{(i-1)}$, para $1 \leq i \leq m$. Retornar
 $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_m]$ e $\mathbf{P} = \text{diag}(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_m)$

No passo (3) do algoritmo AINV anterior, pode-se observar que o método realiza o produto matriz-vector $(\mathbf{A}\mathbf{D}'^{-1}\mathbf{A}^t) \mathbf{z}_i$. Para o presente problema de rede, não é necessário armazenar, nem a matriz $\mathbf{A}$, nem a matriz $\mathbf{A}\mathbf{D}'^{-1}\mathbf{A}^t$. Como a matriz $\mathbf{D}'$ é diagonal pode-se aproveitar a estrutura que possuem tais matrizes na aplicação aos problemas de grande porte.

Com as observações anteriores, pode-se aplicar o algoritmo do gradiente conjugado com um pré-condicionador indefinido para o sistema aumentado (15).

Uma outra alternativa, em vez de aplicar o algoritmo AINV, seria determinar a decomposição de Cholesky da matriz $\mathbf{A}\mathbf{D}'^{-1}\mathbf{A}^t$, ver Castro [2000] para o caso da função objectivo ser linear ou (ver passo (13) da solução do sistema aumentado) resolver o sistema de equações:

$$\mathbf{R} \mathbf{z}_2^{i+1} = -\mathbf{r}_2^{i+1} + \mathbf{E}\mathbf{D}'^{-1} \mathbf{r}_1^{i+1},$$

isto é:

$$(\mathbf{A}\mathbf{D}'_k^{-1}\mathbf{A}^t) \mathbf{z}_2^{i+1} = -\mathbf{r}_2^{i+1} + \mathbf{E}\mathbf{D}'^{-1} \mathbf{r}_1^{i+1},$$

e neste caso, usa-se o esquema apresentado em Júdice, *et al.* [2003].

Com a solução do sistema (15), determinou-se $\mathbf{dx}$ e $\mathbf{dy}$. Usando (13) determina-se então $\mathbf{dz}$ e, novamente, um novo ponto interior é calculado usando a relação (10). O processo repete-se até que um critério de paragem do método primal-dual seja satisfeito.

Para o caso em que a função objectivo seja separável, o problema não linear de múltiplos produtos pode ser resolvido usando o sistema normal de equações, em vez do sistema aumentado, ver Torres [2006].

4 Resultados computacionais

Nesta secção, são apresentados os resultados computacionais para determinar a solução do problema não linear de fluxo em rede, de diferentes dimensões, para diversos produtos e diferentes funções de custos não lineares e não separáveis. As experiências foram realizadas usando o método de pontos interiores primal-dual, cujo correspondente sistema linear foi resolvido usando os diferentes métodos de solução discutidos anteriormente.

Todos os testes computacionais foram realizados em um microcomputador PC Duron, plataforma Windows XP, de 512 MB de memória RAM e 1600 MHz de frequência. Os códigos computacionais foram totalmente escritos em FORTRAN. Foi usada dupla precisão para todos os cálculos computacionais.

Na aplicação do método primal-dual, a sequência do parâmetro de barreira $\{\mu_k\}$ deve convergir para zero tão rápido quanto possível. Várias regras foram estudadas para o decrescimento desse parâmetro μ , ver por exemplo os trabalhos de Waltz *et al.* [2005], Wächter *et al.* [2006], Yamashita e Yabe [2005]. Neste trabalho usou-se a estratégia adoptada por Luksan *et al.* [2005] que segundo os autores desenvolve-se bem na prática. A regra é a seguinte:

$$\begin{aligned}\mu &= \sigma \mathbf{x}^t \mathbf{z} / np \\ \text{onde} \\ \sigma &= 0,1 \min \{0,05 (1 - \rho) / \rho, 2\}^3 \\ &\quad \text{e} \\ \rho &= \frac{\min(x_i z_i)}{x^t z / np}\end{aligned}$$

sendo np o número de variáveis.

Foram analisadas redes de diversas dimensões, baseadas na rede básica do trabalho de Nagurney [1984], p. 476. Essa rede básica consiste de 20 nós e 28 arcos e é usada para determinar a solução do problema de equilíbrio de tráfego em rede de transporte. Essa rede básica foi depois estendida para formar redes de grande porte. Para isso, foi implementado um programa especial FORTRAN para determinar a dimensão da nova rede, isto é, o número de arcos e de nós e os dois vectores que determinam a origem e o destino que definem cada arco da rede.

Em todos os testes computacionais do método primal-dual é necessário partir-se de um ponto inicial, não necessariamente viável. Para determinar o tamanho de passo α , foi usado o valor de $\beta = 0,99995$. Neste caso, aceita-se o primeiro passo α , sem usar qualquer função de mérito nem qualquer filtro, como é usado em problemas gerais da programação não linear.

O critério de paragem na solução do respectivo sistema linear simétrico indefinido foi fixado para:

$$\mathbf{r} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} - \begin{bmatrix} D & E^t \\ E & 0 \end{bmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix}$$

tal que $\|\mathbf{r}\| < \epsilon_1$, ϵ_1 uma determinada tolerância ou quando um número máximo de iterações é alcançado, digamos 600 iterações. Já o critério de paragem do método de pontos interiores é quando o valor da função objectivo no ponto actual é suficientemente próximo do valor da função objectivo no ponto anterior, isto é, a diferença relativa absoluta desses valores das funções objectivo é menor ou igual a 10^{-8} e o valor do parâmetro μ é próximo a zero. Nesse caso, verifica-se se $\|F(\mathbf{w})\| \leq \epsilon_2$, para ϵ_2 uma certa tolerância, por exemplo $\epsilon_2 = 10^{-3}$ ou quando

$$(|\text{gap}| / (1 + |\text{gap}|)) \leq \epsilon_3$$

sendo ϵ_3 é uma certa tolerância que depende do número de produtos.

Aqui, “gap” é a diferença da função primal, do problema (5), $f(\mathbf{x})$, e a função dual $d(\mathbf{x}, \mathbf{y}, \mathbf{z})$ dada por:

$$d(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \mathbf{b}^t \mathbf{y} - (\mathbf{x}^t \nabla f(\mathbf{x}) - f(\mathbf{x})).$$

Foram feitas análises para $K=20, 50, 100, 500$ e 1000 produtos. Como cada variável de decisão $\mathbf{x}_k$, $k=1, \dots, K$ tem dimensão igual ao número de arcos, então a dimensão total de todas essas variáveis de decisão $\mathbf{x} = (\mathbf{x}_k)_{k=1, \dots, K}$ é igual ao número de arcos da rede multiplicado pelo número de produtos. Não foram levados em conta o número de componentes da variável de folga $\mathbf{z}$. Assim, por exemplo, o número total de variáveis para a rede de 390 arcos e 210 nós e 1000 produtos, como é um dos casos estudados, é de 390.000 variáveis.

No presente trabalho foram usados dois tipos de funções de custo não linear não separável, encontrados na literatura, indicados a seguir:

$$\text{i) } f(\mathbf{x}_k) = \sum x_{ki}^2 + \sum_{i < j} x_{ki} x_{kj}, \text{ para } \mathbf{x}_k = (x_{ki}) \text{ para produto } k$$

$$\text{ii) } f(\mathbf{x}_k) = \sum x_{ki}^4 + 0,5 \sum x_{ki}^2 + \sum_{i < j} x_{ki} x_{kj}, \text{ para } \mathbf{x}_k = (x_{ki}) \text{ para produto } k.$$

Shi [2004] usou funções deste tipo, mas trabalhou com o caso menos geral de programação convexa linearmente restringido e com um único produto.

A seguir são apresentados os resultados computacionais para a primeira função não linear.

$$\text{Função de custo } f(\mathbf{x}) = \sum x_i^2 + \sum_{i < j} x_i x_j$$

Tabela 4.1 – Rede de 105 arcos e 60 nós

Número de produtos	μ	gap	Número de iterações	Tempo (segundos)
20	10^{-12}	10^{-15}	17	3,24
50	10^{-12}	10^{-15}	15	13,47
100	10^{-12}	10^{-15}	16	17,24
500	10^{-12}	$0,572201 \times 10^{-5}$	9	19,73
1000	10^{-12}	10^{-15}	14	167,92

Tabela 4.2 – Rede de 390 arcos e 210 nós

Número produtos	μ	gap	Número iterações	Tempo (segundos)
20	10^{-12}	10^{-15}	17	42,97
50	10^{-12}	10^{-15}	18	111,81
100	10^{-12}	10^{-15}	15	203,65
500	10^{-12}	10^{-15}	15	1047,56
1000	10^{-12}	$0,174788 \times 10^{-2}$	15	2066,87

A seguir são apresentados os resultados computacionais para a segunda função.

$$\text{Função de custo } f(x) = \sum x_i^4 + 0,5 \sum x_i^2 + \sum_{i < j} x_i x_j$$

Tabela 4.3 – Rede de 105 arcos e 60 nós

Número produtos	μ	gap	Número iterações	Tempo (segundos)
20	10^{-12}	$0,619884 \times 10^{-5}$	9	0,63
50	10^{-12}	$0,762934 \times 10^{-5}$	9	1,50
100	10^{-12}	$0,114439 \times 10^{-4}$	9	2,94
500	10^{-12}	$0,596457 \times 10^{-1}$	9	14,92
1000	10^{-12}	$0,366077 \times 10^{-3}$	8	26,81

Tabela 4.2 – Rede de 390 arcos e 210 nós

Número produtos	μ	gap	Número iterações	Tempo (segundos)
20	10^{-12}	$0,343285 \times 10^{-2}$	25	37,97
50	10^{-12}	$0,555181 \times 10^{-2}$	25	94,63
100	10^{-12}	$0,532985 \times 10^{-1}$	24	182,24
500	10^{-12}	0,211540081	23	881,55
1000	10^{-12}	0,229695103	24	1845,61

Como pode ser observado, o valor do parâmetro de barreira μ é bastante pequeno em todos os problemas de teste, e o número de iterações necessárias para alcançar uma solução aceitável é razoável.

O algoritmo AINV comportou-se muito bem e é robusto para o caso de determinar a inversa da matriz $\mathbf{B}_k = \mathbf{A} (\mathbf{D}'_k)^{-1} \mathbf{A}^t$, $k=1,\dots,K$. Este algoritmo é um dos passos computacionalmente caro no método de pontos interiores primal – dual. Uma outra alternativa seria determinar a decomposição dessas matrizes usando o algoritmo de Cholesky, ver Castro [2000] para o caso do custo da função objectivo ser linear.

Pode-se observar também, que o método de pontos interiores para a maioria dos casos usando a primeira função, o *gap* é nulo. Já para a segunda função, esse *gap* é próximo a zero.

O tempo computacional, como era de esperar-se, aumenta quando o número de produtos também aumenta, mas que é bastante aceitável.. O algoritmo de pontos interiores aplicado para fluxo em rede poderia ainda ser melhorado com refinamentos adicionais.

5 Conclusões

Neste artigo, o método de pontos interiores primal-dual foi desenvolvido para resolver problemas de fluxo em rede para múltiplos produtos. O algoritmo apresentado explora a estrutura da matriz de restrições de tal forma a não precisar armazenar nem a matriz de restrições nem o produto das matrizes envolvidas, o que pode ser útil para problemas de grande porte.

Foram construídas uma série de problemas não lineares não separáveis, para diferentes número de produtos (até 1000 produtos) e diferentes dimensões da rede.

A solução do sistema simétrico indefinido foi determinada usando o algoritmo do gradiente conjugado pré-condicionado através da diagonalização de cada sub-matriz

$\mathbf{D}_k = [\nabla^2 f_k(\mathbf{x}) + \mathbf{X}_k^{-1}\mathbf{Z}_k]$, $k=1,\dots,K$, obtendo-se a matriz diagonal $\mathbf{D}'$. Com isto foi possível aplicar o algoritmo AINV para determinar, aproximadamente, a inversa da matriz

$\mathbf{A}(\mathbf{D}'_k)^{-1}\mathbf{A}^t$. Esse algoritmo AINV mostrou-se robusto e comportou-se bem na prática.

Igualmente pode-se concluir a eficiência do método de pontos interiores para problemas de grande porte, apesar da solução, usando o *gap*, estar próximo à solução exacta. Pode-se aceitar que são erros satisfatórios para problemas de grande porte.

Futuros trabalhos incluirão o desenvolvimento de novos pré-condicionadores mais eficientes para este tipo de aplicação. Também poderão ser incluídos o uso de algoritmos numéricos em paralelo e/ou a comparação com o método de Cholesky.

6 Referências

- Ahuja, A.; Magnanti, T. & Orlin, J. (1993). *Network Flows: Theory, Algorithms and Applications*, Prentice-Hall, Inc. New Jersey.
- Babonneau, F. & Vial, J. (2005). ACCPM with a nonlinear constraint and an active set strategy to solve nonlinear multicommodity flow problems.
http://www.optimization-online.org/DB_HTML/2005/06/1148.html. p. 1-29.
- Bai, Z. & Wang, Z. (2006). Restrictive preconditioners for conjugate gradient methods for symmetric positive definite linear systems. *Journal of Computational and Applied Mathematics*, 187, 202-226.
- Benzi, M.; Cullum, J. & Tuma, M. (2000). Robust approximate inverse preconditioning for the conjugate gradient method. *SIAM Journal on Scientific Computing*, 22, 4, 1318-1332.
- Benzi, M.; Golub, G. & Liesen, J. (2005). Numerical solution of saddle point problems. *Acta Numerica*, 14, 1-137.
- Bergamaschi, L.; Gondzio, J. & Zilli, G. (2004). Preconditioning indefinite systems in interior point methods for optimization. *Computational Optimization and Applications*, 28, 149-171.
- Bunch, J. & Parlett, B. (1971). Direct methods for solving symmetric indefinite systems of linear equations. *SIAM Journal of Numerical Analysis*, 8, 639-655.
- Castro, J. (2000). A specialized interior – point algorithm for multicommodity network flows. *SIAM Journal on Optimization*, 10, 3, 852 -877.
- Dollar, H. & Wathen, A. (2006). Approximate factorization constraint preconditioners for saddle-point matrices. *SIAM Journal on Scientific Computing*, 27, 5, 1555-1572.
- Durazzi, C. & Ruggiero, V. (2003). Indefinitely preconditioned conjugate gradient method for large sparse equality and inequality constrained quadratic problems. *Numerical Linear Algebra with Applications*, 10, 673-688.
- El-Bakry, A. ; Tapia, R. ; Tsuchiya, T. & Zhang, Y. (1996). On the formulation and theory of Newton interior-point method for nonlinear programming. *Journal of Optimization Theory and Applications*, 89, 507-541.
- Fiacco, A. e McCormick, G. (1968). *Nonlinear programming: Seqüencial unconstrained minimization technique*. John Wiley and Sons.
- Goffin, J.; Gondzio, R.; Sarkissian, R. & Vial, J. (1996). Solving nonlinear multicommodity flow problems by the analytic center cutting plane method. *Mathematical Programming*, 76, 131-154.
- Golub, G.; Wu, X. & Yuan, J. (2001). SOR-like methods for augmented systems. *BIT*, 41, 1, 71-85.

- Júdice, J., Patricio, J., Portugal, L., Resende, M. E Veiga, G. (2003). A study of the preconditioners of network interior point methods. *Computational Optimization and Applications*, 24, 5 – 35.
- Karmarkar, N. (1984). A new polynomial time algorithm for linear programming. *Combinatorica*, 4, 373-395.
- Keller, C.; Gould, N. & Wathen, A. (2000). Constraint preconditioning for indefinite linear systems. *SIAM Journal Matrix Analysis and Applications*, 21, 4, 1300-1317.
- Lawphongpanich, S. (2000). Simplicial with truncated Dantzig-Wolfe decomposition for nonlinear multicommodity network flow problems with side constraints. *Operations Research Letters*, 26, 33-41.
- Luksan, L.; Matonoba, C. & Vlcek, J. (2005). Interior point methods for large-scale nonlinear programming, *Optimization Methods and Software*, 20, 4-5, 569-582.
- Migdalas, A. (2006) . Nonlinear Programming in Telecommunications. In: *Handbook of Optimization in Telecommunications* [edited by M. Resende and P. Pardalos], Springer, 27-66.
- Nagurney, A. (1984). Comparative tests of multimodal traffic equilibrium methods, *Transportation Research*, 18B, 469-485.
- Nagurney, A. (2006). Supernetworks. In: *Handbook of Optimization in Telecommunications* [edited by M. Resende and P. Pardalos], Springer, 1073-1119.
- Nocedal, J. & Wright, S. (1999). *Numerical Optimization*, Springer-Verlag, New York.
- Ouorou, A.; Mahey, P. & Vial, J. (2000). A survey of algorithms for convex multicommodity flow problems. *Management Science*, 46, 1, 126-147.
- Shi, Y. (2004). A projected-steepest-descent potential-reduction algorithm for convex programming problems. *Numerical Linear Algebra with Applications*, 11, 883-893.
- Torres, L. (2006). Modelo não linear de fluxo em rede para multi-produtos. CD-ROM. XIII CLAIO – Congreso Latino-Iberoamericano de Investigación Operativa, 27 – 30 de Noviembre de 2006, Montevideo, Uruguay.
- Wächter A. & Biegler, L. (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming, *Mathematical Programming, Serie A*, 106, 1, 25-57.
- Waltz, R.; Morales, J.; Nocedal, J. & Orban, D. (2005). An interior algorithm for nonlinear optimization that combines line search and trust region steps. *Mathematical Programming, Serie A*, 107, 3, 391-408.
- Wright, S. (1997). *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia, Pa.
- Yamashita, H. & Yabe, H. (2005). Quadratic convergence of a primal -dual interior point method for degenerate nonlinear optimization problems, *Computational Optimization and Applications*, 31, 123-143.

