UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
CURSO DE PÓS-GRADUAÇÃO EM MATEMÁTICA APLICADA

Uma implementação do Método do Gradiente Projetado na solução do problema não-linear de controle do conversor catalítico.

por

João Batista da Paz Carvalho

Dissertação submetida como requisito parcial
para a obtenção do grau de
Mestre em Matemática Aplicada

Prof. Júlio Claeyssen
Orientador

Porto Alegre, Março de 1996.

CIP - CATALOGAÇÃO NA PUBLICAÇÃO

Carvalho, João Batista da Paz

Uma implementação do Método do Gradiente Projetado na solução do problema não-linear de controle do conversor catalítico. / João Batista da Paz Carvalho.—Porto Alegre: CPGMAp da UFRGS, 1996.

78 p.: il.

Dissertação (mestrado) —Universidade Federal do Rio Grande do Sul, Curso de Pós-Graduação em Matemática Aplicada, Porto Alegre, 1996. Orientador: Claeyssen, Júlio

Dissertação: Engenharia Matemática e Matemática Industrial.
controle, fronteira livre, otimização não linear, simulação .

AGRADECIMENTOS

Agradeço ao professor Júlio Claeyssen pela orientação e aconselhamento durante todas as etapas desta tarefa.

Agradeço, com toda sinceridade , ao professor Mark Thompson, por suas sugestões, presteza e interesse para com meu trabalho.

Agradeço ao professor Vilmar Trevisan pelo apoio e pelo incentivo nos meus estudos.

Agradeço à Capes pela bolsa de estudo concedida, um valioso incentivo à presente realização .

Agradeço ao Laboratório de Recursos Computacionais do CPGMap pelo suporte e pela ampla disponibilidade de seu equipamento.

SUMÁRIO

LISTA DE FIGURAS

RESUMO

O presente trabalho trata da formulação , algoritmização e implementação numérica de um problema não-linear de controle de fronteria livre sujeito a restrições também não lineares, definido em [Fri 94] e relativo ao modelo de funcionamento de um conversor catalítico monolítico cerâmico.

São apresentados resultados de algumas simulações numéricas, usando um programa em FORTRAN77, no ambiente de estações de trabalho SUN e DEC alfa 3000.

ABSTRACT

This work describes the formulation, algoritmization and numerical implementation of a non-linear free boundary control problem which is subjected to non-linear restritions also. This problem concerns upon a working model of a ceramic monolitic catalitic converter and is defined in [Fri 94].

We present here results of some numeric simulations, using a FORTRAN77 program, done in workstations SUN and DEC alpha 3000 enviroment.

INTRODUÇÃO

O conversor catalítico é um equipamento situado no sistema de exaustão do automóvel. Após os gases poluentes serem expelidos do motor, passam através do conversor catalítico (catalizador) e tomam parte de reações químicas que os convertem em gases menos nocivos. Atualmente, conversores mais usados são os monolíticos cerâmicos, que são reatores tubulares dentro dos quais os gases fluem reagindo com suas paredes. A figura (1) mostra um esboço desse equipamento.

Em escala industrial, existe a preocupação de melhorar o desempenho desses equipamentos, predizendo a emissão estacionária nos principais regimes de funcionamento do motor, isto é, na auto-estrada, no engarrafamento, na serra, na garagem, etc. Na prática, estuda-se seu comportamento no regime de aquecimento do motor, ocasião em que há maior descarga e concentração de poluentes devido à má combustão .

Um modelo em E.D.Ps não -lineares do catalizador do tipo monolítico cerâmico foi desenvolvido por Cavendish [Cav 80, Cav 85] e melhorado por Friedman [Fri 91]; permite identificar a busca por um melhor desempenho nesses equipamentos como um problema de otimização num espaço de funções contínuas por segmentos. Em termos sucintos, equivale a encontrar uma calibragem ótima que minimize um funcional que contabiliza as concentrações de poluente na saída do equipamento.

O presente trabalho se propõe a implementar numericamente um algoritmo de busca à calibragem ótima do conversor para um regime arbitrado de funcionamento do motor, e segundo um modelo simplificado proposto em [Fri 91].

O primeiro capítulo atenta para a descrição do modelo e formulação do problema de otimização , cuja solução é a própria calibragem ótima que buscamos.

No segundo capítulo, fazemos uma breve análise do modelo não - linear adotado. Nossas argumentações analíticas demandam alguns resultados da Análise Funcional Aplicada, e que por esta razão estão desenvolvidos no Apêndice.

No terceiro capítulo, discutimos algumas técnicas computacionais que se fazem indispensáveis à realização de nossa tarefa. Tais técnicas, que são apresentadas na forma descritiva ou mesmo já na forma algorítmica, se constituem em grandes sub-tarefas numéricas a serem realizadas.

No quarto capítulo, descrevemos a metodologia que nos conduzirá à algoritmização do procedimento global de solução de nosso problema de controle ótimo. Tal metodologia está presente na abundante literatura de Otimização Não -linear dos dias de hoje.

No quinto capítulo, descrevemos e dimensionamos a tarefa computacional envolvida pelo presente trabalho. Nesse sentido discutimos questões como comple- xidade computacional e detalhes da implementação numérica. Também são apresentados os resultados de algumas simulações , bem como evidências computacionais que sobrevêem.

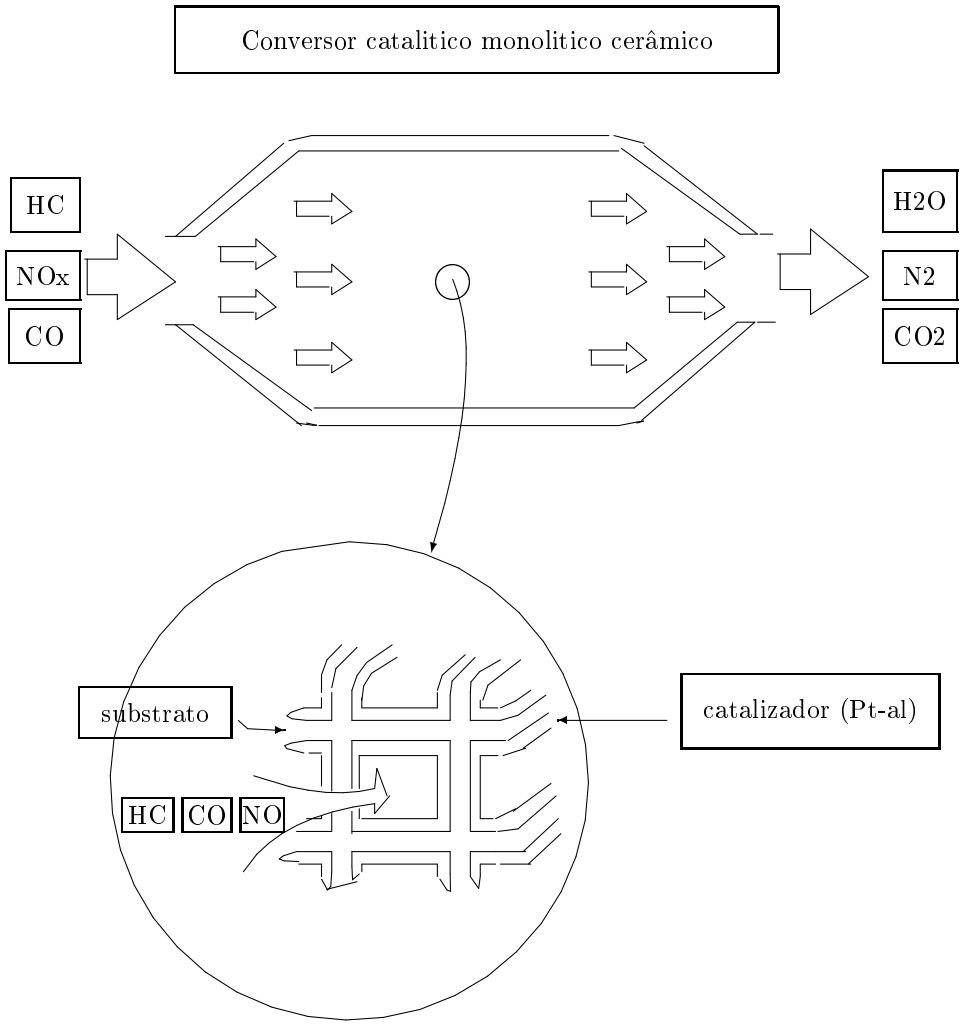

Figura 1: Esquema do conversor catalítico.