Uso da Simulação de Monte Carlo e da Curva de Gatilho na Avaliação de Opções de Venda Americanas

Javier Gutiérrez Castro [†]
Tara K. Nanda Baidya [†]
Fernando A. Lucena Aiube [†]

[†] Departamento de Engenharia Industrial (DEI)
Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)
Rua Marquês de São Vicente 225, C.E.P. 22453-900 – Gávea,
Rio de Janeiro, Brasil.
javiergc@aluno.puc-rio.br
baidya@ind.puc-rio.br
aiube@ind.puc-rio.br

Abstract

In 1973 Black and Scholes [2] published their article on the valuation of European option. Since then, there have been many works extending this work in many directions. One such direction is the valuation of American options. On the matter, no exact analytical formula has been developed yet. In stead, numerical methods have been used in their valuations. Monte Carlo simulation has been the method which has become more and more popular among researchers in this field.

The threshold curve method, used by Grant, Vora and Weeks [7] to value American options, is calculated through Monte Carlo simulation. This is the traditional method used in finance. We propose to modify the methodology of Ibáñez and Zapatero [9], which also uses the threshold curve, to obtain a more efficient and more accurate method than that of Grant, Vora and Weeks [7]. In this work, the described procedures and numerical tests are focused in American Put Options.

Resumo

Em 1973 Black e Sholes [2] publicaram um seminal artigo no qual, pela primeira vez, se avaliava analiticamente uma opção do tipo européia. Desde então, tem surgido uma grande quantidade de trabalhos estendendo esse artigo para diversas áreas e aplicações. O apreçamento de opções americanas é uma das vertentes. Sobre isso, não existe até o momento uma fórmula analítica que permita calcular de maneira exata o preço de uma opção americana. Portanto, métodos numéricos vêm sendo utilizados nesta tarefa. Entre eles, o método da simulação de Monte Carlo tem se tornado o de maior popularidade entre os pesquisadores dessa área.

A curva de gatilho, método utilizado por Grant, Vora e Weeks [7] para avaliar opções americanas, é calculada através da simulação de Monte Carlo. Este é o método tradicional utilizado em Finanças. Nossa proposta consiste em modificar a metodologia desenvolvida por Ibáñez e Zapatero [9], que também calcula a curva de gatilho, para obter um método mais eficiente e mais preciso do que o apresentado por Grant, Vora e Weeks [7]. Neste trabalho, os procedimentos descritos e os testes numéricos realizados, foram orientados para opções de venda americanas.

Keywords: American Put Options, Monte Carlo Simulation, Threshold Curve.

Title: Utilization of Monte Carlo Simulation and Threshold Curve to Value American Put Options.

1 Introdução

Em 1973 Black e Scholes [2] apresentaram um artigo seminal em que foi apreçada analiticamente uma opção financeira do tipo européia. Até então, este era um grande desafio para os pesquisadores nessa área. Mas, no caso de opções americanas, não existe nenhum procedimento analítico que permite avaliá-las com exatidão. Inicialmente a literatura propôs métodos que aproximam analiticamente o valor da opção americana, como é o caso de Barone-Adesi e Whaley [1]. Por outro lado, o método binomial, utilizado por Cox, Ross e Rubinstein [5] para avaliar opções européias, resulta também útil no apreçamento de opções americanas. Mas todos esses procedimentos têm a desvantagem de ficarem restringidos a uma série de condições, tais como ter no máximo uma variável estocástica e o ativo subjacente ser modelado por um processo estocástico de Movimento Geométrico Browniano. Estas restrições inviabilizam sua aplicação em problemas complexos.

Em relação à técnica de simulação de Monte Carlo, existe um histórico relativamente recente no que diz respeito à sua utilização no apreçamento de opções americanas. As primeiras abordagens propostas foram realizadas por Boyle et al. [3] e Broadie e Glasserman [4], tornando-se referências para os trabalhos que posteriormente desenvolveram-se nessa área de pesquisa. Estas metodologias tentam aproximar o preço da opção usando os conceitos da programação dinâmica, isto é, utilizando procedimentos recursivos de cálculo (de trás para frente) a partir de simulações dos valores do ativo subjacente ao longo do tempo.

Neste contexto, um método que tem tido uma ampla difusão devido a sua facilidade de aplicação é o desenvolvido por Grant, Vora e Weeks [7]. Este método tem a particularidade de calcular previamente a curva de gatilho (*threshold curve*) ou curva de preços críticos de exercício, sendo uma das primeiras metodologias que introduziram este conceito no apreçamento de opções americanas por simulação.

A determinação dos instantes ótimos de exercício da opção (ou preços críticos de exercício), ao longo de todo o período de maturação, define o que na literatura se conhece como curva de gatilho. A curva de gatilho é um conceito bastante útil, sobretudo quando são analisadas opções reais. O tratamento de opções reais é análogo ao de opções financeiras, em que o ativo subjacente passa a ser um ativo real, por exemplo, o valor de um projeto. Assim, é possível por meio desta curva identificar o período adequado para realizar um investimento de valor K (preço de exercício). Este investimento ocorrerá quando o valor do projeto atinja um nível igual ou superior àquele definido na curva. Caso exista a possibilidade de abandonar o projeto (obtendo um valor de recuperação K), a opção de abandono deverá ser exercida assim que o valor do projeto seja igual ou inferior àquele definido na curva. Estes são alguns exemplos da utilidade prática da curva de gatilho.

As opções financeiras ou reais do tipo americana podem ser exercidas ao longo do intervalo de tempo que vai de $t_0 = 0$, até $t_N = T$ (tempo de maturidade). O prazo até a maturidade ou vencimento é dividido em N intervalos, sendo que a opção pode ser exercida em qualquer um desses intervalos. Ao utilizar a simulação de Monte Carlo, a modelagem da opção americana assemelha-se a uma opção bermuda, a qual se caracteriza por ter mais de uma data de exercício até o vencimento. Quanto mais intervalos discretos forem considerados no intervalo $[t_0; T]$, melhor será o modelo que descreve o comportamento real de uma opção americana (que se exerce em tempo contínuo e não discreto).

Uma metodologia alternativa é a desenvolvida por Ibáñez e Zapatero [9]. Como feito por Grant, Vora e Weeks [7], eles também determinam primeiramente a curva de gatilho.

A novidade que trazem os autores é uma maneira diferente de calcular os preços críticos de exercício, o que se mostra muito eficiente.

Neste artigo é analisado com especial ênfase o algoritmo de Ibáñez e Zapatero [9]. Por outro lado, foram feitas modificações nesse algoritmo que aprimoram a construção da curva de gatilho. Isto permitiu melhorar a precisão dos resultados usando os testes originais propostos pelos autores. O enfoque dado no trabalho está na avaliação de opções de venda americanas considerando que o ativo subjacente é a única variável estocástica. São comparadas três metodologias: o algoritmo de Grant, Vora e Weeks [7], Ibáñez e Zapatero [9] e Ibáñez e Zapatero modificado (aqui proposto).

O artigo está assim organizado: a seção 2 descreve o algoritmo desenvolvido por Grant, Vora e Weeks [7] aplicando-o a uma série de testes numéricos; a seção 3 apresenta o algoritmo de Ibáñez e Zapatero [9]; a seção 4 aplica o algoritmo descrito na seção 3 aos mesmos testes numéricos da seção 1, detalhando as modificações que serão realizadas na implementação do algoritmo; a seção 5 compara os resultados obtidos com as diferentes metodologias; e a seção 6 apresenta as conclusões e considerações finais.

2 O Método de Grant, Vora e Weeks

2.1 Definição de Curva de Gatilho

O algoritmo de Grant, Vora e Weeks [7] determina primeiramente a curva de gatilho ou fronteira de exercício ótima. Esta é formada pelo conjunto de pontos S_t^* nos quais o valor de manter viva a opção (esperar) é igual ao valor intrínseco (exercer), onde S_t é o valor do ativo subjacente, que é uma variável estocástica, e o sobrescrito * indica que S_t é o preço crítico de exercício no instante t .

Denote-se por $P_t(S_t, K)$ o preço da opção de venda, e $I(S_t, K) = K - S_t$ é o valor intrínseco ou valor da opção de venda quando é exercida, sendo K o preço de exercício. Portanto, na curva de gatilho se dá que $P_t(S_t^*, K) = I(S_t^*, K)$.

Dado um instante inicial t_0 , e um instante T (maturidade da opção ou prazo máximo de exercício), pode-se subdividir o horizonte de tempo $T - t_0$ em N intervalos, com datas de exercício discretas em $\{t_1, t_2, \dots, t_N = T\}$. Em algum instante de tempo t_n , assume-se que o exercício da opção é ótimo se $S_{t_n} \leq S_{t_n}^*$.

Seja r a taxa livre de risco de curto prazo, e Q a medida de probabilidade martingale. Logo, em alguma data t_n ($n \in \{N-1, N-2, \dots, 1\}$), o preço da opção é calculado por:

$$P_{t_n}(S_{t_n}, K) = E_{t_n}^Q [e^{-\int_{t_n}^{\tau^*} r_s ds} I(S_{\tau^*}, K)] \quad (1)$$

Onde $\tau^* \in \{t_{n+1}, t_{n+2}, \dots, T\}$ é o chamado tempo ótimo de parada, definido como o primeiro t_{n+i} no qual $S_{t_{n+i}} \leq S_{t_{n+i}}^*$; noutro caso $\tau^* = \infty$. Em outras palavras, é o primeiro instante em que o preço do ativo “S” fica abaixo da curva de gatilho.

Assumindo que S_t segue um Movimento Geométrico Browniano (MGB), sob a medida martingale, escreve-se: $dS_t/S_t = \alpha dt + \sigma dZ$, onde $\alpha = r - q$ é o *drift* ou tendência neutra ao risco, q é a taxa de dividendos, σ é a volatilidade do preço do ativo, e, $dZ = \varepsilon_t(dt)^{1/2}$ é o incremento do processo padrão de Wiener com $\varepsilon_t \sim \text{NID}(0,1)$. Dado que S_{t_n} é uma variável estocástica a ser gerada por simulação de Monte Carlo, é conveniente utilizar um MGB discretizado, da seguinte forma:

$$S_{t_n+\Delta t} = S_{t_n} \exp \left[\left(\alpha - \sigma^2 / 2 \right) \Delta t + \sigma (\Delta t)^{1/2} \varepsilon_{t_n+\Delta t} \right] \quad (2)$$

Onde Δt é o intervalo de tempo entre t_n e t_{n+1} . Desta maneira, é possível discretizar a equação (1) sob a mesma medida de probabilidade, resultando na seguinte equação:

$$P_{t_n}(S_{t_n}, K) = \frac{1}{M} \sum_{a=1}^A e^{-(\tau_a^* - t_n)r} (K - S_{\tau_a^*}) \quad (3)$$

Onde A é o número total de caminhos Brownianos, dentre os M simulados a partir de um instante de tempo t_n , nos quais num instante de tempo (o primeiro) t_{n+i} tem-se $S_{t_{n+i}} \leq S_{\tau_{n+i}}^*$ (podem existir caminhos que em nenhum t_{n+i} aconteça tal situação). Por conseqüência existem “ A ” períodos τ^* , identificados por τ_a^* ($a=1, \dots, A$) correspondentes a um determinado $S_{\tau_a^*}^*$ em que a mencionada restrição é satisfeita.

2.2 Descrição do algoritmo de Grant, Vora e Weeks

O método de Grant, Vora e Weeks [7] foi um dos primeiros em utilizar o conceito de curva de gatilho, e serviu de base para o algoritmo de Ibáñez e Zapatero [9], que será abordado posteriormente. Nascimento [10] apresenta de forma bastante ampla esta metodologia, inclusive com diversas experimentações numéricas.

Para calcular os preços críticos de exercício, o algoritmo utiliza a condição de valor ótimo (*value matching condition*). A condição é ilustrada a seguir para uma opção de venda americana P_t sobre um ativo base S_t e com preço de exercício K :

$$P_t(S_t^*, K) = K - S_t^* \quad (4)$$

Onde S_t^* representa o preço crítico de exercício do ativo base no instante t . Na maturidade da opção, o valor crítico é dado pelo preço de exercício K :

$$P_t(S_T, K) = \max(K - S_T, 0) \Rightarrow S_T^* = K \quad (5)$$

No instante de tempo $t=T$, o preço crítico de exercício S_t^* é igual ao preço de exercício da opção, representado pela letra K . Por definição, o preço da opção de compra é a diferença entre o preço do ativo S_t e o preço de exercício K (ou seja $S_t - K$). Para a opção de venda temos $K - S_t$. Se no tempo $t=T$ (último período para decidir o exercício) S_t é igual a K , o valor da opção seria zero, portanto, exercer ou não a opção forneceria o mesmo resultado (zero para essa data). Assim, o preço crítico de exercício S_T^* é igual a K .

Na equação (5), T é a data de vencimento da opção. Já para um instante t qualquer antes do vencimento, a decisão ótima depende do conhecimento prévio do preço crítico imediatamente posterior no futuro, $S_{t+\Delta t}^*$, como pode ser visto a seguir:

$$P_t(S_t^*, K) = \max(K - S_t^*, e^{-r\Delta t} E_t[P_{t+\Delta t}(S_{t+\Delta t}^*, K)]) \quad (6)$$

Onde o último termo à direita constitui o valor de continuação, ou o valor de manter a opção viva. E_t é o valor esperado condicional à informação disponível no instante t e Δt é o tamanho de uma discretização do tempo.

A dificuldade surge quando são calculados os preços críticos que dependem de preços futuros. Como a informação futura é desconhecida no instante atual, utiliza-se a simulação de Monte Carlo como auxílio no cálculo destes valores.

Para ilustrar o processo de cálculo do preço crítico de exercício, supõe-se que o valor a ser determinado é $S_{T-\Delta t}^*$, ou seja, o preço crítico de exercício do instante imediatamente anterior ao vencimento da opção. Primeiramente, adota-se como condição inicial $S_{T-\Delta t} = S_T$. A partir de $S_{T-\Delta t}$, simula-se valores para S_T e P_T . Obtém-se então o valor de P_T , utilizando-se a média das simulações executadas. A seguir, verifica-se se $S_{T-\Delta t}^*$ satisfaz a condição de valor ótimo dada pela equação (6), o que significaria escrever:

$$K - S_{T-\Delta t}^* = e^{-r\Delta t} E_{T-\Delta t}[P_T(S_T^*, K)] \quad (7)$$

Caso a condição acima não seja satisfeita, incrementa-se $S_{T-\Delta t}$ de um valor $-\Delta S$ (um valor pequeno) e simula-se novamente valores para S_T e P_T , repetindo-se o procedimento até que o valor crítico $S_{T-\Delta t}^*$ seja encontrado. A curva de gatilho é obtida repetindo-se o procedimento acima, recursivamente, até o instante inicial.

2.3 Resumo do Algoritmo de Grant, Vora e Weeks

O algoritmo Grant, Vora e Weeks [7], para o cálculo do preço de uma opção de venda americana, pode ser resumido nos seguintes passos:

1. Discretiza-se a vida útil da opção em $N=T/\Delta t$ partes, onde Δt é o tamanho de cada intervalo, e adota-se a condição terminal $S_T^* = K$.

2. No instante $T-\Delta t$, adota-se como aproximação de $S_{T-\Delta t}$ um valor igual ou próximo de S_T^* . Em seguida, utiliza-se a simulação de Monte Carlo para se obter diferentes valores de S_T e, conseqüentemente, de P_T . O valor de P_T é calculado através da média das simulações executadas.

3. Verifica-se se a condição de valor ótimo, expressa pela equação (7), é satisfeita. Caso afirmativo, inicia-se o próximo passo. Caso contrário, incrementa-se $S_{T-\Delta t}$ de um valor $-\Delta S$ e repete-se o passo anterior.

4. Repete-se o segundo e o terceiro passos para os instantes anteriores, até chegar a t_0 . Para calcular o preço da opção de venda devem-se simular caminhos em todos os instantes posteriores ao momento avaliado, e aplicando a equação (3), obtém-se esse valor.

5. Uma vez obtida a curva de gatilho, o preço da opção de venda é obtido através da aplicação da equação (3), a partir de simulações do preço do ativo subjacente. Detalhando passo a passo, faz-se:

- (i) Simular uma grande quantidade de caminhos (M) a partir do valor inicial do ativo subjacente $S_{t_0} = S_0$. As simulações são feitas em intervalos de tempos discretos $\{t_1, t_2, \dots, t_N = T\}$.

- (ii) Para cada caminho simulado, no primeiro instante de tempo t_n ($n = 1, 2, \dots, N$) em que o valor do ativo S_{t_n} seja menor ou igual ao ponto crítico da curva de gatilho $S_{t_n}^{(*)}$, será exercida a opção, sendo o preço da opção de venda em t_n o valor intrínseco: $K - S_{t_n}$. A seguir, desconta-se este preço com a taxa livre de risco: $P_m = e^{-(t_n - t_0)r} \cdot (K - S_{t_n})$, onde P_m ($m \in \{1, 2, \dots, M\}$) representa o preço da opção de venda em t_0 para um caminho simulado dentre as 'M' realizações. É provável que existam caminhos nos quais, em todo momento, os preços fiquem acima da curva de gatilho; para estes casos $P_m = 0$, naturalmente.
- (iii) O preço da opção de venda será a média aritmética de todos os P_m 's:

$$P_{t_0}(S_0, K) = 1/M \sum_{m=1}^M P_m \quad (8)$$

A Figura 1 apresenta um exemplo ilustrativo que calcula o preço de uma opção de venda com três caminhos simulados a partir de um preço inicial do ativo $S_0 = 35$, em um horizonte de tempo dividido em seis períodos.