Dentro do catalizador, com os gases de escape a temperaturas acima de 300^oC , processam-se as reações químicas que transformam os gases poluentes em substâncias inofensivas. Seu corpo cerâmico tem minúsculos canais revestidos por uma camada de óxido de alumínio, com grande área superficial, onde também se encontram os metais nobres paládio/molibdênio . Em contato com esses metais, os poluentes CO, HC e NOx transformam-se em água, gás carbônico, nitrogênio e nitrogênio puro.

O catalizador desenvolvido pela Autolatina, no Brasil, é do tipo *Three Way Catalyst*, que transforma os três gases poluentes de uma só vez. Externamente ele é revestido com uma carcaça de aço inoxidável. Em seu interior há duas partes cerâmicas semelhantes a duas colméias e por onde os gases passam. Uma manta conformável funciona como *amortecedor* e preenche o espaço entre as partes cerâmicas e a cápsula de aço. Alguns veículos de teste já rodaram mais de 100.000 km com catalizadores deste tipo.

1 FORMULAÇÃO DO PROBLEMA.

1.1 Um Modelo de reação -difusão .

O conversor catalítico a ser estudado está fundamentado em reações químicas de oxidação do CO (monóxido de carbono), hidrocarbonetos nocivos e H_2. As reações de oxidação , que têm o alumínio-Platina como catalizador, são as seguintes

$$CO + 1/2O_2 \overset{Pt-al}{\longrightarrow} CO_2$$
$$C_3H_6 + 9/2O_2 \overset{Pt-al}{\longrightarrow} 3CO_2 + 3H_2O$$
$$H_2 + 1/2O_2 \overset{Pt-al}{\longrightarrow} H_2O$$

Propileno (C_3H_6) é a espécie de hidrocarboneto onde as reações são rápidas o suficiente para serem importantes na modelagem, razão pela qual é o único hidrocarboneto a ser considerado aqui. Se a intenção fosse considerar também hidrocarbonetos de oxidação lenta, incluiríamos o CH_4 (metano) e a correspondente equação de oxidação

$$CH_4 + 2O_2 \overset{Pt-al}{\longrightarrow} CO_2 + 2H_2O.$$

Sejam

C_i = concentração do composto i

T = temperatura

R_i = taxa de reação específica para o composto i

onde $i = 1, 2, 3, 4$ representam os compostos CO, C_3H_6, H_2, O_2, respectivamente.

As seguintes expressões para as taxas de reação foram obtidas experimentalmente por Voltz et al (1973).

$$R_1 = K_1C_1C_4/G \quad molCO/(cm^2 \cdot s) \tag{1.1}$$
$$R_2 = K_2C_2C_4/G \quad molC_3H_6/(cm^2 \cdot s) \tag{1.2}$$
$$R_3 = K_3C_3C_4/G \quad molH_2/(cm^2 \cdot s) \tag{1.3}$$

onde

$$G = T(1 + L_1C_1 + L_2C_2)^2(1 + L_3C_1^2C_2^2)(1 + L_4C_{NO}^{0.7}) \tag{1.4}$$

$$K_1 = 6.802 \times 10^{16} \exp(-13,108/T)$$
$$K_2 = 1.416 \times 10^{18} \exp(-15,109/T)$$
$$K_3 = 7.443 \times 10^{13} \exp(-19,552/T)$$
$$L_1 = 8.099 \times 10^{6} \exp(409/T)$$
$$L_2 = 2.579 \times 10^{8} \exp(-191/T)$$
$$L_3 = 1.13 \times 10^{21} \exp(9,299/T)$$
$$L_4 = 3.02 \times 10^{1} \exp(-3,733/T)$$

Observamos que NO atua como inibidor, não tomando parte da reação .

A maioria desses equipamentos é projetada para operar em condições balanceadas da razão ar/combustível para a conversão simultânea do monóxido de carbono, hidrocarbonetos e óxidos nitrogenados no escapamento do automóvel. Isto significa que vale a conservação de massa e energia, que nos fornecerá uma expressão para a taxa de reação do oxigênio (O_2).

$$\text{Balanço de Massa: } R_4 = 0.5R_1 + 4.5R_2 + 0.5R_3 \qquad (O_2)$$

No equacionamento a seguir, denotaremos por x a variável de distância ao longo de canais paralelos e t a variável temporal. Definimos ainda

$T_g(x,t)$ = temperatura da massa gasosa que flui pelo catalizador (K);

$T_s(x,t)$ = temperatura do sólido (paredes cerâmicas)(K);

$Cg_i(x,t)$ = concentração do composto i na corrente gasosa (mol/cm^3);

$Cs_i(x,t)$ = concentração do composto i nas paredes sólidas (mol/cm^3);

a = fator de catálise (J/mol);

$h(T_g)$ = coeficiente de difusão do calor $(\frac{J}{cm^2 \cdot s \cdot K})$;

$C(T_s)$ = calor específico do sólido (parede cerâmica)$(\frac{J}{cm^2 \cdot s})$;

$w(t)$ = taxa de fluxo de massa gasosa pelo catalizador $(\frac{J}{cm^2 \cdot s \cdot K})$;

Km_i=coeficiente de transferência de massa para o composto i $(\frac{J}{cm^2 \cdot s \cdot K})$.

Valores específicos destas constantes são encontrados em ([Cav 85]).

Ao ignorarmos a transferência de calor entre canais adjacentes do catalizador, o seguinte Sistema de Equações Diferenciais Parciais é estabelecido

$$C(T_s)\frac{\partial T_s}{\partial t} = \lambda \frac{\partial^2 T_s}{\partial x^2} + h(T_g)(T_g - T_s) + a(R_1 + R_2 + R_3) \tag{1.5}$$

$$w(t)\frac{\partial T_g}{\partial x} = h(T_g)(T_s - T_g) \tag{1.6}$$

$$-w(t)\frac{\partial Cg_1(x,t)}{\partial x} = Km_1(Cg_1 - Cs_1) \tag{1.7}$$

$$-w(t)\frac{\partial Cg_2(x,t)}{\partial x} = Km_2(Cg_2 - Cs_2) \tag{1.8}$$

$$-w(t)\frac{\partial Cg_3(x,t)}{\partial x} = Km_3(Cg_3 - Cs_3) \tag{1.9}$$

$$-w(t)\frac{\partial Cg_4(x,t)}{\partial x} = Km_4(Cg_4 - Cs_4) \tag{1.10}$$

para $0 < x < \infty, t > 0$, onde $x = 0$ é o ponto inicial do conversor. Assumimos que o conversor é muito longo, isto é, ele ocupa todo o intervalo $0 < x < \infty$.

A reação química entre os gases na superfície sólida e o volume gasoso em fluxo é expressa pelas equações algébricas

$$Km_i(T_g)(Cg_i - Cs_i) = aR_i(T_s, C_s) \qquad (i = 1, ...4) \tag{1.11}$$

onde os R_i são as taxas de reação definidas acima.

A função $w(t)$ é conhecida, bem como as condições de fronteira: para $t > 0$

$$T_s(0,t) = S(t) \tag{1.12}$$

$$[Cg_1, Cg_2, Cg_3, Cg_4](0,t) = (C^1, C^2, C^3, C^4)(t) \tag{1.13}$$

$$T_g(0,t) = T_0(t) \tag{1.14}$$

e a condição inicial: para $0 < x < \infty$

$$T_s(x,0) = T_s^0 \tag{1.15}$$

onde T_s^0 constante positiva.

1.2 O Problema de Controle Ótimo

O objetivo do conversor catalítico é diminuir a concentração dos vários tipos de poluente na saída do escapamento dos veículos. Entretanto, a prática nos revela que tal equipamento somente funciona bem quando as temperaturas do sistema motor-catalizador-escapamento são elevadas, e não há nada de espantoso nisso, pois sabemos que esta é uma característica dos processos químicos. Mas o que fazer então , uma vez que as situações em que realmente precisamos de um bom desempenho do catalizador são exatamente aquelas onde a temperatura de suas paredes são baixas ? Tal acontece, por exemplo, na partida a frio que damos em nosso automóvel no início do dia.

Tecnicamente, o problema principal resulta do fato que, sem estratégia de controle alguma, quando ligamos o automóvel pela primeira vez no dia, o equipamento está frio e a temperatura da parede cerâmica é de aproximadamente 300^oK (27^oC) (países tropicais), e nessa situação as taxas de reação são baixas e o processo é ineficiente. Dessa forma, além de haver queima imperfeita do combustível, que é característica da partida a frio de motores de combustão interna e que lança no ar compostos altamente tóxicos, nosso equipamento está frio e não consegue trabalhar bem dado seu princípio de funcionamento. Tal exemplo se aplica a outras situações , onde, da mesma forma, o catalizador não consegue trabalhar bem quando o motor trabalha mal (combustão imperfeita).

A estratégia de controle é simples, ao iniciarmos a partida a frio do automóvel, aquecemos mecanicamente a secção $x = 0$ do modelo catalizador e então podemos controlar as taxas de reação em toda a extensão das paredes graças à propagação do calor pela estrutura cerâmica. Uma vez que podemos controlar tais taxas, podemos, ao menos teoricamente, minimizar a concentração de poluente mesmo enquanto a temperatura do motor ainda é baixa, melhorando o desempenho do catalizador. Quanto maior a temperatura das paredes cerâmicas, que vão aquecendo progressivamente, menor nossa ação controladora, e assim esperamos que no instante final t_0 do intervalo de aquecimento, nossa interferência cesse.

Matematicamente, descrevemos nossa ação controladora por

$$T_s(0,t) = 300 + S(t) \tag{1.16}$$

$$S(t) = \text{função de controle}$$

O objetivo é reduzir a concentração Cg_i em $x = L$. Uma vez que cada espécie de poluente conta diferentemente em um teste normativo para controle de poluição , queremos minimizar um expressão da forma

$$J(S) = \sum_{j=1}^{3} \lambda_j \int_0^{t_0} Cg_j(L,t)dt \qquad , \lambda_1 + \lambda_2 + \lambda_3 = 1 \tag{1.17}$$

onde os λ_j são pesos especificados para cada poluente de concentração Cg_j.

Para cada escolha de função de controle $S(t)$, devemos resolver o sistema de equações (1.5)-(1.10) e então computar a expressão em (1.17), que chamamos de funcional de concentrações (*funcional de custo* no contexto da otimização).

Queremos então , dada a função contínua $w(t)$ no intervalo de aquecimento $(0, t_0)$, que caracteriza o funcionamento do motor (lembramos que $w(t)$ é a taxa de fluxo de massa através do sistema), encontrar a função $S^*(t)$ que minimiza a concentração de poluente na saída do equipamento e que depende não localmente de $w(t)$. Não existe qualquer tipo de retro-alimentação (*feedback*) em nossa estratégia, como se poderia, perfeita e licitamente, pensar. O que nós queremos é apenas uma calibragem ótima $S^*(t)$ que minimize (1.17) para cada $w(t)$, ou seja, para cada regime de funcionamento específico do motor do automóvel.

Vamos introduzir o conjunto $\mathcal{A}$ de funções de controle admissíveis.

$$\mathcal{A} = \{S(t) \in C^1((0, t_0)), 0 \leq S(t) \leq N, \int_0^{t_o} S(t)dt \leq M\} \tag{1.18}$$

As constantes N e M representam restrições mecânicas impostas aos controles $S(t)$, e são dados do problema; em termos práticos, traduzem limitações de natureza material, pois obviamente temperaturas arbitrariamente elevadas não são suportadas, e limitações de caráter energético, dado que nossa quota de energia para executar tal tarefa é limitada e deverá ser reposta, de alguma forma, até a próxima ignição a frio do automóvel, por exemplo.

Assim, estamos interessados em resolver o seguinte problema de Controle Ótimo, proposto em [Fri 94],

$$\text{encontrar } S^* \text{ tal que } J(S^*) = \min_{S \in \mathcal{A}} J(S).$$

1.3 Um modelo simplificado

A solução do problema de controle para o modelo (1.5)-(1.10) demanda um esforço computacional muito grande. Nosso objetivo aqui é considerar uma versão mais simples de (1.5)-(1.10) mas que ainda mantenha os aspectos teóricos e computacionais importantes desse modelo.

Hipóteses:

1. $h(T_g) = 0$: negligenciamos a difusão da temperatura da massa gasosa, uma vez que o fluxo através do catalizador é muito grande e então os efeitos difusivos podem ser desprezados.

2. temos apenas um elemento poluente com concentração $c = c_s$ no sólido e $c = c_g$ no volume gasoso.

3. as taxas de reação são nulas em $T_s = 300^o K$: $R(T_s, c_s) = (T_s - 300)c_s$

4. $\lambda = 1, a = 1, C(T_s) \equiv C, K_{m_i} \equiv 1$

Para fins de re-equacionamento, escrevemos $T = T_s - 300, u = c_g$, e então temos

$$C\frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial x^2} + Tc_s$$

$$-w(t)\frac{\partial u}{\partial x} = u - c_s$$

$$u - c_s = Tc_s \Leftrightarrow K_{m_i}(T_g)(c_g - c_s) = a(T_s - 300)c_s \Rightarrow c_s = \frac{u}{1+T}$$

Segue o sistema de E.D.P. não-lineares

$$C\frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial x^2} + \frac{T}{1+T}u, \qquad 0 < x < 2L, 0 < t \leq t_0 \tag{1.19}$$

$$-w(t)\frac{\partial u}{\partial x} = \frac{T}{1+T}u \tag{1.20}$$

para $0 < x < 2L, 0 < t \leq t_0$, e que têm acopladas as condições de fronteira:

$$T(x,0) = 0 \qquad\qquad u(0,t) = u_0 > 0 \tag{1.21}$$

$$T(0,t) = S(t) \qquad (termo\ de\ controle) \tag{1.22}$$

$$T(x,t) \to 0\ ao\ x \to \infty \qquad\qquad 0 \leq t \leq t_0 \tag{1.23}$$

Nosso problema então é encontrar o controle $S^*(t)$ que minimize o funcional

$$J(S) = \int_0^{t_0} u(L,t)dt$$

onde $S(t)$ e $u(x,t)$ se relacionam pelas equações (1.19)-(1.22) definidas agora no domínio limitado $M = (0,2L) \times (0,t_0)$ e onde (1.23) é substituída pela equação

$$T(2L,t) = 0 \quad \forall t \in [0,t_0]; \tag{1.24}$$

tais ajustes em relação a (1.19)-(1.23) são necessários já pensando-se na algoritmização de um procedimento de solução de tal problema de fronteira.

De (1.20)-(1.21) obtemos que

$$u(x,t) = u_0 \exp\left[-\int_0^x \frac{1}{w(t)} \cdot \frac{T}{1+T}(y,t)dy\right]$$

e lembramos que $w(t) > 0 \ \ \forall t \in [0,t_0]$.