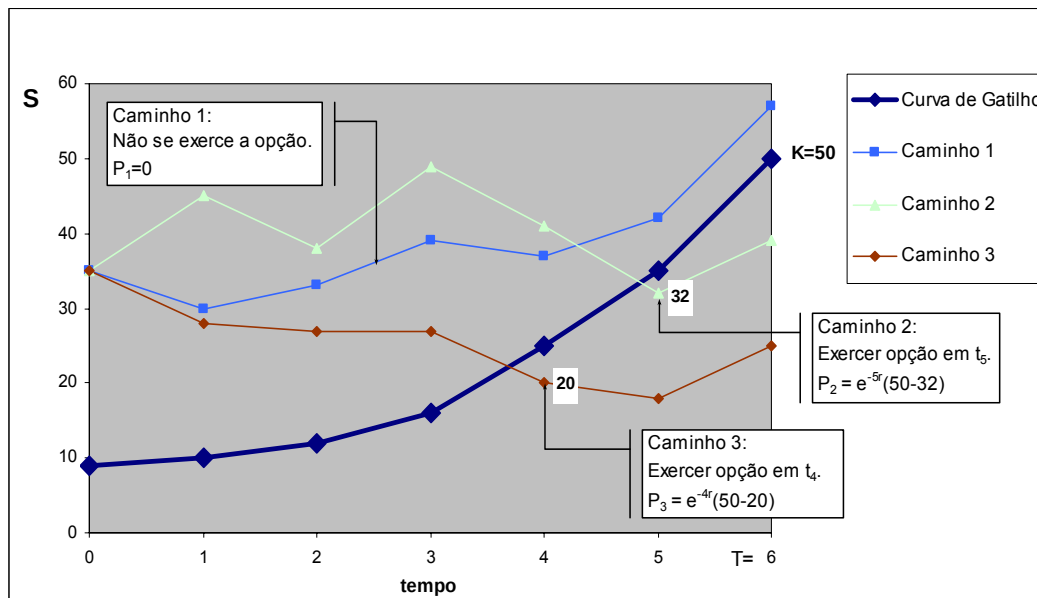


Figura 1: Exemplo de Cálculo do Preço da Opção de Venda. Faz-se a média aritmética dos três P_m 's. $P_0(S_0 = 35, K = 50) = (0 + e^{-5r}(50 - 32) + e^{-4r}(50 - 20))/3$.

2.4 Testes Numéricos com o algoritmo de Grant, Vora e Weeks

Para testar a metodologia de Grant, Vora e Weeks [7] consideraram-se testes numéricos com os mesmos parâmetros dos realizados por Huang, Subrahmanyam e Yu [6], que utilizaram o método binomial, e cujos resultados são usados como *benchmark*.

Dado que esse algoritmo não especifica claramente o valor que os incrementos $-\Delta S$ devem ter, estipulou-se um valor muito pequeno igual a 0,01 (na seção 3 será visto este valor se fez igual ao erro ξ utilizado no passo 5 do algoritmo de Ibáñez e Zapatero [9], só para uniformizar o erro de convergência). Foram empregadas 100.000 simulações para o cálculo da curva de gatilho a cada tentativa de encontrar os preços críticos de exercício e outras 100.000 para o preço do ativo subjacente que permite calcular o preço da opção de

venda tendo já tracejada a curva de gatilho. Repete-se esta última etapa 50 vezes, com o fim de obter um preço médio da opção e um desvio padrão. Na Tabela 1 exibem-se os resultados.

A seguir são apresentadas as explicações referentes à Tabela 1:

- O número de datas de exercício refere-se à quantidade de intervalos discretos em que o espaço de tempo compreendido entre $t_0=0$ e $t_N = T$ foi subdividido. Foram consideradas 5 e 25 datas de exercício para efeitos de comparar com os resultados obtidos por Ibáñez e Zapatero [9], os quais fizeram os testes somente com essas datas.
- Para exemplificar, o primeiro teste numérico tem como parâmetros $K=35$, $\sigma=20\%$ ao ano, $T=1$ mês (0,0833 anos). O preço médio da opção de venda calculado pelo algoritmo de Grant, Vora e Weeks modificado (com 5 datas de exercício) resultou no mesmo valor do benchmark, com um desvio padrão de 0,0002.
- A coluna “% diferença” mede a porcentagem, em valor absoluto, em que o preço da opção de venda calculado por meio das simulações se distancia do valor verdadeiro. Sua fórmula é: $|\text{Preço Opção de Venda} - \text{Opção de Venda Verdadeira}| / (\text{Opção de Venda Verdadeira})$.

Tabela 1: Resultados do Preço da Opção de Venda aplicando o algoritmo Grant, Vora e Weeks ($S_0 = 40$; $r = 0,0488$; $q = 0$)

K	σ	T (anos)	Opção de Venda Verdadeira*	5 datas de exercício			25 datas de exercício		
				Preço Opção de Venda	Desvio Padrão	% diferença	Preço Opção de Venda	Desvio Padrão	% diferença
35	0,2	0,0833	0,0062	0,0062	0,0002	0,00%	0,0061	0,0002	1,61%
35	0,2	0,5833	0,4328	0,4270	0,0031	1,34%	0,4300	0,0032	0,65%
40	0,2	0,0833	0,8522	0,8485	0,0024	0,43%	0,8499	0,0022	0,27%
40	0,2	0,5833	1,9904	1,9643	0,0056	1,31%	1,9841	0,0040	0,32%
45	0,2	0,0833	5,0000	4,9654	0,0009	0,69%	4,9927	0,0001	0,15%
45	0,2	0,5833	5,2670	5,2077	0,0066	1,13%	5,2559	0,0063	0,21%
35	0,4	0,0833	0,2466	0,2463	0,0026	0,12%	0,2457	0,0026	0,36%
35	0,4	0,5833	2,1549	2,1382	0,0084	0,77%	2,1496	0,0091	0,25%
40	0,4	0,0833	1,7681	1,7645	0,0049	0,20%	1,7651	0,0038	0,17%
40	0,4	0,5833	4,3526	4,3240	0,0066	0,66%	4,3457	0,0086	0,16%
45	0,4	0,0833	5,2868	5,2781	0,0055	0,16%	5,2781	0,0055	0,16%
45	0,4	0,5833	7,3830	7,3415	0,0089	0,56%	7,3733	0,0079	0,13%
				MAPE		0,6156%	MAPE		0,3698%
				RMSE		2,6379%	RMSE		0,6355%

* Os preços verdadeiros da opção de venda americana em cada teste numérico (12 em total) são os obtidos por Huang, Subrahmanyam e Yu [8] usando um modelo binomial com 10.000 passos. Estes resultados servem como *benchmark* para o cálculo das medidas de erro: MAPE e RMSE.

- O **MAPE** é uma medida estatística do erro para um conjunto de testes. No total foram realizados 12 testes numéricos (um em cada linha da Tabela 1). Em cada teste a “% de diferença” varia. Portanto, uma maneira de consolidar uma medida de erro para um conjunto de testes realizados sob certos parâmetros comuns e um determinado número de datas de exercício, é por meio desta medida. Numericamente, o MAPE é a média aritmética da coluna “% diferença” (explicada no parágrafo anterior).
- O **RMSE** é outra medida estatística do erro de um conjunto de testes, que para o caso de 12 testes sua fórmula seria:
$$RMSE = \sqrt{\sum_{i=1}^{12} (\text{Preço Put}_i - \text{Put Verdadeira}_i)^2 / 12}$$
 Em outras palavras, é a raiz quadrada do erro médio quadrático.

- Chame-se de “experimento” ao conjunto de testes numéricos realizados sob uma certa quantidade de datas de exercício. Por exemplo, a Tabela 1 exhibe dois experimentos, os quais agrupam 12 diferentes testes numéricos. Em cada experimento o conjunto de testes numéricos são os mesmos, só varia o valor do parâmetro “número de datas de exercício”.
- Quanto menor for a percentagem nestas medidas de erro, melhor será a aproximação para os valores de referência ou *benchmarks*. Nos experimentos apresentados na Tabela 1 nota-se que um maior número de datas de exercício (esta é a variável que muda de um experimento para outro) faz que os valores do MAPE e do RMSE sejam reduzidos.

3 O algoritmo de Ibáñez e Zapatero

A seguir será descrita a metodologia de Ibáñez e Zapatero [9] para o cálculo do preço de opções americanas via Simulação de Monte Carlo, e que também utiliza o conceito de curva de gatilho.

3.1 Determinar a fronteira de exercício ótima um período antes da maturidade

Dado que o preço crítico de exercício no instante de tempo $t_N = T$ ($S_{t_N}^*$) é igual a K (como explicado na seção 2.2), o trabalho então, concentra-se em calcular recursivamente os outros pontos da curva de gatilho. Inicia-se então pelo preço crítico de exercício no período t_{N-1} .

Passo 1: Deseja-se achar o ponto $S_{t_{N-1}}^*$. Começa-se com um ponto inicial $S_{t_{N-1}}^{(1)}$ escolhido arbitrariamente. Normalmente toma-se este valor igual ao preço de exercício K .

Passo 2: Depois, calcula-se o preço da opção: $P_{t_{N-1}}(S_{t_{N-1}}^{(1)}, K)$ em t_{N-1} , aplicando para isso a simulação de Monte Carlo conforme a equação (3). Por outro lado, e só no instante t_{N-1} , seria também possível empregar a conhecida fórmula de Black e Scholes [2], visto que entre t_{N-1} e t_N existe um só período. Assim, tomando como valor inicial do ativo $S_{t_{N-1}}^{(1)}$ e o preço de exercício K em t_N , calcula-se o preço da opção de venda com a exatidão que fornece esta fórmula, sem ser necessário (neste particular caso) realizar as simulações.

Passo 3: Para encontrar um novo preço $S_{t_{N-1}}^{(2)}$ que se aproxime mais do $S_{t_{N-1}}^*$, é necessário encontrar uma regra eficiente para ir de um ponto de aproximação a outro. Uma forma bastante rápida para convergir ao preço crítico de exercício é utilizar o método de aproximações sucessivas de Newton. A convexidade da função preço da opção garante a convergência até o ponto fixado $S_{t_{N-1}}^*$. Assim:

Para $s = 1, 2, 3, \dots, *$, usa-se a aproximação:

$$P_{t_{N-1}}(S_{t_{N-1}}^{(s)}, K) + P_S(S_{t_{N-1}}^{(s)})(S_{t_{N-1}}^{(s+1)} - S_{t_{N-1}}^{(s)}) = I(S_{t_{N-1}}^{(s+1)}, K) \quad (9)$$

onde $P_S(S_{t_{N-1}}^{(s)}) = \frac{\partial P(S_{t_{N-1}}^{(s)}, K)}{\partial S}$. Reorganizando a equação (9) encontra-se o valor $S_{t_{N-1}}^{(s+1)}$:

$$S_{t_{N-1}}^{(s+1)} = \left(K - P_{t_{N-1}}(S_{t_{N-1}}^{(s)}, K) + P_S(S_{t_{N-1}}^{(s)})S_{t_{N-1}}^{(s)} \right) \div \left(1 + P_S(S_{t_{N-1}}^{(s)}) \right) \quad (10)$$

Para resolver a equação (10) deve-se calcular antes $P_S(S_{t_{N-1}}^{(s)})$, que é a principal dificuldade na aplicação do método de Newton.

O cálculo analítico de $P_S(S_{t_{N-1}}^{(s)})$ para uma opção de venda americana, a cada iteração 's', não é possível (só em opções europeias existe expressão analítica para $P_S(S_{t_{N-1}}^{(s)})$). É um fato amplamente conhecido na literatura de Finanças que à medida que S_{t_n} (ne $\{N-1, N-2, \dots, 1\}$) se aproxima mais do preço crítico de exercício $S_{t_n}^*$, a derivada se inclina mais tendendo a -1. Por esta razão, Ibáñez e Zapatero [9] sugerem inicializar as iterações com um valor $P_S(S_{t_{N-1}}^{(1)}) = -0.60$, e gradativamente realizar incrementos (negativos) a cada iteração até um máximo de $P_S(S_{t_{N-1}}^{(*)}) = -0.90$.

Passo 4: Achado $S_{t_{N-1}}^{(2)}$ calcula-se o preço da opção $P_{t_{N-1}}(S_{t_{N-1}}^{(2)}, K)$, aplicando a equação (3), ou a equação de Black e Scholes (lembrado que esta só serve no período t_{N-1}). Após, calcula-se uma nova aproximação $S_{t_{N-1}}^{(3)}$ usando a equação (10).

Passo 5: Repete-se o procedimento 's' vezes até convergir ao valor $S_{t_{N-1}}^{(*)} = S_{t_{N-1}}^{(s)}$, sendo que $|S_{t_{N-1}}^{(s)} - S_{t_{N-1}}^{(s-1)}| < \xi$, para algum número ξ muito pequeno.

A convergência para o ponto fixado $S_{t_{N-1}}^{(*)}$ se realiza de maneira monotônica, isto é, para uma opção de venda tem-se que: $S_{t_{N-1}}^{(1)} > S_{t_{N-1}}^{(2)} > \dots > S_{t_{N-1}}^{(*)}$. As iterações finalizam quando se encontra $S_{t_{N-1}}^{(s-1)} - S_{t_{N-1}}^{(s)} < \xi$ ou quando exista uma mudança no sinal da convergência: $S_{t_{N-1}}^{(s)} > S_{t_{N-1}}^{(s-1)}$. Em qualquer dos dois casos estima-se o ponto médio das duas últimas iterações: $\hat{S}_{t_{N-1}}^{(*)} = (S_{t_{N-1}}^{(s)} + S_{t_{N-1}}^{(s-1)})/2$, sendo esta a estimativa do valor de $S_{t_{N-1}}^{(*)}$.

3.2 Determinar a fronteira de exercício ótima um período antes da maturidade

Seguidamente, repete-se o mesmo procedimento para os pontos $t_{N-2}, t_{N-1}, \dots, t_0$ (de maneira recursiva). A cada ponto, sugere-se reiniciar o algoritmo tomando como preço inicial do ativo $S_{t_n}^{(1)} = S_{t_{n+1}}^{(*)}$ (o ponto fixado do período à frente, calculado previamente).

Assim, no final encontra-se um conjunto de pontos $S^{(*)} = \{S_{t_0}^{(*)}, S_{t_1}^{(*)}, \dots, S_{t_{N-1}}^{(*)}, S_{t_N=T}^{(*)}\}$, que formam a fronteira de exercício ótima ou curva de gatilho: $S_{t_n}^{(*)} = \hat{F}_{t_n}(t_n)$. Uma maneira eficiente de achar a função $\hat{F}_{t_n}$ é fazendo uma regressão quadrática ou cúbica do conjunto $S^{(*)}$ com os tempos $\{t_0, t_1, \dots, T\}$, embora resulte também aceitável fazer uma simples interpolação entre dois períodos discretos.

A utilidade da curva de gatilho está no fato que, no primeiro instante de tempo em que o valor do ativo subjacente "S" fique abaixo desta curva, dever-se-á optar pelo exercício, sendo o preço da opção de venda em algum instante t_n o valor intrínseco: $K - S_{t_n}$.

3.3 Determinar a fronteira de exercício ótima um período antes da maturidade

Uma vez traçada a curva de gatilho, para calcular o preço da opção de venda, ao igual que em Grant, Vora e Weeks [7] (seção 2.3), aplica-se a equação (3) a partir de simulações do ativo subjacente.

4 O algoritmo de Ibáñez e Zapatero modificado

Nesta seção apresentamos uma proposta de aprimoramento do algoritmo de Ibáñez e Zapatero [9]. Posteriormente realizamos os experimentos numéricos.

4.1 Aprimoramento no cálculo da curva de gatilho

A primeira melhoria que se pode efetuar no cálculo do ponto $S_{t_{N-1}}^{(*)}$ é empregar a fórmula de Black e Scholes [2] para calcular o preço exato da opção de venda nos passos 2 e 4 do algoritmo, dado que existe um único período até a maturidade da opção. Ibáñez e Zapatero [9] também fazem esta sugestão. A novidade estaria no passo 3, onde deve ser estimado $P_S(S_{t_{N-1}}^{(s)})$. Quando existe só um período até a maturidade da opção, o valor da derivada do preço da opção com relação ao preço do ativo subjacente é calculado por uma expressão analítica fechada proveniente da fórmula de Black e Scholes [2] dada por:

$$P_S(S_{t_{N-1}}^{(s)}) = e^{-q(T-t_{N-1})} [N(d_1) - 1] \quad (11)$$

onde:

$$d_1 = \left[\ln \left(\frac{S_{t_{N-1}}^{(s)}}{K} \right) + \left(r - q + \frac{1}{2} \sigma^2 \right) (T - t_{N-1}) \right] \frac{1}{\sigma \sqrt{T - t_{N-1}}}$$

$N(.)$ = função distribuição normal padrão acumulada.

q = taxa de dividendos.

r = taxa de desconto livre de risco.

K = preço de exercício.

σ = volatilidade do preço do ativo subjacente.

Portanto, no passo 3 do algoritmo, referente ao período t_{N-1} é conveniente usar a expressão analítica exata de $P_S(S_{t_{N-1}}^{(s)})$. Isto permitirá uma melhor convergência para o ponto $S_{t_{N-1}}^{(*)}$.

Ibáñez e Zapatero [9] sugerem aplicar o algoritmo em duas etapas para obter um melhor cálculo da curva de gatilho nos períodos $t_{N-1}, t_{N-2}, \dots, t_0$. Seguindo esta sugestão, no algoritmo de Ibáñez e Zapatero modificado também se aplicam as duas etapas nos mesmos períodos com exceção do ponto fixado no período t_{N-1} , que já foi calculado empregando valores analíticos exatos do preço da opção de venda e da sua derivada. A seguir se explica cada etapa.

(i) Primeira etapa

Determinar uma curva de gatilho simulando uma quantidade não muito grande de caminhos Brownianos aleatórios, por exemplo, 5.000, a cada vez que se utilize a equação (3) nos passos 2 e 4. Para efeitos de redução de variância, as 5.000 simulações devem estar constituídas por 2.500 simulações mais seus respectivos valores antitéticos.

Variável antitética é uma das principais e mais simples técnicas para reduzir variância. Consiste em gerar uma variável estocástica negativamente correlacionada à variável de estado do ativo objeto. Assim, cada trajetória deve ser associada a um par de seqüências, isto é, duas trajetórias negativamente correlacionadas. Para maior informação sugere-se consultar Frota [6].

Melhoria proposta: No passo 3 do algoritmo, Ibáñez e Zapatero [9] não fornecem maiores detalhes de como fazer os acréscimos de $P_S(S_{t_n}^{(s)})$ a cada iteração. Eles sugerem começar com um valor $P_S(S_{t_{N-1}}^{(1)}) = -0.60$, e depois fazer incrementos (negativos) a cada iteração até um valor máximo de $P_S(S_{t_{N-1}}^{(*)}) = -0.90$. Mas, que valores devem ter tais incrementos?

A curva de gatilho a ser calculada nesta etapa é uma primeira tentativa de aproximar os preços críticos $\{S_{t_0}^{(*)}, S_{t_1}^{(*)}, \dots, S_{t_{N-3}}^{(*)}, S_{t_{N-2}}^{(*)}\}$, assim, observou-se através de diversos experimentos numéricos, que uma maneira rápida de achar a convergência é realizando acréscimos de -0.05 ($-\Delta = -0.05$) começando com um valor de $P_S(S_{t_n}^{(1)}) = -0.60$. De acordo com as experimentações realizadas, na terceira ou quarta iteração já encontra-se o preço crítico de exercício, com um erro $\xi = 0.01$ (erro de convergência descrito no passo 5 do algoritmo, que é do mesmo valor considerado por Ibáñez e Zapatero [9] nos testes que realizaram).

(ii) Segunda etapa

Os preços críticos de exercício achados na primeira etapa, servirão de ponto de partida $S_{t_n}^{(1)}$ a cada t_n ($n = \{0, 1, \dots, N-2\}$) em um novo cálculo da curva de gatilho. Assim, $S_{t_n}^{(1)}$ da etapa 2 é igual a $S_{t_n}^{(*)}$ da etapa 1. Estes pontos iniciais estão muito mais próximos do verdadeiro valor $S_{t_n}^{(*)}$, e conseqüentemente $P_S(S_{t_n}^{(s)})$ se aproxima mais de -1 . Portanto, ao aplicar novamente o algoritmo espera-se obter uma melhor aproximação da curva de gatilho.

Para melhorar ainda mais a precisão, foram simulados 100.000 caminhos Brownianos aleatórios (50.000 mais seus respectivos valores antitéticos) a cada vez que se utilize a equação (3) nos passos 2 e 4.

Melhoria proposta: Nesta etapa, Ibáñez e Zapatero [9] não falam como fazer os acréscimos em $P_S(S_{t_n}^{(s)})$. Então, visto que os pontos iniciais encontram-se muito próximos dos preços críticos de exercício, diversos testes práticos efetuados indicam que, começando as iterações com um $P_S(S_{t_n}^{(1)}) = -0.85$ e fazendo posteriormente incrementos ($-\Delta$) pequenos de -0.01 , consegue-se uma boa aproximação. No máximo em 10 iterações encontrar-se-á o novo preço crítico de exercício, considerando um erro $\xi = 0.01$.

4.2 Experimentos utilizando o algoritmo de Ibáñez e Zapatero modificado

Ibáñez e Zapatero [9] fizeram testes numéricos com os mesmos parâmetros dos realizados por Huang, Subrahmanyam e Yu [6], que utilizaram o método binomial, e cujos resultados são usados como *benchmark*. O algoritmo de Ibáñez e Zapatero modificado também replica esses testes e logo ambas as metodologias são comparadas.

Na Tabela 2 são exibidos doze testes numéricos nos quais é calculado o preço da opção de venda americana. Os parâmetros comuns são os seguintes: $r = 0,0488$ anual; $q=0$. Aplicam-se as duas etapas (como foram descritas) na determinação da curva de gatilho. Uma vez obtida a curva, para calcular o preço da opção de venda simulam-se 100.000 caminhos Brownianos (50.000 com seus valores antitéticos) a partir do preço inicial do ativo $S_0 = 40$. Simulam-se estes caminhos 50 vezes (sob a mesma curva de gatilho calculada previamente), e o preço da opção de venda provém da média aritmética do preço da opção obtido a cada vez em que se realizaram as 100.000 simulações de caminhos.

Tabela 2: Resultados do algoritmo de Ibáñez e Zapatero modificado
($S_0 = 40$; $r = 0,0488$; $q = 0$)

K	σ	T (anos)	Opção de Venda Verdadeira*	5 datas de exercício			25 datas de exercício		
				Preço Opção de Venda	Desvio Padrão	% diferença	Preço Opção de Venda	Desvio Padrão	% diferença
35	0,2	0,0833	0,0062	0,0062	0,0003	0,00%	0,0062	0,0002	0,00%
35	0,2	0,5833	0,4328	0,4263	0,0035	1,50%	0,4315	0,0038	0,30%
40	0,2	0,0833	0,8522	0,8487	0,0023	0,41%	0,8512	0,0019	0,12%
40	0,2	0,5833	1,9904	1,9654	0,0046	1,26%	1,9851	0,0050	0,27%
45	0,2	0,0833	5,0000	4,9659	0,0011	0,68%	4,9927	0,0002	0,15%
45	0,2	0,5833	5,2670	5,2076	0,0053	1,13%	5,2562	0,0058	0,21%
35	0,4	0,0833	0,2466	0,2465	0,0027	0,04%	0,2467	0,0026	0,04%
35	0,4	0,5833	2,1549	2,1408	0,0075	0,65%	2,1528	0,0089	0,10%
40	0,4	0,0833	1,7681	1,7654	0,0040	0,15%	1,7676	0,0046	0,03%
40	0,4	0,5833	4,3526	4,3243	0,0093	0,65%	4,3477	0,0084	0,11%
45	0,4	0,0833	5,2868	5,2791	0,0054	0,15%	5,2853	0,0045	0,03%
45	0,4	0,5833	7,3830	7,3406	0,0090	0,57%	7,3748	0,0094	0,11%
				MAPE	0,5997%			MAPE	0,1211%
				RMSE	2,6196%			RMSE	0,4991%

* Os preços verdadeiros da opção de venda americana em cada teste numérico (12 em total) são os obtidos por Huang, Subrahmanyam e Yu [8] usando um modelo binomial com 10.000 passos. Estes resultados servem como *benchmark* para o cálculo das medidas de erro: MAPE e RMSE.

4.3 Comparação dos resultados em Ibáñez e Zapatero (I&Z) e I&Z modificado

Nesta seção se realiza uma comparação baseada nos resultados das medidas de erro entre os algoritmos de Ibáñez e Zapatero [9] e Ibáñez e Zapatero modificado, sendo que em ambos os casos são considerados os testes numéricos da Tabela 2. Na Tabela 3, são exibidos tais resultados.

Tabela 3: Medidas de erro: Ibáñez e Zapatero (I&Z) versus I & Z Modificado

Algoritmo utilizado	5 datas de exercício		25 datas de exercício	
	MAPE	RMSE	MAPE	RMSE
Ibáñez e Zapatero (I & Z)*	1,0818%	2,7295%	0,6151%	0,9567%
I & Z Modificado**	0,5997%	2,6196%	0,1211%	0,4991%

Fonte: * Elaboração própria a partir dos resultados obtidos por I&Z.

** Resultados provenientes da Tabela 2.

Observa-se na Tabela 3, que o algoritmo I&Z modificado gera menores erros em ambas as datas de exercício. Desta maneira, verifica-se que as modificações realizadas no algoritmo de Ibáñez e Zapatero [9] foram apropriadas. Esta melhora na precisão dos resultados é devido ao aprimoramento realizado na determinação da curva de gatilho. Na estimação do ponto crítico de exercício do penúltimo período, por exemplo, ao invés de simular caminhos entre uma iteração e outra até atingir o preço crítico de exercício, emprega-se a fórmula de Black e Scholes [2], e, além disso, ao utilizar as aproximações sucessivas de Newton usou-se a fórmula analítica da derivada da opção. Por outro lado, na estimação dos outros pontos críticos de exercício, o fato de que em cada uma das duas etapas de aplicação do algoritmo ter sido utilizado diferentes valores iniciais da derivada e dos acréscimos (passo 3 do algoritmo), contribuiu também para obter uma melhor aproximação.

4.4 Análise de sensibilidade do algoritmo de I&Z modificado

Nesta seção são realizadas análises de sensibilidade de certos parâmetros, como o número de datas de exercício, o número de simulações para obter o preço da opção e o cálculo de uma curva de gatilho média. A seguir detalham-se estas experimentações.

4.4.1. Número de Datas de Exercício

Na Tabela 3, observa-se que o número de datas de exercício melhora a precisão dos resultados, obtendo menores valores para os erros.

Foram realizados experimentos adicionais alterando-se o valor desta variável. Os testes numéricos são sempre os exibidos na Tabela 2, mantendo os mesmos parâmetros para o cálculo da curva de gatilho e para o cálculo do preço da opção de venda. Os resultados dessas experimentações são exibidos a seguir na Figura 2.

De acordo com a Figura 2, nota-se que existe uma significativa redução das medidas de erro à medida que o número de datas de exercício vai crescendo até um valor aproximado de 25. A partir daí a redução das medidas de erro continua, mas já não é tão significativo o ganho obtido pelo acréscimo de maior quantidade de datas.

O tempo de processamento computacional cresce à medida que se incrementa o número de datas de exercício. A Figura 3 exibe os tempos médios despendidos no cálculo do preço da opção de venda americana num teste numérico, em função do número de datas de exercício.

Em relação ao tempo de processamento computacional nota-se uma significativa elevação à medida que o número de datas de exercício aumenta. Observa-se que, o tempo médio que se leva em computar um teste numérico com 25 datas de exercício é de 65,16 segundos (1 minuto aprox.), mas ao se dobrar o número de datas de exercício (50 datas) o tempo de processamento sobe para 185,40 segundos (3 minutos aproximadamente). O ganho na melhora do erro de passar de 25 para 50 datas talvez não justifique o custo computacional.

O programa que computou os diferentes testes numéricos foi desenvolvido em MatLab, rodado em um computador Pentium IV de 2,8 GHz e 480 MB de RAM.

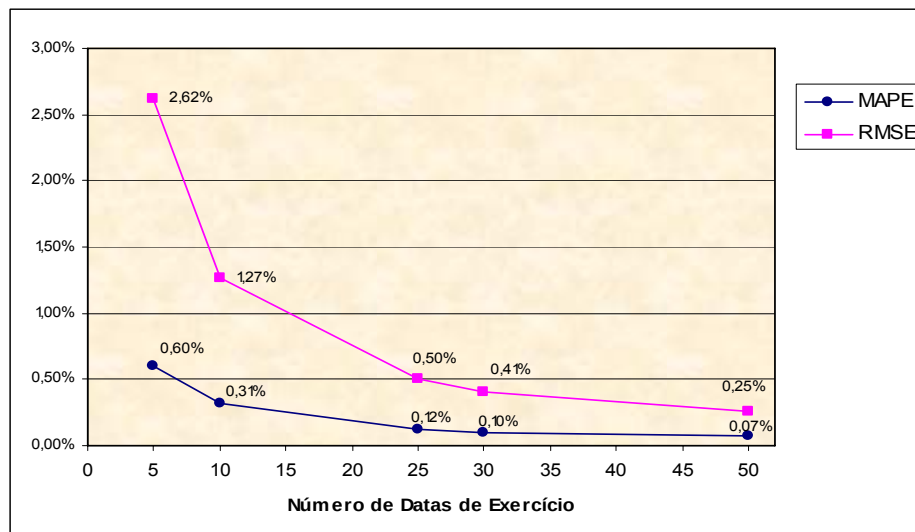


Figura 2: Valor das medidas de erro em função do número de datas de exercício

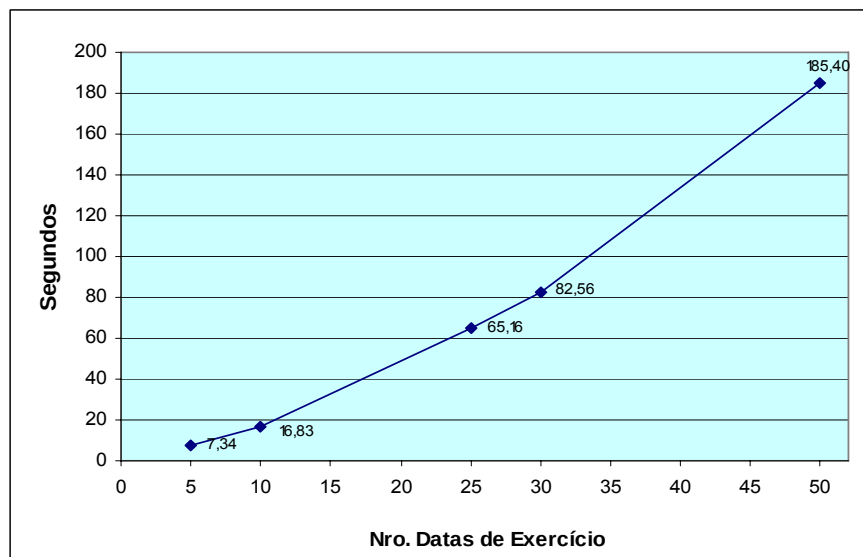


Figura 3: Tempo computacional em função do número de datas de exercício

4.4.2. Outras Análises de Sensibilidade

Foram realizados experimentos numéricos para analisar o comportamento de algumas variáveis, por exemplo, o número de simulações de caminhos Brownianos do preço do ativo subjacente, que permite calcular o preço da opção de venda americana uma vez que se tenha determinado a curva de gatilho. Dividindo o tempo até a maturidade em 25 datas de exercício, constatou-se que a partir de 10.000 simulações consegue-se uma estabilidade nas medidas de erro MAPE e RMSE, e, em relação ao tempo de processamento computacional, existe uma diferença de poucos segundos entre o que se consome com 10.000 e 100.000 simulações. No entanto, optou-se por fazer os testes com

100.000 simulações, pois Ibáñez e Zapatero [9] utilizaram essa quantidade de simulações. Desta forma, e para efeitos de comparação, manteve-se o mesmo valor do parâmetro.

Adicionalmente realizou-se um experimento para analisar se a média de várias curvas de gatilho melhoraria a precisão dos resultados. Os resultados mostraram que o fato de calcular várias curvas de gatilho e com estas se obter uma curva média (tirando a média dos valores dos preços críticos de exercício em cada instante de tempo), não implicou em uma melhora nas medidas de erro. Conclui-se então que basta calcular uma única curva aplicando sempre as duas etapas descritas. A média de várias curvas de gatilho resultou ser desnecessária aumentando exponencialmente o tempo de processamento computacional.

5 Análise dos resultados e comparações das metodologias testadas

Apresenta-se na Figura 4 um resumo das medidas de erro para os três algoritmos testados, sendo que o tempo até a maturidade foi dividido em 25 datas de exercício, o que mostrou ser uma quantidade razoável para conseguir uma boa precisão nos resultados. Nota-se que o algoritmo I&Z modificado ofereceu as menores medidas de erro.

No que diz respeito ao tempo de processamento computacional (ver Figura 5) os algoritmos de Ibáñez e Zapatero [9] e Ibáñez e Zapatero modificado empregam tempos bastante similares, mas o algoritmo de Grant, Vora e Weeks [7] demanda um tempo muito maior, isto porque não existe uma regra específica que acelere a convergência para os pontos críticos de exercício. Sempre utiliza-se o mesmo valor $-\Delta S$ e conseqüentemente em determinados períodos são requeridas muitas iterações até aproximar o ponto crítico.