Daí, evidencia-se que o termo

$$\frac{T}{1+T}u = f(T) = f(x,t)$$

em (1.19) é realmente um operador em T. Assim, podemos considerar, para propósitos de análise, a equação (1.19) como sendo da forma de uma equação de difusão não linear

$$\frac{\partial T}{\partial t} = \frac{1}{C}\frac{\partial^2 T}{\partial x^2} + f(T). \tag{1.25}$$

2 ANÁLISE DO MODELO.

No capítulo anterior, foi indicado que a parte evolutiva do modelo simplificado desenvolvido na seção (1.3) pode ser considerada como uma equação do tipo parabólico não linear

$$T_t = \frac{1}{C} \triangle T + f(T)$$

onde $\triangle$ denota o Laplaciano n-dimensional e onde $C > 0$ é a constante de difusão do calor, característica do problema. Este tipo de equação tem conquistado atenção matemática devido a suas variadas e abrangentes aplicações em engenharia, química e física, tanto no caso transiente quanto no caso estacionário

$$\frac{1}{C} \triangle T + f(T) = 0.$$

Maiores referências podem ser encontradas em [Smo 83].

Nosso primeiro propósito será um estudo sobre a existência e a unicidade de soluções para o problema inicial de fronteira já definido na seção (1.3) para cada controle $S(t) \in C((0, t_0))$ e para $0 < x < 2L, 0 < t \leq t_0$

$$C\frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial x^2} + \frac{T}{1+T}u \tag{2.1}$$

$$-w(t)\frac{\partial u}{\partial x} = \frac{T}{1+T}u \tag{2.2}$$

$$T(x,0) = 0, \qquad u(0,t) = u_o > 0, \qquad T(0,t) = S(t) \tag{2.3}$$

$$T(2L,t) = 0 \ \ \forall t \in (0,t_0). \tag{2.4}$$

Mostraremos na seção (2.1) que existe uma equivalência entre o problema recém citado (quando desacoplado das concentrações u) e o problema definido pela equação integral

$$T(x,t) = \int_0^t \int_0^{2L} G(x, t-\tau, y) f(T(y,\tau)) dy d\tau - \frac{1}{C}\int_0^t G_y(x, t-\tau, 0) S(\tau) d\tau \tag{2.5}$$

onde G é um núcleo de Green definido sobre $M = (0, 2L) \times (0, t_0)$, para funções $f(T)$ e $S(t)$ continuamente diferenciáveis.

Nosso objetivo aqui é estabelecer existência e unicidade para (2.1)-(2.4). Muito embora a metodologia clássica nos conduza à questões de existência, estimativas e regularidade de soluções de problemas iniciais de valores na fronteira, no contexto dos espaços funcionais de Sobolev, e esteja amparada em alguma variante do Princípio do Máximo para equações diferenciais parabólicas para assegurar a unicidade, nossa estratégia será estabelecida num contexto e numa metodologia alternativa. Nosso alvo passa a ser então o estudo da formulação integral e sua análise através da *Teoria das Equações Integrais de Hammerstein* e no contexto dos espaços funcionais ordenados.

Nosso segundo propósito será apresentar alguns resultados e lemas concernentes à questão do problema de otimização estabelecido na seção (1.3), e no contexto da Teoria dos Espaços de Banach Convexos.

2.1 Uma representação integral desacoplada.

Considere o problema de contorno

$$\frac{\partial T}{\partial t} = \frac{1}{C}\frac{\partial^2 T}{\partial x^2} + f(x,t) \qquad 0 < x < 2L, t > 0 \tag{2.6}$$

$$T(0,t) = h(t), \qquad t > 0 \tag{2.7}$$

$$T(x,0) = T_0(x), \qquad 0 < x < 2L \tag{2.8}$$

$$T(2L,t) = 0 \tag{2.9}$$

o qual possui uma função de Green (ver [But 82])

$$G_0(x,t,y) = \frac{1}{L}\sum_{n=1}^{\infty} \sin\frac{n\pi x}{2L}\sin\frac{n\pi y}{2L}\exp\left[-\frac{(n\pi)^2}{4L^2}\frac{1}{C}\cdot t\right] \tag{2.10}$$

onde $C > 0, f(x,t), h(t), T_o(x)$ são funções continuamente diferenciáveis.

A seguir, estabeleceremos uma representação integral para a solução do problema de contorno (2.6)-(2.9), e então derivaremos uma expressão integral para a solução das equações (2.1)-(2.4).

2.1.1 Lema (representação integral).

Para funções $f(x,t), h(t), T_0(x)$ suficientemente regulares, a solução das equações (2.6)-(2.9) tem a forma

$$T(x,t) = \int_0^t \int_0^{2L} G(x,t-\tau;t) f(y,\tau)\,dy\,d\tau - \frac{1}{C}\int_0^t G_y(x,t-\tau;0)h(\tau)\,d\tau + \int_0^{2L} G(x,t;y)T_0(y)\,dy \tag{2.11}$$

Prova: Definimos, para $\overline{t} \in (0,t_0)$

$$w(x,\epsilon) = \int_0^{2L} G(x,\epsilon,y)T(y,\overline{t})\,dy.$$

Pela definição de núcleo, temos

$$T(x,\overline{t}) = \lim_{\epsilon\to 0} w(x,\epsilon) \tag{2.12}$$

e então observamos que $G_t(x,t,y) - G_{xx}(x,t,y) = 0 \quad \forall x,y \in (0,2L), t \in\ (0,t_0)$ implica que

$$w_\epsilon(x,\epsilon) - w_{xx}(x,\epsilon) = 0 \qquad \forall x, \epsilon > 0$$

$$w(x,0) = T(x,\overline{t})$$

Fixamos $x, \overline{t}, \epsilon > 0, \Omega = (0,2L)\times(0,\overline{t})$.

Definimos

$$v(y,t) = G(x,\overline{t}+\epsilon-t,y), 0 \le y \le 2L, 0 \le t \le \overline{t}$$

Temos então

$$\int_\Omega v(y,t)f(y,t)dydt = \int_\Omega v(y,t)\left(T_t(y,t) - \frac{1}{C}T_{yy}(y,t)\right)dydt =$$
$$-\int_0^{\overline{t}}\int_0^{2L} T\left(v_t + \frac{1}{C}v_{yy}\right)dydt + \int_0^{2L} T(y,\overline{t})v(y,\overline{t})dy - \int_0^{2L} T(y,0)v(y,0)dy -$$
$$\frac{1}{C}\int_0^{\overline{t}} \left[v(0,t)T_y(0,t) - T(0,t)v_y(0,t) + v(2L,t)T_y(2L,t) - T(2L,t)v_y(2L,t)\right]dt =$$
$$-0 + \int_0^{2L} G(x,\epsilon,y)T(y,\overline{t})dy - \int_0^{2L} G(x,\overline{t}+\epsilon,y)T_0(y)dy -$$
$$\frac{1}{C}\int_0^{\overline{t}} \left[G(x,\overline{t}-t+\epsilon,0)T_y(0,t) - G_y(x,\overline{t}-t+\epsilon,0)h(t)\right]dt + 0 - 0$$

e desta forma podemos escrever

$$\int_0^{2L} G(x,\epsilon,y)T(y,\overline{t})dy = \int_0^{\overline{t}}\int_0^{2L} G(x,\overline{t}+\epsilon-t,y)f(y,t)dydt +$$
$$\int_0^{2L} G(x,\overline{t}+\epsilon,y)T_0(y)dy - \frac{1}{C}\int_0^{\overline{t}} G_y(x,\overline{t}+\epsilon-t,0)h(t)dt$$

e finalmente, fazendo $\epsilon \to 0$ e usando (2.12) temos

$$T(x,\overline{t}) = \int_0^{\overline{t}}\int_0^{2L} G(x,\overline{t}-t,y)f(y,t)dydt + \int_0^{2L} G(x,\overline{t},y)T_0(y)dy - \frac{1}{C}\int_0^{\overline{t}} G_y(x,\overline{t}-t,0)h(t)dt$$

e então a solução $T(x,t)$ de (2.6)-(2.9) é dada por (2.11) , o que demonstra o Lema.