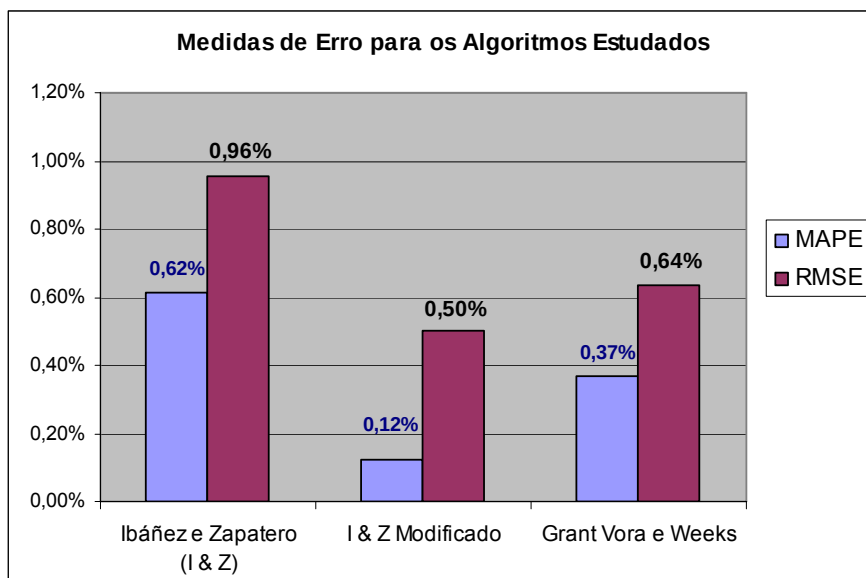


Figura 4: Medidas de Erro nos três algoritmos abordados

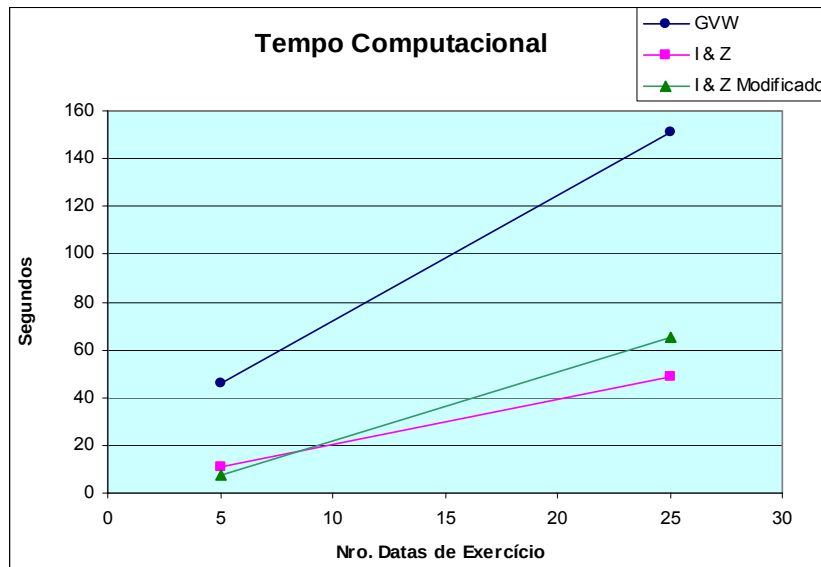


Figura 5: Tempo computacional em função do número de datas de exercício

6 Conclusões e Considerações Finais

O grande apelo na utilização do algoritmo de Grant, Vora e Weeks [7], deve-se ao fato de que esta metodologia permite o cálculo prévio de uma curva de gatilho, a qual é muito relevante, especialmente quando são avaliadas opções reais. A curva de gatilho permite identificar o momento ótimo de exercer a opção, tanto em opções financeiras bem como em opções de projetos de investimento, por exemplo, investir ou abandonar o projeto.

No presente trabalho foram avaliados os algoritmos de Grant, Vora e Weeks [7] e Ibáñez e Zapatero [9]. Este último também determina uma curva de gatilho, e portanto, ambos os algoritmos são possíveis de serem comparados em quanto à eficiência na determinação da curva de gatilho e na exatidão dos resultados. Adicionalmente foram propostas modificações no algoritmo de Ibáñez e Zapatero [9], de forma a obter uma melhor precisão no cálculo dos preços das opções de venda americanas. O resultado final dos testes é exibido na Figura 4.

Por outro lado, observou-se que o número de datas de exercício é um parâmetro que influencia de maneira relevante na precisão dos resultados (ver Figura 2). É lógico que ao aumentar as datas de exercício os resultados fiquem mais próximos dos verdadeiros, já que representa melhor o comportamento real de uma opção americana, na qual o exercício dá-se em tempo contínuo e não por intervalos de tempo discretos (opção bermuda).

A proposta feita para melhorar a convergência para os preços críticos de exercício, estabelecendo valores iniciais das derivadas do preço da opção e dos incrementos a cada iteração (passo 3 do algoritmo Ibáñez e Zapatero [9]), mostrou ser bastante eficiente, permitindo melhorar os resultados obtidos por Ibáñez e Zapatero [9]. Da mesma forma, o algoritmo de Ibáñez e Zapatero modificado obteve erros bem menores do que Grant, Vora e Weeks [7]. Todos estes algoritmos têm em comum a necessidade de determinar previamente a curva de gatilho. O que varia é como as aproximações para os preços críticos de exercício são realizadas.

Com relação ao tempo computacional, o algoritmo de Ibáñez e Zapatero modificado consome um tempo muito menor do que a metodologia de Grant, Vora e Weeks [7], sendo esta última não prática em problemas complexos.

7 Referências

- [1] Barone-Adesi, G. & Whaley, R.E. "Efficient Analytic Approximation of American Option Value". *Journal of Finance*, v.42, 1987, pp.301-320.
- [2] Black, F. & Scholes, M. "The Pricing of Options and Corporate Liabilities". *Journal of Political Economy*, v.81, 1973, pp.637-659.
- [3] Boyle, P. & Broadie, M. & Glasserman, P. "Monte Carlo Methods for Security Pricing". *Journal of Economic Dynamics and Control*, v.21, 1997, pp.1267-1321.
- [4] Broadie, M. & Glasserman, P. "Pricing American-Style Securities Using Simulation". *Journal of Economic Dynamics and Control*, v.21, 1997, pp.1323-1352.
- [5] Cox, J.C. & Ross, S.A. & Rubinstein, M. "Option Pricing: A Simplified Approach". *Journal of Financial Economics*, n.7, 1979, pp.229-263.
- [6] Frota, A.E.F. "Avaliação de Opções Americanas Tradicionais e Complexas". *Dissertação de Mestrado*, Departamento de Engenharia Industrial, PUC-Rio, Rio de Janeiro, 2003.
- [7] Grant, D. & Vora, G. & Weeks, D.E. "Path-Dependent Options: Extending the Monte Carlo Simulation Approach". *Management Science*, v.43, 1997, pp.1589-1602.
- [8] Huang, J. & Subrahmanyam, M.G. & Yu, G.G. "Pricing and Hedging American Options; A Recursive Integration Method". *Review of Financial Studies*, v.9, 1996, pp.277-300.
- [9] Ibáñez, A. & Zapatero, F. "Monte Carlo Valuation of American Options Through Computation of the Optimal Exercise Frontier". *Journal of Financial and Quantitative Analysis*, v.39, 2004, pp.253-275.
- [10] Nascimento, A. F. "Avaliação de Investimentos em Tecnologia da Informação: uma Perspectiva de Opções Reais". *Dissertação de Mestrado*, Departamento de Engenharia Industrial, PUC-Rio, Rio de Janeiro, 2005.

Minimização da Energia Potencial de Sistemas Estruturais Não-Lineares: uma comparação de algoritmos numéricos de programação não-linear

Saete Souza de Oliveira Buffoni [†]

Aline Michelly Silva Moreira [‡]

Milena de Andrade Sacramento [‡]

[†] Universidade Federal Fluminense – UFF
Brasil

saete@vm.uff.br

<http://www.professores.uff.br/saete>

milenasacramento@vm.uff.br

aline_michelly@vm.uff.br

Abstract

The main aim of the present work is to develop a formulation and some strategies for implementation and comparison of methods in nonlinear programming without restriction applied to the calculation of displacements in structural elements, presenting as example the analysis and calculation of the displacements in a truss element through the formulation of the minimum potential energy. The displacements had been calculated through the following methods: Univariate, Gradient Descent Method, Powell Conjugated Directions Method, Fletcher-Reeves Method, Standard Newton Raphson Method, Modified Newton Raphson Method, Rank One Correction Method, Rank Two Correction Method and BFGS Method. The algorithms had been implemented in Fortran language. After the solution of the problem it carries through comparisons between the methods.

Resumo

O presente trabalho trata do estudo e implementação de métodos em programação não-linear sem restrição utilizando os métodos Univariante, Máximo Declive, Powell, Fletcher-Reeves, Newton Raphson Padrão, Newton Raphson Modificado, Rank One Correction, Rank Two Correction e BFGS. O algoritmo é implementado em linguagem fortran. Em seguida a implementação dos métodos executa-se um exemplo para o cálculo dos deslocamentos em uma treliça através da formulação da energia potencial mínima, no qual a função-objetivo é uma função não-linear das variáveis de decisão. Em seguida a solução do problema realiza-se uma comparação entre os métodos.

Keywords: Mathematical programming, optimization without restrictions, Truss Elements, Algorithms.

Title: Potential Energy Minimization of Nonlinear Structural Systems: a comparison of numerical algorithms for nonlinear programming.

1 Introdução

Otimizar é melhorar o que já existe, projetar o novo com mais eficiência e menor custo. A otimização visa determinar a melhor configuração de projeto sem ter que testar todas as possibilidades envolvidas. Problemas de otimização são caracterizados por situações em que se deseja maximizar ou minimizar uma função numérica de várias variáveis, num contexto em que podem existir restrições. Tanto as funções acima mencionadas como as restrições dependem dos valores assumidos pelas variáveis de projeto ao longo do procedimento de otimização.

Pode-se aplicar otimização em várias áreas, como por exemplo, no projeto de sistemas ou componentes, planejamento e análise de operações, problemas de otimização de estruturas, otimização de forma, controle de sistemas dinâmicos. A otimização tem como vantagens diminuir o tempo dedicado ao projeto, possibilitar o tratamento simultâneo de uma grande quantidade de variáveis e restrições de difícil visualização gráfica e/ou tabular, possibilitar a obtenção de algo melhor, obtenção de soluções não tradicionais, menor custo. Como limitações se tem o aumento do tempo computacional quando aumenta-se o número de variáveis de projeto, pode surgir funções descontínuas que apresentam lenta convergência, funções com presença de muitos mínimos locais onde o mínimo global raramente é obtido.

De acordo com vários autores dentre os quais se citam Gill, P. E, Murray, W. & Wright M. (1981), Filho (2004) e Friedlander (1994) uma das subáreas da Pesquisa Operacional é a Programação Matemática, que é dividida em Programação Linear e Programação Não-Linear com ou sem restrições. Quando se otimiza um problema, trabalha-se com um modelo que é uma representação simplificada do real, o qual abrange apenas as variáveis mais relevantes e que exercem maior impacto sobre a solução, porém, a maioria dos problemas possui não-linearidades que é uma melhor representação da realidade.

O presente trabalho apresenta o cálculo dos deslocamentos em uma treliça através da formulação da energia potencial mínima, nesse problema a função-objetivo é uma função não-linear das variáveis de decisão.

2 Métodos de Programação Não-Linear sem Restrições

Os métodos de programação não-linear sem restrição se dividem em três métodos. Os métodos de ordem zero que são utilizados quando o valor da função é obtido com precisão pobre. Pode ser uma função não diferenciável ou altamente não-linear e, portanto, os valores das derivadas (ou gradientes) não são confiáveis e não devem ser utilizados, pois é difícil obter as derivadas de forma precisa. Dessa forma, somente o valor da função-objetivo é utilizado. Entre os mais importantes métodos de ordem zero, há o método univariante e o método das direções conjugadas de Powell. Os métodos de primeira ordem utilizam os valores da função-objetivo e de suas derivadas (gradientes) em relação às variáveis de projeto. Exemplos clássicos destes métodos são o “Steepest Descent”, ou máximo declive, método dos gradientes conjugados e método de Fletcher e Reeves e por fim os métodos de segunda ordem que utilizam os valores da função-objetivo, de suas derivadas e também da matriz hessiana. Os exemplos mais importantes são os métodos de Newton Raphson Padrão e Quase-Newton.

Os métodos existentes para resolver problemas de Programação Não-Linear sem restrições procuram, inicialmente, encontrar uma direção d a seguir que reduza a função-objetivo, às vezes chamada de direção de busca. Uma vez obtida essa direção, decidem o quanto “andar” nessa direção, t . Através desse procedimento, a cada passo, um problema de encontrar n variáveis x é reduzido a um problema de encontrar uma variável t , como está descrito na equação a seguir:

$$x = x_o + t_o d_o \Rightarrow f(x) = f(x_o + t_o d_o) = f(t) \quad (1)$$

onde x_0 é o ponto inicial. O problema de encontrar t pode ser resolvido fazendo uso de técnicas de minimização de uma função de uma variável que são de fácil implementação. Essa etapa é denominada busca unidimensional, e, quando a direção d coincide com a direção de um dos eixos coordenados é denominada busca univariada. O algoritmo seguido pelos métodos de otimização sem restrições é apresentado a seguir:

- Encontrar x_0 e do que reduzam a função-objetivo;
- Encontrar t na direção d_0 que minimize f (busca unidimensional) e obter $x = x_0 + t_0 d_0$;
- Verificar a convergência, e, se satisfeita, pare ($x = x^*$);
- Caso contrário, $x_{i+1} = x$, $i = i+1$ (controle do número de iterações) e voltar à etapa inicial.

Detalhes sobre este assunto podem ser encontrados em Friedlander (1994)

3 Métodos Implementados

3.1 Método Univariante

Por ser um método para resolução de problemas de otimização sem restrições, deve-se trabalhar com uma direção de busca d , que minimize a função-objetivo f , um incremento t (o quanto andar) e um ponto inicial x_0 .

- Algoritmo:

- 1- Determinar se t é positivo ou negativo, sendo que seu valor deve ser bem pequeno;
- 2- Partindo de um ponto inicial x_0 escolhido e conhecendo-se t , encontrar o valor de x_{k+1} para uma dada direção e_k na qual se caminha. Verificar se $f(x_{k+1})$ decresce no sentido positivo ou negativo para este e_k , calculando-se $f(x_{k+1}) = f(x_k \pm t e_k)$;
- 3- Se $f(x_{k+1}) < f(x_k)$, continuar nesta direção, “avançando” mais um incremento, ou seja, dobrar t e testar novamente, voltando ao passo 2. Enquanto f estiver sendo minimizada, a direção e_k é mantida. Caso contrário, parar no último valor x_i bem sucedido na direção e_k e escolher uma nova direção e_{k+1} para continuar a “caminhada”, voltando portanto, ao passo 1, com t original;
- 4- Parar quando o valor da função f ficar constante, ou seja, quando nenhuma direção e_k melhorar f .

3.2 Método do Máximo Declive

Apesar de ser um método semelhante ao método univariante, sua particularidade é o fato de que a “caminhada” em busca do ponto ótimo ocorre sempre na direção do gradiente da função-objetivo, direção de busca d_k . Além disso, o tamanho do passo com o qual se caminha, t_k , é determinado a cada iteração através de uma busca linear na direção d_k .

- Algoritmo:

- 1- Determinar x_0 , ponto a partir do qual se inicia a busca pelo ponto ótimo;
- 2- Sendo $\nabla f(x_k)$, o gradiente da função-objetivo $f(x)$ no k -ésimo ponto x_k , a direção de busca na qual se caminha partindo de x_k é o versor de um vetor de sentido oposto ao gradiente de $f(x)$, sendo esta dada por:

$$d_k = - \frac{\nabla f(x_k)}{|\nabla f(x_k)|} \quad (2)$$

- 3- Em cada iteração obtém-se um ponto seguinte x_{k+1} dado por:

$$x_{k+1} = x_k + t_k d_k \quad (3)$$

Sendo t_k determinado por busca linear na direção d_k e $t_k > 0$;

- 4- Dessa forma há a geração de sucessivos pontos nos quais o valor da função-objetivo $f(x)$ se torna cada vez menor:

$$f(\mathbf{x}_{k+1}) = f(\mathbf{x}_k + t_k \mathbf{d}_k) \quad (4)$$

5- O algoritmo termina quando os valores assumidos pela função-objetivo nos sucessivos pontos gerados forem muito próximos, devido à proximidade destes pontos, $\mathbf{x}_k$ e $\mathbf{x}_{k+1}$. Ou seja:

$$\mathbf{x}_{k+1} \approx \mathbf{x}_k \quad (5)$$

$$f(\mathbf{x}_{k+1}) \approx f(\mathbf{x}_k) \quad (6)$$

Assim $|\nabla f(\mathbf{x}_k)| \approx 0$

3.3 Método das Direções Conjugadas de Powell

Uma maneira de acelerar o método Univariante é utilizar o chamado “movimento padrão” (*pattern move*) e esta foi a metodologia implementada por Powell.

Algoritmo:

1- Minimizar a função $f(\mathbf{x})$ com busca linear ao longo das direções $\mathbf{g}_k$,

$$\mathbf{g}_k = -\frac{\nabla f(\mathbf{x}_k)}{|\nabla f(\mathbf{x}_k)|} \quad (7)$$

partindo do ponto inicial $\mathbf{x}_{k0}$ de forma a gerar os pontos $\mathbf{x}_{k1}, \dots, \mathbf{x}_{kn}$;

2- Após encerrar a busca linear, encontrar o índice m , normalmente $m=n$, correspondente à direção em que a função-objetivo apresenta maior decréscimo, indo de $\mathbf{x}_{km-1}$ para $\mathbf{x}_{km}$;

3- Calcular a direção “padrão” $\mathbf{d}_k = \mathbf{x}_k - \mathbf{x}_{k-m}$ e determinar o tamanho do passo t_k que minimize $f(\mathbf{x})$, tal que

$$\mathbf{x} = \mathbf{x}_{k0} + t_k \mathbf{d}_k \quad (8)$$

Minimizar $f(\mathbf{x})$ com busca linear nesta direção “padrão” $\mathbf{d}_k$ e com o passo t_k ;

4- Descartar uma das direções, substituí-la por $\mathbf{d}_k$ na próxima etapa de minimizações e repetir as etapas (2) e (3).

3.4 Método de Fletcher-Reeves

O método de Fletcher-Reeves é um método do gradiente conjugado para funções não-quadráticas, que não necessita da avaliação da matriz hessiana $H(\mathbf{x})$, sendo esta substituída por uma busca linear (“*Linear Search*”).

- Algoritmo:

1. Determinar o ponto inicial $\mathbf{x}_0$ e calcular a direção de busca $\mathbf{d}_k$, dada por:

$$\mathbf{d}_k = \mathbf{d}_0 = -\mathbf{g}_0 \quad (9)$$

Onde,

$$\mathbf{g}_0 = \nabla f(\mathbf{x}_0) \quad (10)$$

2. Para $k = 0, 1, \dots, n-1$ faz-se:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + t_k \mathbf{d}_k \quad (11)$$

Onde t_k é o tamanho do passo, sendo este determinado por busca linear;

3. Ao obter-se o ponto $\mathbf{x}_{k+1}$, calcula-se o gradiente da função-objetivo $\mathbf{f}$ no ponto $\mathbf{x}_{k+1}$, dado por:

$$\mathbf{g}_{k+1} = \nabla f(\mathbf{x}_{k+1}) \quad (12)$$

4. Para determinar a nova direção de busca, $\mathbf{d}_{k+1}$, faz-se:

$$\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_k \mathbf{d}_k \quad (13)$$

Onde,

$$\beta_k = \frac{\mathbf{g}_{k+1}^t \mathbf{g}_{k+1}}{\mathbf{g}_k^t \mathbf{g}_k} \quad (14)$$

Enquanto $k \neq n-1$;

5. Ao obter-se $\mathbf{x}_n$, substituir $\mathbf{x}_0$ por $\mathbf{x}_n$ e retornar ao passo 1.

3.5 Método de Newton Raphson Padrão

O princípio deste método é minimizar uma função f através de uma aproximação local por uma função quadrática. As aproximações quadráticas ganham importância à medida que se aproximam do ponto ótimo do problema, sendo melhores do que as lineares. Esse método trabalha com $\lambda_k = 1$, sem “linear search”(busca linear).

Para minimizar quadrática $q(x_{k+1})$ devemos ter $\frac{\partial q}{\partial x_{k+1}} = 0$, logo :

$$\nabla_k + H_k (x_{k+1} - x_k) = 0 \quad (15)$$

ou

$$x_{k+1} = x_k - H_k^{-1} \nabla_k \quad (16)$$

Esta equação é a forma pura do método de Newton. O valor de H_k^{-1} é interpretado como uma correção na direção oposta ao gradiente da função, de forma a acelerar o processo iterativo. Concluindo que a direção de busca será:

$$d_k = -H_k^{-1} \nabla_k \quad (17)$$

3.6 Método de Newton Raphson Modificado

Esse método consiste na mesma formulação que o Newton Raphson Padrão, porém evita a atualização da matriz hessiana H_k em cada passo:

$$d_k = -H_0^{-1} \nabla_k \quad (18)$$

Uma modificação do método de Newton-Raphson é que ele pode ser implementado com “line search” ao invés de $t_k = 1$

3.7 Método Correção de Posto Um (Rank One Correction)

Primeiramente, examina-se a fórmula de recorrência.

$$S_{k+1} = S_k + a_k Z_k Z_k^t \quad (19)$$

Que preserva simetria. Para $i = k$, a expressão (19) seria:

$$\delta_k = S_{k+1} q_k = S_k q_k + a_k Z_k Z_k^t q_k \quad (20)$$

ou

$$Z_k = \frac{\delta_k - S_k q_k}{a_k Z_k^t q_k} \quad (21)$$

O produto interno por q_k^t em (20) fornece:

$$q_k^t \delta_k - q_k^t S_k q_k = a_k (Z_k^t q_k)^2 \quad (22)$$

Substituindo (21) em (19) tem-se:

$$S_{k+1} = S_k + \frac{(\delta_k - S_k q_k)(\delta_k - S_k q_k)^t}{q_k^t (\delta_k - S_k q_k)} \quad (23)$$

3.8 Método Rank Two Correction

O algoritmo DFP segue as mesmas idéias do algoritmo de “Rank One Correction”, mas nele a inversa da Hessiana é atualizada pela soma de duas matrizes simétricas de posto 1 e é denominada de rank two.

O algoritmo se inicia com uma aproximação de $H^{-1} \rightarrow S_0$ no ponto x_0 . Seguem-se os mesmos passos do rank one, mas a atualização de S_{k+1} é feita desse modo:

$$S_{k+1} = S_k + \frac{\delta_k \delta_k^t}{\delta_k^t q_k} - \frac{S_k q_k q_k^t S_k}{q_k^t S_k q_k} \quad (24)$$

3.9 Método BFGS

Método de Broyden-Fletcher-Goldfarb-Shanno (BFGS) para problemas de minimização irrestrita. Esse método possui o algoritmo mais bem sucedido para atualização da inversa. Ele garante as seguintes propriedades:

S_k simétrica $\rightarrow S_{k+1}$ simétrica
 S_k positiva definida $\rightarrow S_{k+1}$ positiva definida

A fórmula de recorrência é a seguinte:

$$S_{k+1} = \left[I - \frac{\delta_k q_k^t}{\delta_k^t q_k} \right] S_k \left[I - \frac{q_k \delta_k^t}{\delta_k^t q_k} \right] + \frac{\delta_k \delta_k^t}{\delta_k^t q_k} \quad (25)$$

O método BFGS é denominado muitas vezes na literatura como método DFP complementar. Um processo de busca linear para determinação de a_k e parte integrante do algoritmo.

4 Resolução de Um Elemento de Treliça

Deseja-se determinar os deslocamentos (u_1 e u_2) da treliça não-linear mostrada na Figura 1 através da formulação da energia potencial mínima, onde L é o comprimento da barra na direção 1, A é a área da seção transversal, E_1 é o módulo de elasticidade na direção 1, E_2 é o módulo de elasticidade na direção 2 e P é a carga aplicada no ponto B. Os valores das variáveis citadas são dados de entrada no programa implementado e são definidos por:

$$L = 400 \text{ cm}; A = 4 \text{ cm}^2; E_1 = 2 \times 10^6 \text{ kN/cm}^2; E_2 = -2.5 \times 10^8 \text{ kN/cm}^2, P = 8000 \text{ Kn} \quad (26)$$

A relação tensão-deformação não-linear é dada por:

$$\sigma = E_1 \varepsilon + E_2 \varepsilon^2 \quad (27)$$

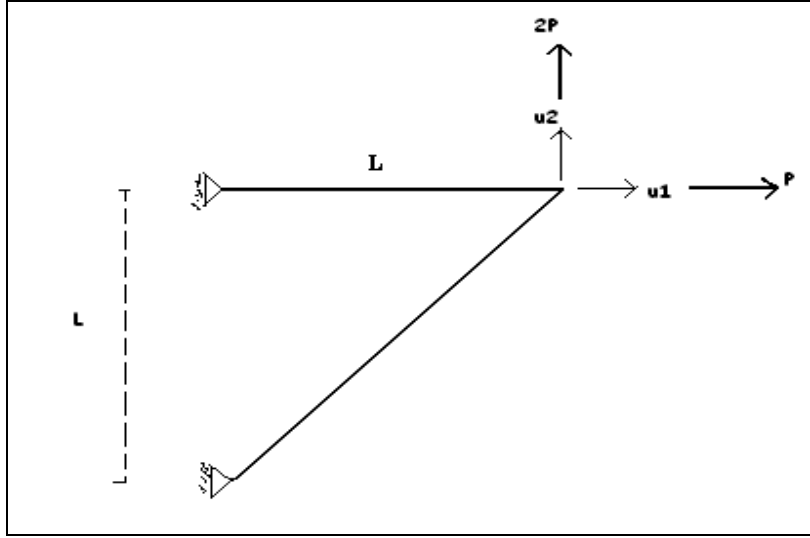


Figura 1: Treliça Não-Linear (Sacramento & Buffoni (2006))

A energia potencial total de um sistema é obtida através do somatório das energias de deformação, U , (Timoshenko & Gere, [4]). As energias são expressas por unidade de volume do material. Considerando-se um elemento sujeito à tensão e deformação, tem-se:

$$u = \int_0^{\varepsilon} \sigma d\varepsilon \quad (28)$$

$$U = \int u dV \quad (29)$$

Utilizando-se as equações de compatibilidade chegam-se as seguintes expressões:

$$\varepsilon_1 = \frac{\Delta L_1}{L_1} = \frac{u_1}{L} \quad (30)$$

$$\varepsilon_2 = \frac{\Delta L_2}{L_2} = \frac{(\cos 45^\circ u_1 + \cos 45^\circ u_2)}{L_2} = \left(\frac{\sqrt{2}}{2} u_1 + \frac{\sqrt{2}}{2} u_2 \right) \frac{1}{L\sqrt{2}} \quad (31)$$

$$\varepsilon_2 = (u_1 + u_2) \frac{1}{2L} \quad (32)$$

Fazendo uso das expressões (29), (31) e (32), chega-se às seguintes expressões para os deslocamentos:

$$U_{1,2} = AL \left(\frac{1}{2} \frac{E_1 u_1^2}{L^2} + \frac{1}{3} \frac{E_2 u_1^3}{L^3} \right) + AL\sqrt{2} \left(\frac{E_1 (u_1 + u_2)^2}{8 L^2} + \frac{E_2 (u_1 + u_2)^3}{24 L^3} \right) \quad (33)$$

Dessa forma, a energia potencial é dada pela seguinte expressão:

$$\Pi = AL \left(\frac{1}{2} \frac{E_1 u_1^2}{L^2} + \frac{1}{3} \frac{E_2 u_1^3}{L^3} \right) + AL\sqrt{2} \left(\frac{E_1 (u_1 + u_2)^2}{8 L^2} + \frac{E_2 (u_1 + u_2)^3}{24 L^3} \right) - 2Pu_1 - Pu_2 \quad (34)$$

onde L é o comprimento da barra na direção 1, A é a área da seção transversal, E_1 é o módulo de elasticidade na direção 1, E_2 é o módulo de elasticidade na direção 2 e P é a carga aplicada. A Figura 2 apresenta as curvas de nível para a energia potencial da expressão (34) e verifica-se que esta é mínima próxima aos pontos $u_1 = 0.5$ e $u_2 = 1$.

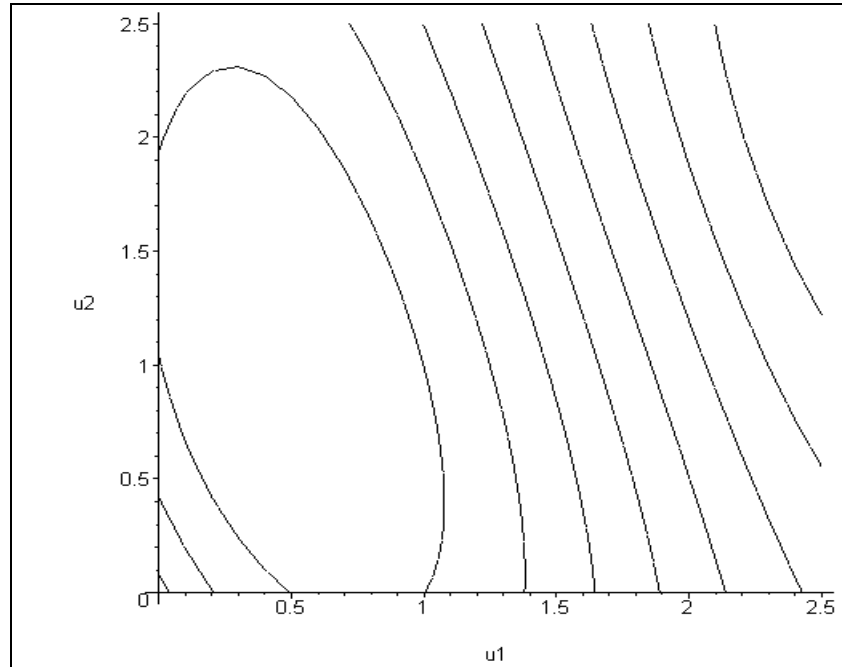


Figura 2: Curvas de nível da energia potencial expressa em (34).

4.1 Resolução de Um Elemento de Treliça Utilizando os Métodos Estudados

O objetivo é verificar a convergência dos métodos estudados para os valores de deslocamentos da treliça nas direções 1 e 2 que tornam a energia potencial mínima, onde os valores exatos apresentam-se na Figura 2. Dessa forma, realizaram-se testes para o ponto de partida é $\mathbf{x}_0 = (0,0)$, ou seja, $u_1 = 0$ e $u_2 = 0$.

4.1.1 Método Univariante

Com um passo $t = 0.000001$, o método apresentou os resultados da Tabela 1.

Tabela 1: Método Univariante – Ponto de partida (1,3).

ITERAÇÃO	u_1	u_2	Energia Potencial Mínima (Π)
1	0	0	-0.01599999
2	1E-006	0	-0.01599999
3	3E-006	0	-0.04799988
4	7E-006	0	-0.1119993

ITERAÇÃO	u_1	u_2	Energia Potencial Mínima (Π)
5	1.5E-005	0	-0.239997
6	3.1E-005	0	-0.495987
7	6.3E-005	0	-1.007946
8	0.000127	0	-2.031782
9	0.000255	0	-4.07912
10	0.000511	0	-8.172466
11	0.001023	0	-16.35384
12	0.002047	0	-32.6953
13	0.004095	0	-65.29319
14	0.008191	0	-130.1492
15	0.016383	0	-258.5058
16	0.032767	0	-509.8255
17	0.065535	0	-991.1171
18	0.131071	0	-1870.121
19	0.262143	0	-3308.304
20	0.524287	0	-5021.3
21	0.524287	1E-006	-5021.305
22	0.524287	3E-006	-5021.314
23	0.524287	7E-006	-5021.333
24	0.524287	1.5E-005	-5021.369
25	0.524287	3.1E-005	-5021.443
26	0.524287	6.3E-005	-5021.59
27	0.524287	0.000127	-5021.884
28	0.524287	0.000255	-5022.472
29	0.524287	0.000511	-5023.648
30	0.524287	0.001023	-5025.999
31	0.524287	0.002047	-5030.697
32	0.524287	0.004095	-5040.073
33	0.524287	0.008191	-5058.752
34	0.524287	0.016383	-5095.812
35	0.524287	0.032767	-5168.75
36	0.524287	0.065535	-5309.934
37	0.524287	0.131071	-5573.801
38	0.524287	0.262143	-6029.705
39	0.524287	0.524287	-6671.613
40	0.524287	1.048575	-7015.154
41	0.524286	1.048575	-7015.155
42	0.524284	1.048575	-7015.157
43	0.52428	1.048575	-7015.162
44	0.524272	1.048575	-7015.171
45	0.524256	1.048575	-7015.189
46	0.524224	1.048575	-7015.226
47	0.52416	1.048575	-7015.3

ITERAÇÃO	u_1	u_2	Energia Potencial Mínima (Π)
48	0.524032	1.048575	-7015.448
49	0.523776	1.048575	-7015.742
50	0.523264	1.048575	-7016.328
51	0.52224	1.048575	-7017.485
52	0.520192	1.048575	-7019.746
53	0.516096	1.048575	-7024.052
54	0.507904	1.048575	-7031.799
55	0.49152	1.048575	-7043.806
56	0.458752	1.048575	-7053.644
57	0.458752	1.048574	-7053.644
58	0.458752	1.048572	-7053.645
59	0.458752	1.048568	-7053.645
60	0.458752	1.04856	-7053.647
61	0.458752	1.048544	-7053.649
62	0.458752	1.048512	-7053.654
63	0.458752	1.048448	-7053.663
64	0.458752	1.04832	-7053.682
65	0.458752	1.048064	-7053.72
66	0.458752	1.047552	-7053.794
67	0.458752	1.046528	-7053.94
68	0.458752	1.04448	-7054.22
69	0.458752	1.040384	-7054.732
70	0.458752	1.032192	-7055.568
71	0.458752	1.015808	-7056.478
72	0.458752	1.015807	-7056.478
73	0.458752	1.015805	-7056.478
74	0.458752	1.015801	-7056.478
75	0.458752	1.015793	-7056.478
76	0.458752	1.015777	-7056.478
77	0.458752	1.015745	-7056.479
78	0.458752	1.015681	-7056.481
79	0.458752	1.015553	-7056.484
80	0.458752	1.015297	-7056.49
81	0.458752	1.014785	-7056.501
82	0.458752	1.013761	-7056.52
83	0.458752	1.011713	-7056.546
84	0.458752	1.007617	-7056.549
85	0.458752	1.007618	-7056.549
86	0.458752	1.00762	-7056.549
87	0.458752	1.007624	-7056.549
88	0.458752	1.007632	-7056.549
89	0.458752	1.007648	-7056.55
90	0.458752	1.00768	-7056.55

ITERAÇÃO	u_1	u_2	Energia Potencial Mínima (Π)
91	0.458752	1.007744	-7056.55
92	0.458752	1.007872	-7056.551
93	0.458752	1.008128	-7056.552
94	0.458752	1.00864	-7056.554
95	0.458752	1.009664	-7056.555
96	0.458752	1.009663	-7056.555
97	0.458752	1.009661	-7056.555
98	0.458752	1.009657	-7056.555
99	0.458752	1.009649	-7056.556
100	0.458752	1.009633	-7056.556
101	0.458752	1.009601	-7056.556
102	0.458752	1.009537	-7056.556
103	0.458752	1.009409	-7056.556
104	0.458752	1.00941	-7056.556
105	0.458752	1.009412	-7056.556
106	0.458752	1.009416	-7056.556
107	0.458752	1.009424	-7056.556
108	0.458752	1.009423	-7056.556
109	0.458752	1.009421	-7056.556
110	0.458752	1.00942	-7056.556

A Figura 3 mostra o caminho descrito pelo método Univariante.