$\diamond$

Como, por (2.2), obtemos a expresão

$$u(x,t) = u_0 \exp\left(-\int_0^x \frac{1}{w(t)}\frac{T}{1+T}(y,t)dy\right) \tag{2.13}$$

e podemos escrever então $\dfrac{1}{C}\dfrac{Tu}{1+T} = f(T)$, onde f é um funcional de T, devido à sua dependência não-local.

O Lema acima justifica então utilizarmos a representação integral implícita

$$T(x,t) = \int_0^t\int_0^{2L} G_0(x,t-\tau;y)f_0(T)(y,\tau)dyd\tau - k(x,t) \tag{2.14}$$

onde $k(x,t) = \dfrac{1}{C}\displaystyle\int_0^t G_y(x,t-\tau;0)S(\tau)d\tau$, para a solução do modelo (2.1)-(2.4), e onde o termo

$$f_0(T(x,t)) = \frac{u_0}{C}\frac{T}{1+T}(x,t)\exp\left[-\int_0^x \frac{T}{1+T}(s,t)ds\right] \tag{2.15}$$

é obtido substituindo-se (2.13) em (2.1).

2.2 Uma aplicação da Teoria das Equações de Hammerstein.

Nesta seção , mostraremos como a Teoria das Equações de Hammerstein, desenvolvida no contexto dos Espaços de Banach Ordenados, pode ser aplicada com vistas a fornecer resultados que assegurem a existência e a unicidade da solução de nosso problema integral (2.14) para todo controle $S(t) \in C^1((0,t_0))$, e para $f_0(T)$ definido por (2.15), G_0 definido em (2.10).

Podemos reduzir a equação (2.14) a forma de Hammerstein

$$V(x,t) = \int_0^t \int_0^{2L} G(x,t-\tau;y)h(V(y,\tau))dyd\tau, V \in X = C(\overline{M}), \tag{2.16}$$

onde $V(x,t) = T(x,t) + k(x,t), h(V) = f_0(V-k)$. Além disso, a redução acima preserva como estrutura da aplicação $h : X \longrightarrow X$ algumas hipóteses fundamentais, feitas originalmente sobre $f(T)$, que vão garantir a unicidade de soluções de (2.16), como monotonicidade estrita e sublinearidade estrita, e uma vez que tais características não se perdem por translações .

De maneira abstrata, temos uma equação do tipo

$$V = KH(V) \tag{2.17}$$

onde

$$(Kv)(x) = \int_M G(x,y)v(y)dy \tag{2.18}$$

$$H(v)(x) = h(x,v(x)) \tag{2.19}$$

algumas vezes chamado *operador de Nemyckii*.

Primeiro, introduziremos uma relação de ordem no espaço funcional de Banach $X = C(\overline{M})$. Para tal, definimos um cone de ordem X_+ e então diremos

$$U \leq V \text{ se } V - U \in X_+, \quad U, V \in X.$$
$$U \ll V \text{ se } V - U \in int(X_+), \quad U, V \in X.$$

Podemos então definir sub e supersoluções U_0 e V_0 por

$$U_0 \leq KH(U_0)$$
$$V_0 \geq KH(V_0)$$

e cuja existência é garantida por meio do Lema (A-1.1.17) em apêndice.

Uma necessidade que surge é garantir a validade as seguintes hipóteses:

(N) X_+, o cone de ordem sobre X, é normal;

(P) $K : D(K) \to X$ é um operador fortemente positivo.

A primeira hipótese, que traduz uma compatibilidade entre a definição de cone de ordem e a norma introduzida sobre X e significa que existe um número $c > 0$ tal que

$$0 \leq V \leq U \Rightarrow \|V\| \leq c\|U\| \quad \forall U, V \in X,$$

implicará (proposição (A-1.1.4)) que todo intervalo de ordem $[U,V]$ é limitado.

A segunda hipótese, que significa que (ver definição (A-1.1.3))

$$V \geq 0 \Rightarrow K(V) \gg 0, \quad V \in D(K),$$

não é natural no contexto de problemas de fronteira, uma vez que a imagem do operador K não contém um ponto interior ao cone de ordem natural $C_+(\overline{M})$, mas sim um ponto na sua

fronteira, uma vez que o núcleo $G_0(x,t,y)$ definido em (2.10) é positivo dado que tal característica simplesmente traduz a validade do Princípio do Máximo sobre M para problemas parabólicos. [1] Para validarmos ambas hipóteses acima, é um procedimento clássico definir o espaço

$$X_e = \{x \in X : \text{existe real } a > 0 \text{ tal que } -ae \leq x \leq ae\},$$

com norma $\|x\|_e = inf\{a > 0 : -ae \leq x \leq ae\}$ e cone de ordem $X_+^e = X_+ \bigcap X_e$, onde $X_+ = C_+(M)$, e $e > 0$ é a solução de (2.14) para $f \equiv 1$. Definimos então $K : D(K) \to X_e$, e neste novo contexto K é fortemente positivo.

Segundo, dados U_0, V_0 sub e supersoluções , respectivamente,e sob a hipótese de H monótono crescente ($U \leq V \Rightarrow H(U) \leq H(V)$) , definimos uma sequência de subsoluções U_n e uma sequência de supersoluções V_n:

$$U_{n+1} = KH(U_n), \qquad V_{n+1} = KH(V_n) \tag{2.20}$$

que convergirão , pelo teorema (A-1.1.14) á menor solução U e á maior solução V de (2.16), respectivamente, como sequências monótonas e limitadas sup ou inferiormente.

O terceiro passo, a unicidade dos limites U_n e V_n, será assegurada sob as hipóteses de monotonicidade crescente estrita do operador H (assegurada pelo Lema (A-1.1.19) para $L \leq 1/2$), positividade forte do operador compacto linear K e sublinearidade estrita do operador H (Lema (A-1.1.18)): $\forall \lambda \in (0,1)$

$$H(\lambda V) > \lambda H(V)$$

(ver definição (A-1.1.11)). Assim, a unicidade da solução da equação (2.16), e consequentemente da equação (2.14), é assegurada para $L \leq 1/2$.

2.3 Sobre a existência e a unicidade do controle ótimo.

2.3.1 Um problema de controle simplificado e sua solução .

Pode ser encontrada em [Fri 94] a derivação da solução ótima $S^*(t)$ para um caso simplificado com hipóteses bastante fortes e num espaço mais amplo

$$\mathcal{A} = \{S(t), 0 \leq t \leq t_o; S(t) \textit{ é continua por trechos e } 0 \leq S(t) \leq N, \int_0^{t_0} S(t)dt \leq M\}$$

que agora descreveremos. Trocaremos (2.1) pela equação do calor abaixo (onde $C = 1, L = \infty$ por simplicidade)

$$\frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial x^2}, \qquad 0 < x < \infty, t > 0 \tag{2.21}$$

[1] Não apresentaremos neste trabalho uma prova do Princípio do Máximo para equações parabólicas, uma vez que tal noção está intrínseca nas hipóteses e resultados de positividade do apêndice . Referências podem ser encontradas em [Zei1 85], pag 337, problemas 7.2d e 7.2g .

A hipótese assumida é que o termo $Tu(y,t)$ é muito pequeno se comparado com $\frac{\partial T}{\partial t}$ ou $\frac{\partial^2 T}{\partial x^2}$. Também assumimos que $T(x,t) \to 0$ ao $x \to \infty$. Então, pela representação integral da solução da equação do calor não -homogênea (2.11)

$$T(x,t) = \int_0^t \frac{x}{\sqrt{4\pi}(t-\tau)^{3/2}} exp\left(-\frac{x^2}{4(t-\tau)}\right) S(\tau) d\tau \tag{2.22}$$

A seguir, simplificamos a expressão do funcional

$$J(S) = \int_0^{t_0} exp\left(-\int_0^L \frac{T(y,t)}{1+T(y,t)} dy\right) dt \tag{2.23}$$

observando que podemos escrever

$$J(S) \simeq u_0 \int_0^{t_0} [1 - T(y,t)dy]\, dt$$

e que para chegarmos a tal aproximação fizemos a hipótese forte que $T(x,t) \ll 1$ e então

$$A = \frac{T(x,t)}{1+T(x,t)}$$

é pequeno tal que $\exp(-A) \simeq 1 - A$.

Definindo o funcional

$$J_0(S) = u_0 \int_0^{t_0} \int_0^{\infty} T(y,t) dy dt \tag{2.24}$$

nosso problema então é equivalente a

- encontrar S^* tal que $J_0(S^*) = \max_{S \in \mathcal{A}} J_0(S)$.

Assumimos que $Nt_0 > M$, isto significa que não há energia M suficiente para manter a temperatura $T(0,t)$ em seu valor máximo admissível N durante todo o intervalo $[0, t_0]$. Assim, devemos gastar a energia disponível sabiamente.

2.3.2 Teorema (Solução Ótima)

Existe uma única solução ótima $S_0(t)$ que minimiza (2.24) e ela é dada por:

$$S_0(t) = \begin{cases} N, & \text{se } 0 \le t \le \tilde{t} \\ 0, & \text{se } \tilde{t} < t \le t_0 \end{cases} \tag{2.25}$$

onde $N\tilde{t} = M$.
Prova:

Se substituirmos (2.22) em (2.24) obteremos

$$J_0(S) = u_0 \int_0^{t_0} dt \int_0^t S(\tau) d\tau \int_0^{\infty} \frac{x}{\sqrt{4\pi}(t-\tau)^{3/2}} exp\left(-\frac{x^2}{4(t-\tau)}\right) dx =$$

$$= \frac{2u_0}{\sqrt{4\pi}} \int_0^{t_0} dt \int_0^t S(\tau) \left[\frac{exp\left(-\frac{x^2}{4(t-\tau)}\right)}{\sqrt{t-\tau}} \right]_{x=0}^{x=\infty} d\tau$$

$$= \frac{2u_0}{\sqrt{4\pi}} \int_0^{t_0} S(\tau) d\tau \int_\tau^{t_0} \frac{dt}{\sqrt{t-\tau}} = \frac{4u_0}{\sqrt{4\pi}} \int_0^{t_0} \sqrt{t_0-\tau} S(\tau) d\tau$$

Uma vez que a função $\sqrt{t_0 - \tau}$ é estritamente monótona decrescente em τ, a última integral é maximizada se e somente se $S(t)$ satisfaz (2.25).

2.3.3 Existência e unicidade no caso geral.

Colocaremos nesta subseção algumas questões relacionadas à existência e unicidade de nosso objetivo neste trabalho: o controle ótimo $S^*(t)$ que minimiza

$$J(S) = \int_0^{t_0} u(L,t) dt \tag{2.26}$$

onde $S(t)$ e $u(x,t)$ se relacionam pelo sistema em $(0,2L) \times (0,t_0)$

$$C\frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial x^2} + \frac{T}{1+T} u \tag{2.27}$$

$$-w(t)\frac{\partial u}{\partial x} = \frac{T}{1+T} u \tag{2.28}$$

$$T(x,0) = 0 \qquad u(0,t) = u_0 > 0, \qquad T(0,t) = S(t) \tag{2.29}$$

$$T(2L,t) = 0, 0 < t \leq t_0 \tag{2.30}$$

onde $w(t) \geq r > 0 \quad \forall t \in (0,t_0)$ e sobre o espaço de controles admissíveis

$$\mathcal{A} = \{S(t) \in C^1((0,t_0), 0 \leq S(t) \leq N, \int_0^{t_0} S(t) dt \leq M\} \tag{2.31}$$

já mencionado nos capítulos anteriores.

Sob a hipótese de $\mathcal{A}$ *fracamente sequencialmente compacto*, podemos assegurar que o funcional $J(S)$ possui máximo e mínimo sobre $\mathcal{A}$. Maiores detalhes, inclusive a definição de conjuntos fracamente compactos, podem ser encontrados em [Zei3 85], pág 152.

Conclusões acerca da unicidade do controle ótimo demandariam hipóteses clássicas de convexidade estrita do funcional $J(S)$, as quais não podem ser asseguradas dada a complexidade da aplicação $u : S \mapsto u(S)$. A unicidade do controle ótimo $S^*(t) \in \mathcal{A}$, portanto e até o presente momento, não pode ser assegurada.

3 AS TÉCNICAS COMPUTACIONAIS.

O problema de controle ótimo definido em (1.3), por ser um problema de otimização sujeita à restrições em um espaço infinito-dimensional, o espaço $\mathcal{A}$ então definido, precisará necessariamente ser reformulado a fim de ser resolvido algoritmicamente por meio de procedimentos computacionais de busca ao controle ótimo $S^*(t) \in \mathcal{A}$.

Dessa forma, faz-se necessário aproximar o espaço $\mathcal{A}$ dos controladores admissíveis de modo a termos um problema computacionalmente resolvível e cuja solução é $\tilde{S^*}(t)$. Para tal, seguindo um procedimento clássico em matemática aplicada, definimos os espaços de interpolação np-segmentada $\tilde{\mathcal{A}}_{np}$, $2(np+1)$ dimensionais, densos no espaço $\mathcal{A}$, ou seja, toda função $S(t) \in \mathcal{A}$ pode ser arbitrariamente aproximada por uma sequência de funções $\tilde{S}(t) \in \bigcup_{np} \tilde{\mathcal{A}}_{np}$. Desta forma, a solução $S^*(t)$ pode ser satisfatoriamente aproximada pela solução $\tilde{S}^*(t) \in \tilde{\mathcal{A}}_{np}$ para np suficientemente grande, e neste sentido justificamos ver em $\tilde{S}^*(t)$ uma solução para o problema definido em (1.3).

A seção (3.1) descreve a ponte que conduz e norteia a estratégia acima, e assim possibilita a reformulação de (1.3), apresentada na seção (3.2). A seção (3.3) descreve a técnica de discretização empregada na solução computacional a que nos propomos. As seções (3.4)-(3.5) descrevem algumas tarefas cuja implementação se faz necessária. Os espaços $\tilde{\mathcal{A}}_{np}$ serão definidos como o universo de interpoladoras cúbicas segmentadas definidas no intervalo $[0, t_0]$ e que pertencem ao espaço $\mathcal{A}$ de controles admissíveis, e definiremos como $\mathcal{A}^c_{np}$ (às vezes referido como $\mathcal{A}^c$ simplesmente) o espaço de coeficientes nodais e tangenciais associado a $\tilde{\mathcal{A}}_{np}$. Desta forma, cada função $S(t) \in \mathcal{A}$ é aproximada por uma cúbica np-segmentada $\tilde{S}(t) \in \tilde{\mathcal{A}}_{np}$, à qual associamos bi-univocamente um vetor de coeficientes nodais e tangenciais $c_s \in \mathcal{A}^c_{np} \subseteq \Re^{2(np+1)}$. A proposta deste capítulo é descrever um pouco detalhadamente as principais técnicas e métodos numéricos envolvidos na algoritmização de nossa tarefa.

Uma notação que será frequente nas seções seguintes, e que agora será estabelecida, é que o símbolo # indica algum tipo de aproximação numérica, por exemplo,

$$I_S = \frac{[f(0) + 4f(1/2) + f(1)]}{6}.1 = \# \int_0^1 f(x)dx \qquad \text{(Simpson)}$$

representa uma aproximação numérica, obtida pela quadratura de Simpson, para o valor exato da integral I_S.

3.1 Um Operador de Interpolação Cúbica em $C^1([0, L])$.

Vamos definir um operador em $C^1([0, L])$.

a) Interpoladora cúbica de 2 nós.