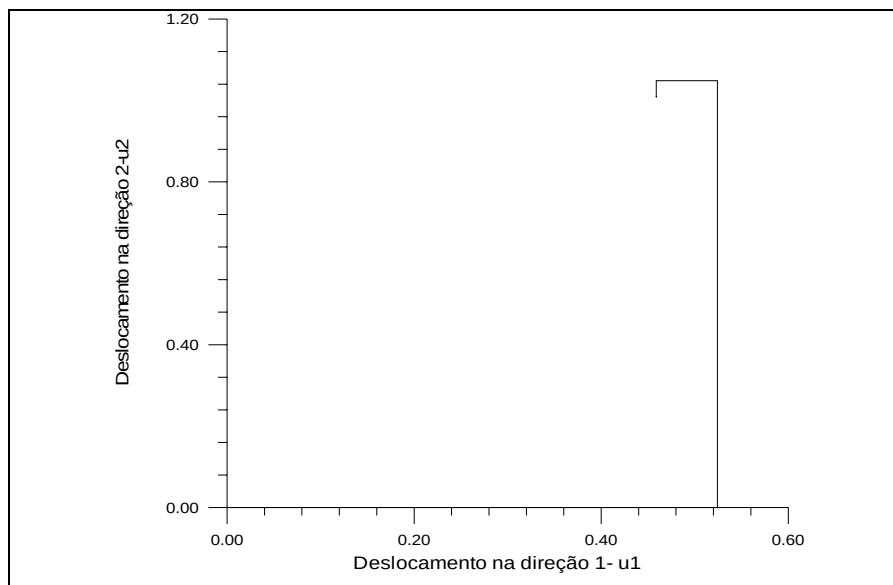


Figura 3: Convergência do método Univariante para o ponto onde a energia potencial é mínima. Ponto de partida (0,0).

4.1.2 Método do Máximo Declive

Com um passo $t = 0.000001$, o método apresentou os resultados da Tabela 2.

Tabela 2: Método do Máximo Declive – Ponto de partida (0,0).

ITERAÇÃO	u_1	u_2	Energia Potencial Mínima (Π)
1	0	0	0
2	0.4689374	0.2344687	-5773.422
3	0.6541881	0.4199456	-6507.465
4	0.4453029	0.9008249	-7024.204
5	0.5016987	0.9342088	-7047.586
6	0.4709735	0.9455975	-7052.063
7	0.4814727	0.9581753	-7054.520
8	0.4637894	0.9857623	-7056.416
9	0.4710845	0.9894892	-7057.094
10	0.4695538	0.9975370	-7057.239
11	0.4686248	0.9979678	-7057.243
12	0.4689887	0.9983280	-7057.245
13	0.4684940	0.9992246	-7057.247
14	0.4687252	0.9993346	-7057.247
15	0.4686671	0.9993617	-7057.248
16	0.4686701	0.9994896	-7057.248
17	0.4686383	0.9994926	-7057.248
18	0.4686443	0.9995240	-7057.248
19	0.4686284	0.9995253	-7057.248
20	0.4686344	0.9995305	-7057.248

A Figura 4 apresenta o caminho descrito pelo método do máximo declive.

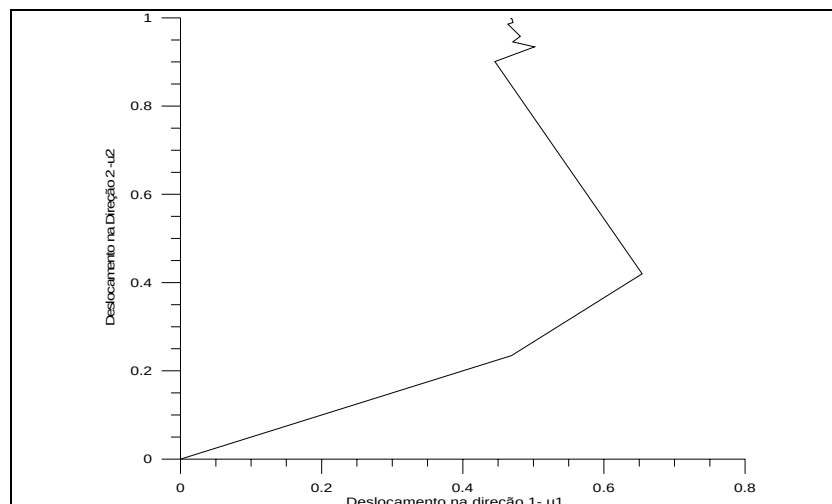


Figura 4: Convergência do método do Máximo Declive para o ponto onde a energia potencial é mínima. Ponto de partida (0,0).

4.1.3 Método de Powell

Com um passo $t = 0.000001$, o método apresentou os resultados da Tabela 3.

Tabela 3: Método de Powell – Ponto de partida (0,0).

ITERAÇÃO	u_1	u_2	Energia Potencial Mínima (Π)
1	0	0	-
2	0.626612	0.402244	-6488.465
3	0.514226	0.903053	-7037.739
4	0.497187	0.934596	-7048.978
5	0.481182	0.960045	-7054.741
6	0.470234	0.989313	-7057.087
7	0.469695	0.993630	-7057.195
8	0.469463	0.997154	-7057.238
9	0.469067	0.998240	-7057.245
10	0.468873	0.998791	-7057.247
11	0.468873	0.999233	-7057.247
12	0.468657	0.999377	-7057.248
13	0.468646	0.999447	-7057.248
14	0.468642	0.999503	-7057.248
15	0.468637	0.999518	-7057.248
16	0.468634	0.999533	-7057.248
17	0.468632	0.999538	-7057.248
18	0.468632	0.999538	-7057.248
19	0.468631	0.999541	-7057.248
20	0.468631	0.999541	-7057.248
21	0.468631	0.999541	-7057.248
22	0.468630	0.999543	-7057.248
23	0.468630	0.999543	-7057.248
24	0.468630	0.999543	-7057.248
25	0.468630	0.999543	-7057.248
26	0.468630	0.999543	-7057.248
27	0.468630	0.999543	-7057.248
28	0.468630	0.999543	-7057.248
29	0.468630	0.999543	-7057.248
30	0.468630	0.999543	-7057.248

A Figura 5 mostra o caminho descrito pelo método de Powell.

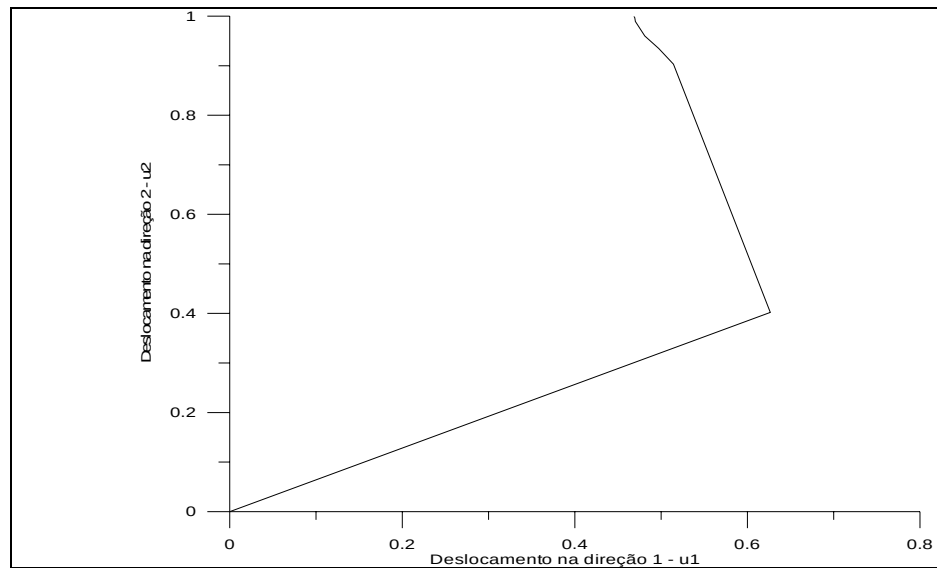


Figura 5: Convergência do método de Powell para o ponto onde a energia potencial é mínima. Ponto de partida (0,0).

4.1.4 Método de Fletcher-Reeves

Com um passo $t = 0.000001$, o método apresentou os seguintes resultados:

Tabela 4: Método de Fletcher-Reeves – Ponto de partida (0,0).

ITERAÇÃO	u_1	u_2	Energia Potencial Mínima (Π)
1	0	0	0
2	0.7502999	0.3751500	0
3	0.5788475	0.4959164	-6303.224
4	0.6029188	0.7042455	-6655.573
5	0.5116172	0.7558123	-6883.635
6	0.5385898	0.8571415	-6964.260
7	0.4902307	0.8773943	-7013.171
8	0.5022644	0.9284234	-7034.250
9	0.4783418	0.9391426	-7046.619
10	0.485991	0.964216	-7051.618
11	0.473804	0.969040	-7054.507
12	0.477007	0.981750	-7055.825
13	0.471011	0.984394	-7056.582
14	0.472972	0.990647	-7056.895
15	0.469928	0.999288	-7057.075
16	0.470723	0.995037	-7057.158
17	0.469231	0.995714	-7057.205
18	0.469718	0.997278	-7057.225
19	0.468961	0.997591	-7057.237

ITERAÇÃO	u_1	u_2	Energia Potencial Mínima (Π)
20	0.469154	0.998387	-7057.242
21	0.4687856	0.998564	-7057.245
22	0.4689022	0.998957	-7057.246
23	0.4687154	0.999041	-7057.247
24	0.4687619	0.999241	-7057.247
25	0.4686714	0.991858	-7057.247
26	0.4686976	0.999387	-7057.247
27	0.4686523	0.999411	-7057.248
28	0.4686630	0.999461	-7057.248
29	0.4686412	0.999475	-7057.248
30	0.4686464	0.999500	-7057.248
31	0.4686358	0.999507	-7057.248
32	0.4686398	0.999532	-7057.248
33	0.4686334	0.999532	-7057.248

A Figura 6 mostra o caminho descrito pelo método de Fletcher Reeves.

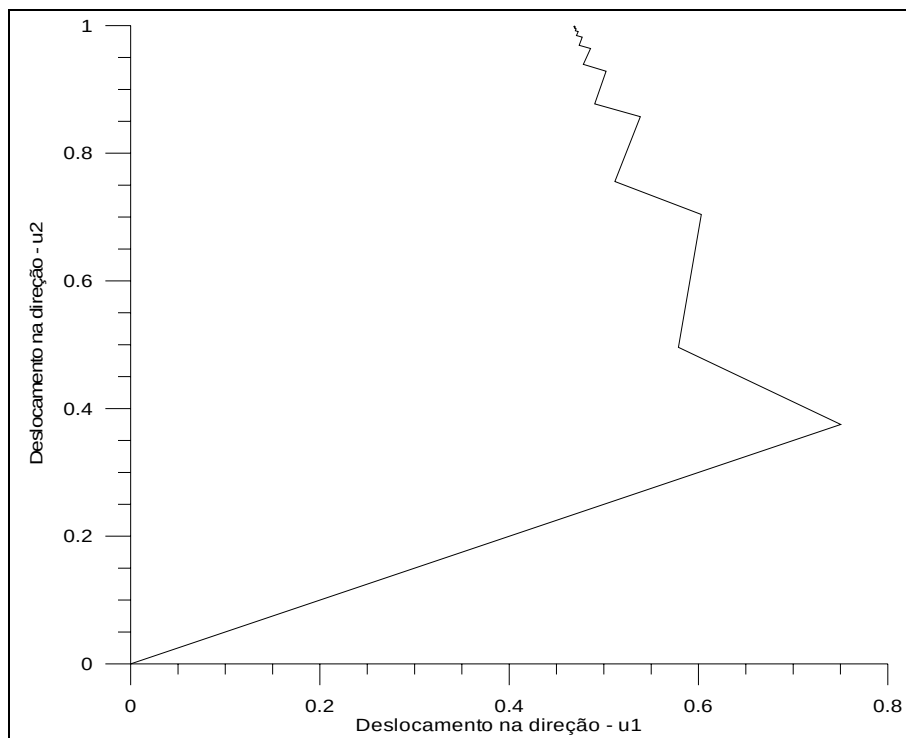


Figura 6: Convergência do método de Fletcher-Reeves para o ponto onde a energia potencial é mínima. Ponto de partida (0,0).

4.1.5 Método de Newton Raphson Padrão

A Tabela 5 apresenta os resultados para um passo $t = 1$.

Tabela 5: Método de Newton Raphson Padrão – Ponto de partida (0,0).

Iteração	u_1	u_2	Energia Potencial Mínima (Π)
1	0	0	0
2	0.4	0.7313709	-6792.15
3	0.4666667	0.9740878	-7055.774
4	0.4686275	0.9993314	-7057.247
5	0.4686292	0.9995434	-7057.248

A Figura 7 mostra o caminho descrito pelo método de Newton Raphson Padrão.

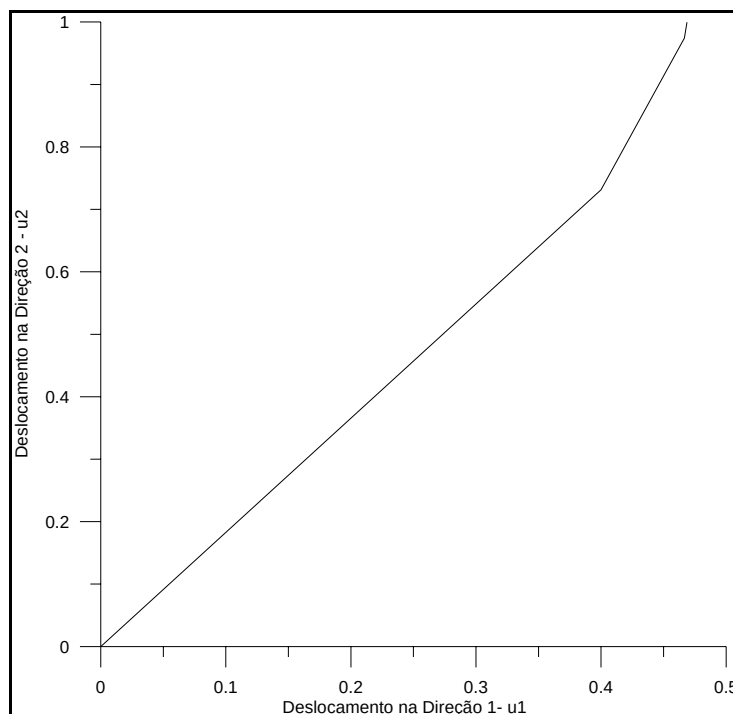


Figura 7: Convergência do método de Newton Raphson Padrão para o ponto onde a energia potencial é mínima. Ponto de partida (0,0).

4.1.6 Método Newton Raphson Modificado

A Tabela 6 apresenta os resultados para um passo $t = 1$.

Tabela 6: Método de Newton Raphson Modificado – Ponto de partida (0,0).

ITERAÇÃO	u_1	u_2	Energia Potencial Mínima (Π)
1	0	0	0
2	0.4	0.7313709	-6792.15
3	0.45	0.8813709	-7018.028
4	0.4632813	0.9450503	-7050.114
5	0.4670717	0.974205	-7055.839

ITERAÇÃO	u_1	u_2	Energia Potencial Mínima (Π)
6	0.4681737	0.9877719	-7056.959
7	0.4684958	0.9940903	-7057.188
8	0.4685901	0.9970242	-7057.235
9	0.4686177	0.9983821	-7057.245
10	0.4686258	0.9990089	-7057.247
11	0.4686282	0.9992976	-7057.247
12	0.4686289	0.9994305	-7057.248
13	0.4686291	0.9994915	-7057.248
14	0.4686291	0.9995196	-7057.248
15	0.4686291	0.9995324	-7057.248

A Figura 8 mostra o caminho descrito pelo método de Newton Raphson Modificado.

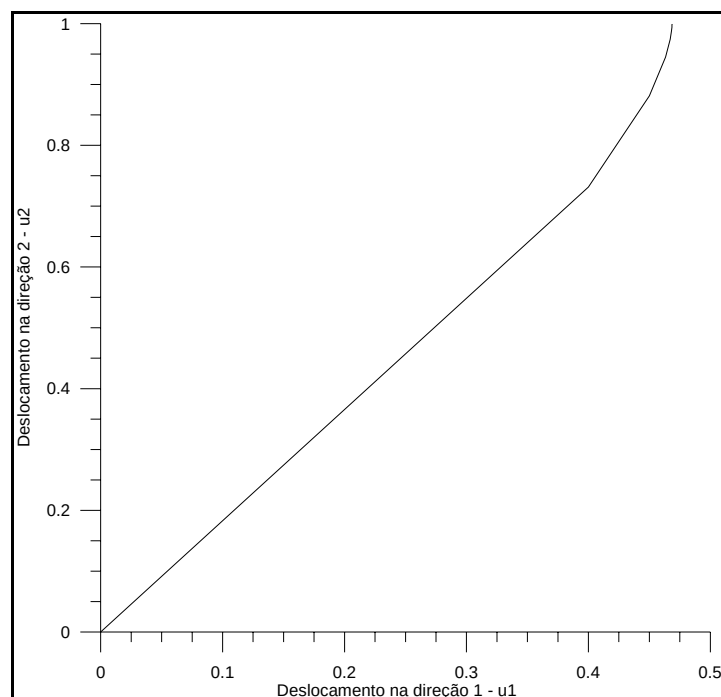


Figura 8: Convergência do método de Newton Raphson Modificado para o ponto onde a energia potencial é mínima. Ponto de partida (0,0).

4.1.7 Método Correção de Posto Um (Rank One Correction)

Com um passo $t = 0.000001$, o método apresentou os resultados da Tabela 7.

Tabela 7: Método Rank One Correction– Ponto de partida (0,0).

ITERAÇÃO	u_1	u_2	Energia Potencial Mínima (Π)
1	0	0	0
2	.5031503	.9199736	-7044.990
3	.5036367	.9208747	-7044.993
4	.4966970	.9165216	-7045.883
5	.5017961	.9101100	-7043.422

A Figura 9 mostra o caminho descrito pelo método Rank One Correction.

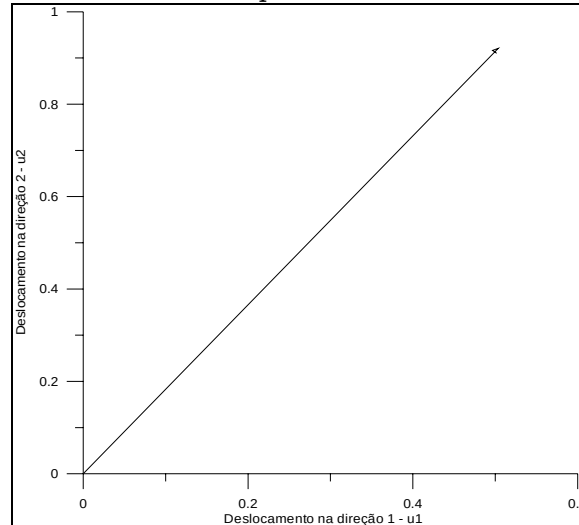


Figura 9: Convergência do método Rank One Correction para o ponto onde a energia potencial é mínima. Ponto de partida (0,0).

4.1.8 Método Correção de Posto Dois (Rank Two Correction)

Com um passo $t = 0.000001$, o método apresentou os seguintes resultados:

Tabela 8: Método Rank Two Correction – Ponto de partida (0,0).

ITERAÇÃO	u_1	u_2	Energia Potencial Mínima (Π)
1	0	0	0
2	.4689374	.2344687	-5773.422
3	.7016331	.3551842	-6350.297
4	.5859184	.4167470	-6511.552
5	0.4888443	0.5048179	-6584.202
6	0.5261520	1.027777	-7020.407
7	0.5106777	0.9640939	-7044.817
8	0.4925829	0.9367751	-7050.315
9	0.4780165	0.9442755	-7052.563
10	0.4640229	0.9739052	-7055.337
11	0.4677314	1.006463	-7057.172
12	0.4683698	0.9982956	-7057.243
13	0.4688688	0.9981811	-7057.245
14	0.4690156	0.9983908	-7057.245
15	0.4688557	0.9994022	-7057.247
16	0.4686606	0.9995679	-7057.247
17	0.4686338	0.9995854	-7057.247

A Figura 10 mostra o caminho descrito pelo método Rank Two Correction.

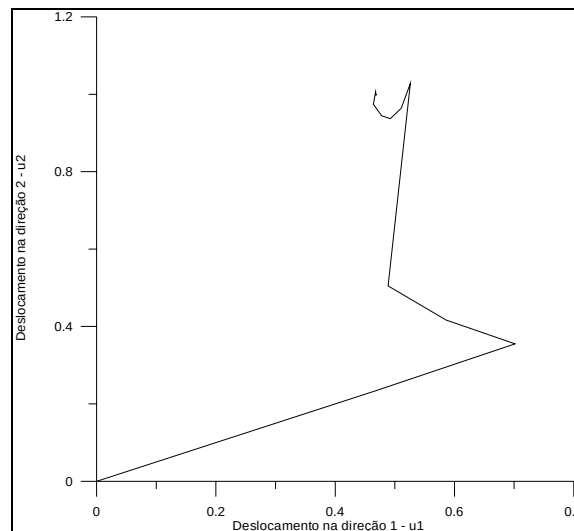


Figura 10: Convergência do método Rank One Correction para o ponto onde a energia potencial é mínima. Ponto de partida (0,0).

4.1.9 Método BFGS

Com um passo $t = 0.000001$, o método apresentou os resultados da Tabela 9.

Tabela 9: Método BFGS – Ponto de partida (0,0).

ITERAÇÃO	u_1	u_2	Energia Potencial Mínima (Π)
1	0	0	0
2	0.4689374	0.2344687	-5773.422
3	0.3274954	0.4551804	-5892.787
4	0.589635	0.4536538	-6583.731
5	0.4506282	0.6759072	-6816.933
6	0.3990856	0.9329341	-6985.949
7	0.4409088	0.8824784	-7010.541
8	0.4620554	0.8574471	-7013.439
9	0.4217083	0.9821547	-7033.454
10	0.4607974	0.9295524	-7045.051
11	0.4645898	0.9949785	-7056.99
12	0.4669828	0.9916543	-7057.054
13	0.4710787	0.9916351	-7057.148
14	0.4686867	0.9949601	-7057.208
15	0.4677857	0.9967992	-7057.218
16	0.468425	0.9959994	-7057.22
17	0.4691035	0.9979317	-7057.243
18	0.4684253	0.998699	-7057.245
19	0.4686766	0.9986504	-7057.246
20	0.4687946	0.9991486	-7057.247
21	0.4685989	-7057.247	0.9993136
22	0.4686617	0.9993017	-7057.247
23	0.4685958	0.9994114	-7057.247
24	0.4686512	0.9993793	-7057.247
25	0.4686287	0.9994021	-7057.247
26	0.4686349	0.999397	-7057.247

A Figura 11 mostra o caminho descrito pelo método BFGS.

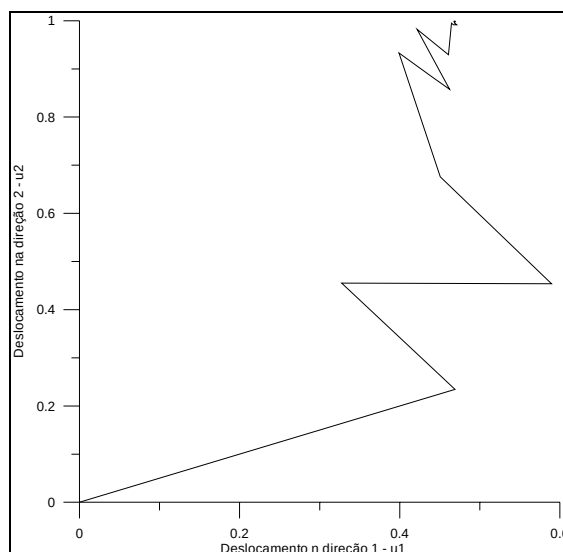


Figura 11 - Convergência do método BFGS. Ponto de partida (0,0).

5 Conclusões

Através dos resultados obtidos, concluiu-se que os deslocamentos u_1 e u_2 obtidos através de todos os métodos são compatíveis com a solução exata apresentada. Nota-se também que o método univariante, é um pouco “lento”, sendo necessárias muitas iterações para se chegar a uma solução ótima do problema. O método de Powell mostra-se mais eficiente que o univariante, pois a solução exata foi obtida com menos iterações. Já os métodos de Newton Raphson Padrão e Rank One Correction se mostraram mais eficientes do ponto de vista do número de iterações até se chegar a solução do problema. A Tabela 10 apresenta uma comparação entre os métodos.

Tabela 10: Comparação entre os métodos.

Método	Ponto de Partida	Número de Iterações	Deslocamentos		Energia Potencial Mínima (Π)
			u_1	u_2	
Univariante	(0,0)	110	0.458752	1.00942	-7056.556
Máximo Declive	(0,0)	20	0.4686344	0.9995305	-7057.248
Powell	(0,0)	30	0.4686304	0.9995436	-7057.248
Método de Fletcher – Reeves	(0,0)	33	0.4686334	0.999532	-7057.248
Newton Raphson Padrão	(0,0)	5	0.4686292	0.9995434	-7057,248
Newton Raphson Modificado	(0,0)	15	0.4686291	0.9995324	-7057.248
Correção de Posto Um (Rank One Correction)	(0,0)	5	0.5017961	0.9101100	-7043.422
Correção de Posto Dois (Rank Two Correction)	(0,0)	(0,0)	17	0.4686338	0.9995854
BFGS	(0,0)	26	0.4686349	0.999397	-7057.247

6 Agradecimentos

Sinceros agradecimentos à Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro – FAPERJ, pelo apoio financeiro.

7 Referências

- Filho, J. V. C. (2004) Pesquisa Operacional – Técnicas de Otimização Aplicadas a Sistemas Agroindustriais. 2ª. Edição. Editora Atlas. São Paulo.
- Friedlander, A. (1994) Elementos de Programação Não-Linear. Editora Unicamp. Campinas. São Paulo.
- Gill, P. E; Murray, W.; Wright, M. (1981) Practical optimization. Academic Press. Nova York.
- Timoshenko, S. P.; Gere, J. E. (1994) Mecânica dos Sólidos. LTC Editora. Rio de Janeiro.
- Sacramento, M.A., Buffoni, S.S.O. (2006) Programação Não-Linear Aplicada ao Cálculo de Deslocamentos em Elementos de Treliça. Profundão – 10º Encontro de Engenharia de Produção da UFRJ. Rio de Janeiro.

REVISTA INVESTIGAÇÃO OPERACIONAL

Política Editorial

Investigação Operacional (IO) é a revista científica da APDIO - Associação Portuguesa de Investigação Operacional. A política editorial da IO é publicar artigos originais e de elevada qualidade que contribuam para a teoria, metodologia, técnicas e software de Investigação Operacional e a sua aplicação a diferentes campos. A Revista também publica artigos com revisões relevantes de temas de IO. Casos de sucesso na aplicação a problemas práticos são especialmente bem vindos.

Processo de Aceitação

Todos os manuscritos submetidos para publicação são revistos e aceites apenas com base na avaliação da sua qualidade, importância e adequação à política editorial. Será responsabilidade do Editor interpretar a avaliação dos revisores. A contribuição de cada artigo deve estar claramente evidenciada na Introdução. Critérios como a relação com literatura existente, comprimento e estilo do artigo são tidos em consideração. Uma indicação clara da viabilidade de aceitação do artigo é habitualmente dada na primeira fase de revisão do artigo.

Será requerido aos autores de um artigo aceite que transfiram os direitos de autoria para a APDIO, que assegurará a mais ampla disseminação possível de informação. Os volumes da Revista são publicados em papel, e distribuídos a todos os associados da APDIO, e em formato electrónico na rede SciELO - Scientific Electronic Library Online.

Resumos dos Artigos indexados em

IAOR - International Abstracts in Operations Research

Instruções aos Autores

1. *Submeter artigos* para publicação ao editor principal, de preferência por e-mail em Microsoft Word ou "Portable Document Format" (PDF) para jfo@fe.up.pt, ou por correio normal (quatro cópias) para o seguinte endereço: Prof. José Fernando Oliveira, Departamento de Engenharia Electrotécnica e de Computadores, Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 PORTO, Portugal.
2. *Língua.* Os artigos devem ser escritos em Português, Inglês ou Espanhol.
3. *Os Manuscritos* devem ser impressos. Numerar as páginas consecutivamente.
4. *A primeira página* do manuscrito escrito em português ou em espanhol deve ter a seguinte informação: (a) Título; (b) nome, e-mail e afiliação institucional dos autores; (c) um resumo; (d) palavras-chave; (e) título em inglês (f) um resumo em inglês; (g) palavras-chave em inglês; (h) identificação do autor correspondente. Se o manuscrito for escrito em inglês, a primeira página deve ter a seguinte informação: (a) Título em inglês; (b) nome, e-mail e afiliação institucional dos autores; (c) um resumo em inglês; (d) palavras-chave em inglês; (e) identificação do autor correspondente.

5. *Agradecimentos*, incluindo informação sobre apoios, dever ser colocados imediatamente antes da secção de referências.
6. *Notas de rodapé* devem ser evitadas.
7. *Formulas* que são referenciadas devem ser numeradas consecutivamente ao longo do manuscrito como (1), (2), etc. do lado direito.
8. *Figuras*, incluindo grafos e diagramas, devem ser numerados consecutivamente em numeração árabe.
9. *Tabelas* devem ser numeradas consecutivamente em numeração árabe.
10. *Referências*. Citar apenas as mais relevantes e listar só as que são citadas no texto. Indicar as citações no texto através de parênteses rectos, e.g., [4]. No final do artigo listar as referências alfabeticamente por apelido do primeiro autor e numerá-las consecutivamente, de acordo com o seguinte formato: Artigos: autore(s), título, nome e volume da revista (ou livro, mas neste caso incluir o nome dos editores), ano e páginas. Livros: Autor(es), título, editor, ano.
11. *Artigos aceites* devem ser enviados pelo autor ao editor, de preferência na forma de um ficheiro fonte em LaTeX com ficheiros EPS para as figuras, juntamente com um ficheiro PDF ou Postscript. Em alternativa, ficheiros fonte em Word são também aceites. Para garantir uma boa qualidade gráfica, as figuras devem ser em formato vectorial; formatos raster como JPG, BMP, GIF, etc. devem ser evitados.
12. *Provas dos artigos* serão enviadas por e-mail como ficheiros PDF para o autor correspondente. Corrigir as provas cuidadosamente, e restringir as correcções apenas aos pontos em que as provas diferem do manuscrito. Desvios à versão aceite pelo editor são apenas possíveis com a autorização prévia e explícita do editor. Trinta separatas de cada artigo são enviados gratuitamente ao autor correspondente.

Informação sobre a Publicação

Investigação Operacional (ISSN 0874-5161) está registada na Secretaria de Estado da Comunicação Social sob o número 108335. Os volumes da Revista são publicados em papel, e distribuídos a todos os associados da APDIO, e em formato electrónico na rede SciELO - Scientific Electronic Library Online. O preço da assinatura anual é de 25 euros. Os volumes são enviados por correio normal. Informação adicional sobre assinaturas pode ser solicitada ao Secretariado da APDIO- CESUR, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 LISBOA, Portugal. Tel. +351 218 407 455 - www.apdio.pt - apdio@cesur.civil.ist.utl.pt

JOURNAL INVESTIGAÇÃO OPERACIONAL

Editorial Policy

Investigação Operacional (IO) is the scientific journal of APDIO - Associação Portuguesa de Investigação Operacional (the Portuguese Operational Research Association). The editorial policy of IO is to publish high quality and original articles that contribute to theory, methodology, techniques and software of Operational Research (OR) and its application to different fields. It also publishes articles with relevant reviews of OR subjects. Cases of successful application of OR to practical problems are specially welcome.

Acceptance Process

All manuscripts submitted for publication are refereed and accepted only on the basis of its quality, importance and adequacy to the editorial policy. It will be the responsibility of the Editor to interpret the referee's assessment. The contribution of each paper should be clearly stated in the introduction. Criteria such as relationship with existing literature, length and style are taken into account. A clear indication on the suitability of a manuscript is usually provided after the first round of refereeing. The authors of an accepted paper will be asked to transfer its copyright to the publisher, which will ensure the widest possible dissemination of information. The volumes of the journal are published in hardcopies, which are distributed to all APDIO associates, and in electronic format in SciELO - Scientific Electronic Library Online.

Articles are abstracted/indexed in

IAOR - International Abstracts in Operations Research

Instructions to Authors

1. *Submit papers* for publication to the main editor, preferably by e-mail in Microsoft Word or "Portable Document Format"(PDF) to jfo@fe.up.pt, or by ordinary mail (four copies) to the following address: Prof. José Fernando Oliveira, Departamento de Engenharia Electrotécnica e de Computadores, Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 PORTO, Portugal.
2. *Language.* Papers must be written in Portuguese, English or Spanish.
3. *Manuscripts* should be typewritten or typeset. Number the pages consecutively.
4. *The first page* of the manuscript written in English should contain the following information: (a) Title; (b) names, e-mails and institutional affiliations of the authors; (c) an abstract; (d) keywords (f) identification of the corresponding author.
5. *Acknowledgements*, including support information, should be placed prior to the references section.
6. *Footnotes* should be avoided.

7. *Formulas* that are referred to should be numbered consecutively throughout the manuscript as (1), (2), etc. on the right.
8. *Figures*, including graphs and diagrams, should be numbered consecutively in Arabic numbers.
9. *Tables* should be numbered consecutively in Arabic numbers.
10. *References*. Cite only the most relevant references and list only those cited in the text. Indicate citations in the text by bracketed numbers, e.g., [4]. At the end of the paper list the references alphabetically by the surname of the first author and number them consecutively, according to the following formats: Articles: author(s), title, name and number of the journal (or book, but in this case include the editors names), year, pages. Books: Author(s), title, publisher, year.
11. *Accepted papers* are to be sent by the author to the editor, preferably in the form of a source file in LaTeX and EPS files for the figures together with a PDF or postscript file. Alternatively, source files in Word are also accepted. To ensure good publishing quality the figures should be in vector formats; raster formats like JPG, BMP, GIF, etc. should be avoided.
12. *Page proofs* will be e-mailed as a PDF file to the corresponding author. Correct proofs carefully, and restrict corrections to points at which the proof is at variance with the manuscript. Deviations from the version accepted by the editor are only possible with the prior and explicit approval of the editor. Thirty offprints of each paper are supplied free of charge to the corresponding author.

Publication information

Investigação Operacional (ISSN 0874-5161) is registered in the *Secretaria de Estado da Comunicação Social* under number 108335. The volumes of the journal are published in hardcopies, which are distributed free of charge to all APDIO associates, and in electronic format in SciELO - Scientific Electronic Library Online. Subscription price is 25 euros. Issues are sent by standard mail. Additional subscription information is available upon request from APDIO Secretariat - CESUR, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 LISBOA, Portugal. Tel. +351 218 407 455 - www.apdio.pt - apdio@cesur.civil.ist.utl.pt

ÍNDICE

M. Sellito, J. Ribeiro, F. Petrillo <i>Modelo não-linear de longo prazo para a potência requerida do sistema brasileiro de eletricidade</i>	1
L. Meza, J. Mello, E Gomes, A. Fernandes <i>Seleção de variáveis em DEA aplicada a uma análise do mercado de energia eléctrica</i>	21
J. Catalão, S. Mariano, V. Mendes, L. Ferreira <i>Optimização da Exploração de Recursos Hídricos considerando o Efeito de Queda</i>	37
L. Torres Guardia e J. Soares de Mello <i>Experiências computacionais com modelos de fluxo para múltiplos produtos com funções de custo não lineares e não separáveis</i>	53
J. Castro, T. Baidya, F. Aiube <i>Uso da Simulação de Monte Carlo e da Curva de Gatilho na Avaliação de Opções de Venda Americanas</i>	67
S. Buffoni, A. Moreira, M. Sacramento <i>Minimização da Energia Potencial de Sistemas Estruturais Não-Lineares: uma comparação de algoritmos numéricos de programação não-linear</i>	85