Seja $f(x)$ uma função contínua, com derivada primeira contínua no intervalo real $[a, b]$, ou seja $f(x) \in C^1([a, b])$.

Sejam dados os parâmetros

$$\begin{aligned} y_a &= f(a), & y_b &= f(b) \\ t_a &= \tfrac{df}{dx}(a) = f'(a), & y_b &= \tfrac{df}{dx}(b) = f'(b). \end{aligned}$$

Definimos a interpoladora cúbica no intervalo $[a,b]$ e sobre o espaço de Banach $C^1([0,L])$

$$\Phi_{[a,b]}[f](x) = y_a \frac{(x-b)(2a-b-x)-2(x-a)^2}{(b-a)^2} + 3y_b \frac{(x-a)^2}{(b-a)^2} - t_a \frac{(x-a)(2x-a-b)}{b-a} \tag{3.1}$$

$$-t_b \frac{(x-a)^2}{b-a} + 2\left[\frac{t_b+t_a}{2} - \frac{y_b-y_a}{b-a}\right]\frac{(x-a)^3}{(b-a)^2}$$

que verifica

$$\Phi_{[a,b]}[f](a) = y_a = f(a), \qquad \Phi_{[a,b]}[f](b) = y_b = f(b) \tag{3.2}$$

$$\frac{d\Phi_{[a,b]}}{dx}[f](a) = t_a = f'(a) \tag{3.3}$$

$$\frac{d\Phi_{[a,b]}}{dx}[f](b) = t_b = f'(b) \tag{3.4}$$

onde

$$\frac{d\Phi_{[a,b]}}{dx}[f](x) = \frac{6(y_b-y_a)(x-a)}{(b-a)^2} - \frac{t_a(4x-3a-b)+2t_b(x-a)}{h} + 6\left[\frac{t_b+t_a}{2} - \frac{y_b-y_a}{b-a}\right]\frac{(x-a)^3}{(b-a)^2} \tag{3.5}$$

$$\frac{d^2\Phi_{[a,b]}}{dx^2}[f](x) = \frac{6(y_b-y_a)}{(b-a)^2} - \frac{4t_a+2t_b}{b-a} + 12\left[\frac{t_b+t_a}{2} - \frac{y_b-y_a}{b-a}\right]\frac{(x-a)}{(b-a)^2} \tag{3.6}$$

Seja $\Psi_{[a,b]}[f] = \int_a^b \Phi_{[a,b]}(x)dx$ a integral de Φ no intervalo $[a,b]$.

$$\Psi_{[a,b]}[f] = -\frac{y_a}{h^2}\int_a^b (x-b)(x-a+h)dx +$$

$$\left[\frac{-2y_a(x-a)^3}{3h^2} + \frac{y_b(x-a)^3}{h^2} - \frac{t_b(x-a)^3}{3h} + \frac{(t_b+t_a)(x-a)^4}{4h^2} - \frac{2(y_b-y_a)(x-a)^4}{4h^3}\right]_a^b -$$

$$\frac{t_a}{h}\int_a^b (x-a)(2x-a-b)dx = \frac{2y_a h}{3} - \frac{2y_a h}{3} + hy_b - \frac{h^2 t_b}{3} + \frac{h^2(t_b+t_a)}{4} - \frac{h(y_b-y_a)}{2} - \frac{h^2 t_a}{6}$$

e então segue

$$\Psi_{[a,b]}[f] = \frac{y_a+y_b}{2}(b-a) + \frac{t_a-t_b}{12}(b-a)^2 \tag{3.7}$$

b) Interpoladora Cúbica nos (n+1) nós $\{x_0, x_1, ..., x_n\}$

Seja a partição regular, de espaçamento h, do intervalo $[0,L]$
$\{x_0 = 0, x_1, x_2, ..., x_{n-1}, x_n = L\}$.

Sendo $y_i = f(x_i), d_i = f'(x_i)$, definimos

$$\Phi^n_{[0,L]}[f](x) = y_{i-1}\frac{(x-x_i)(2x_{i-1}-x_i-x)-2(x-x_{i-1})^2}{h^2} + 3y_i\frac{(x-x_{i-1})^2}{h^2} \tag{3.8}$$

$$-d_{i-1}\frac{(x-x_{i-1})(2x-x_{i-1}-x_i)}{h} - d_i\frac{(x-x_{i-1})^2}{h} + 2(\frac{d_i+d_{i-1}}{2} - \frac{y_i-y_{i-1}}{h})\frac{(x-x_{i-1})^3}{h^2}$$

para algum $i, 1 \le i \le n$, tal que $x_{i-1} \le x < x_i$.

A interpoladora está determinada pelos $2(n+1)$ valores nodais
$\{y_0, y_1, ..., y_n, d_0, d_1, ..., d_n\}$.

Desta forma está estabelecida a correspondência

$$[y_0, y_1, ..., y_n, d_0, d_1, ..., d_n] \longleftrightarrow \text{Interpoladora Cúbica para } f(x) \text{ no } [0, L].$$

Mais precisamente, dizemos Interpoladora Cúbica Segmentada no $[0, L]$, uma vez que está definida em cada um dos segmentos $[x_{i-1}, x_i]$, portanto não tendo uma definição global em todo o intervalo $[0, L]$.

c) A integral nos $(n+1)$ nós $\{x_0, x_1, ..., x_n\}$.

Seja $\Psi_L = \int_0^L \Phi_{[0,L]}[f](x)dx$. Usando a expressão (3.7) obtemos

$$\Psi = \left[\frac{y_0}{2} + y_1 + y_2 + ... + y_{n-1} + \frac{y_n}{2}\right] h + [d_0 - d_n]\frac{h^2}{12} \tag{3.9}$$

d) A dependência paramétrica de máximos locais da interpoladora.

Dados os parâmetros de interpolação $a, b, y_a, y_b, t_a, t_b, h = b - a$, sabemos que a interpoladora $\phi(x)$ definida em (3.1) exibe um máximo local $\overline{x}$ no intervalo (a, b).

Queremos determinar

$$dFy_a = \frac{d\Phi}{dy_a}(\overline{x}), dFy_b = \frac{d\Phi}{dy_b}(\overline{x}), dFt_a = \frac{d\Phi}{dt_a}(\overline{x}), dFt_b = \frac{d\Phi}{dt_b}(\overline{x})$$

a fim de computar os normais de superfície, conforme será explicado na subseção (4.3.3).

Equacionando $\frac{d\Phi}{dx}(\overline{x}) = 0$, e usando a expressão (3.5) temos

$$6h(y_b - y_a)(\overline{x} - a) - h^2 t_a(4\overline{x} - 3a - b) - 2h^2 t_b(\overline{x} - a) + 3h(t_b + t_a)(\overline{x} - a)^2 - 6(y_b - y_a)(\overline{x} - a)^2 = 0$$

e então

$$[3h(t_b + t_a) - 6(y_b - y_a)]\overline{x}^2 + [6h(y_b - y_a) - 4h^2 t_a - 2h^2 t_b - 6ha(t_b + t_a) + 12a(y_b - y_a)]\overline{x} + [(3a + b)h^2 t_a + 2ah^2 t_b - 6ha(y_b - y_a) + 3ha^2(t_b + t_a) - 6a^2(y_b - y_a)] = 0$$

definimos

$$P = 3h(t_b + t_a) - 6(y_b - y_a)$$

$$Q = 6h(y_b - y_a) - 2h^2(2t_a + t_b) - 6ha(t_b + t_a) + 12a(y_b - y_a)$$

$$R = (3a + b)h^2 t_a + 2ah^2 t_b - 6ha(y_b - y_a) + 3ha^2(t_b + t_a) - 6a^2(y_b - y_a)$$

e temos então

$$P\overline{x}^2 + Q\overline{x} + R = 0 \Leftrightarrow \overline{x} = \frac{-Q \pm \sqrt{Q^2 - 4PR}}{2P}. \tag{3.10}$$

Prosseguimos, derivando as equações que definem P, Q, R, respectivamente,

$$\frac{dP}{dy_a} = 6 \qquad \frac{dP}{dy_b} = -6 \qquad \frac{dP}{dt_a} = 3h \qquad \frac{dP}{dt_b} = 3h$$
$$\frac{dQ}{dy_a} = -(6h+12a) \qquad \frac{dQ}{dy_b} = 6h+12a \qquad \frac{dQ}{dt_a} = -6ha-4h^2 \qquad \frac{dQ}{dt_b} = -6ha-2h^2$$
$$\frac{dR}{dy_a} = 6ha+6a^2 \qquad \frac{dR}{dy_b} = -6ha-6a^2 \qquad \frac{dR}{dt_a} = (3a+b)h^2+3ha^2 \qquad \frac{dR}{dt_b} = 2ah^2+3ha^2$$

A derivação implícita de (3.10) permite-nos escrever

$$\overline{x}^2 + 2P\overline{x}\frac{d\overline{x}}{dP} + Q\frac{d\overline{x}}{dP} = 0 \Rightarrow \frac{d\overline{x}}{dP} = \frac{-\overline{x}^2}{2P\overline{x}+Q} \tag{3.11}$$

$$2P\overline{x}\frac{d\overline{x}}{dQ} + \overline{x} + Q\frac{d\overline{x}}{dQ} = 0 \Rightarrow \frac{d\overline{x}}{dQ} = \frac{-\overline{x}}{2P\overline{x}+Q} \tag{3.12}$$

$$2P\overline{x}\frac{d\overline{x}}{dR} + Q\frac{d\overline{x}}{dR} + 1 = 0 \Rightarrow \frac{d\overline{x}}{dR} = \frac{-1}{2P\overline{x}+Q} \tag{3.13}$$

e pela derivada da composição , temos

$$\frac{d\overline{x}}{dy_a} = \frac{d\overline{x}}{dP}\frac{dP}{dy_a} + \frac{d\overline{x}}{dQ}\frac{dQ}{dy_a} + \frac{d\overline{x}}{dR}\frac{dR}{dy_a} \qquad \frac{d\overline{x}}{dy_b} = \frac{d\overline{x}}{dP}\frac{dP}{dy_b} + \frac{d\overline{x}}{dQ}\frac{dQ}{dy_b} + \frac{d\overline{x}}{dR}\frac{dR}{dy_b}$$
$$\frac{d\overline{x}}{dt_a} = \frac{d\overline{x}}{dP}\frac{dP}{dt_a} + \frac{d\overline{x}}{dQ}\frac{dQ}{dt_a} + \frac{d\overline{x}}{dR}\frac{dR}{dt_a} \qquad \frac{d\overline{x}}{dt_b} = \frac{d\overline{x}}{dP}\frac{dP}{dt_b} + \frac{d\overline{x}}{dQ}\frac{dQ}{dt_b} + \frac{d\overline{x}}{dR}\frac{dR}{dt_b}$$

Prosseguindo, obtemos

$$dFy_a = \frac{d\phi}{dy_a}(\overline{x}) = \frac{(\overline{x}-b)(2a-b-\overline{x}) - 2(\overline{x}-a)^2}{h^2} + \phi'(\overline{x})\cdot\frac{d\overline{x}}{dy_a} + \frac{2(\overline{x}-a)^3}{h^3} \tag{3.14}$$

$$dFy_b = \frac{d\phi}{dy_b}(\overline{x}) = \frac{3(\overline{x}-a)^2h - 2(\overline{x}-a)^3}{h^3} + \phi'(\overline{x})\cdot\frac{d\overline{x}}{dy_b} \tag{3.15}$$

$$dFt_a = \frac{d\phi}{dt_a}(\overline{x}) = \frac{(\overline{x}-a)^3 - (\overline{x}-a)(2x-a-b)h}{h^2} + \phi'(\overline{x})\cdot\frac{d\overline{x}}{dt_a} \tag{3.16}$$

$$dFt_b = \frac{d\phi}{dt_b}(\overline{x}) = \frac{(\overline{x}-a)^3 - h(\overline{x}-a)^2}{h^2} + \phi'(\overline{x})\cdot\frac{d\overline{x}}{dt_b} \tag{3.17}$$

3.2 Problema de otimização em espaço de dimensão finita.

Seja $\Phi^{np}_{[0,L]}[f](t)$ operador de interpolação cúbica np-segmentada no intervalo $[0, L]$. Notaremos

$$\tilde{S}(t) = \Phi^{np}[S](t) = \Phi^{np}(t)\cdot c_S, c_S = [c_0, c_1, ..., c_{np}, c_{np+1}, ..., c_{2np+1}]^T \in \Re^{2(np+1)} \tag{3.18}$$

Nosso problema é encontrar $c_{s^*} \in \Re^{2(np+1)}$ tal que

$$\tilde{S}^*(t) = \Phi^{np}(t)\cdot c_{S^*}$$

é a projeção da solução S^* do problema de controle ótimo definido na seção (1.3) no espaço das Interpoladoras Cúbicas np-segmentadas no intervalo $[0, L]$, ou seja,

$$J(\Phi^{np} \cdot c_{S^*}) = \min_{c_S \in \mathcal{A}^c} J(\Phi^{np} \cdot c_S)$$

onde definimos

$$\mathcal{A}^c = \{c_S \in \Re^{2(np+1)} : \min_{0 \le t \le t_0} \Phi^{np}(t) \cdot c_s \ge 0; \max_{0 \le t \le t_0} \Phi^{np}(t) \cdot c_s \le N; \int_0^{t_0} \Phi^{np}(t) \cdot c_s dt \le M\}$$

Nosso problema de controle é, desta forma, reescrito:

- encontrar $c_{S^*} \in \mathcal{A}^c \subset \Re^{2(np+1)}$ tal que

$$\tilde{J}(c_{S^*}) = \min_{c_S \in \mathcal{A}^c} \tilde{J}(c_S) \stackrel{def}{=} \min_{\Phi \cdot c_S \in A} J(\Phi \cdot c_S) \stackrel{def}{=} \min_{\tilde{S} \in A} J(\tilde{S}) = \min_{\tilde{S} \in \mathcal{A}} \int_0^{t_0} u(L,t)dt \tag{3.19}$$

e onde $u(x,t)$ e $\tilde{S}(t)$ se relacionam pelo sistema não linear em $(0,2L) \times (0,t_0)$:

$$C\frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial x^2} + \frac{T}{1+T}u$$

$$-\tilde{w}(t)\frac{\partial u}{\partial x} = \frac{T}{1+T}u$$

$T(x,0) = 0$ $\qquad$ $u(0,t) = u_o > 0$ $\qquad$ $\tilde{w}(t)$ função dada

$T(0,t) = \tilde{S}(t)$

$T(2L,t) = 0 \quad \forall t \in [0,t_0]$

3.3 Resolução do sistema de equações por diferenças finitas.

Nosso próximo passo é estabelecer como resolver numericamente as Equações Diferenciais Parciais que definem de uma forma implícita as concentrações $u(x,t)$ em termos da função de controle $\tilde{S}(t)$. Tais equações foram definidas na seção (3.2) e agora têm condições de contorno mais adequadas à implementação numérica.

$$C\frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial x^2} + \frac{T}{1+T}u \tag{3.20}$$

$$-\tilde{w}(t)\frac{\partial u}{\partial x} = \frac{T}{1+T}u \tag{3.21}$$

$$T(x,0) = 0, \qquad 0 \le x \le 2L \tag{3.22}$$

$$T(0,t) = \tilde{S}(t), \qquad 0 \le t \le t_0 \tag{3.23}$$

$$u(0,t) = u_o > 0, \qquad 0 \le t \le t_0 \tag{3.24}$$

$$T(2L,t) = T'(2L,t) = 0, \qquad 0 \le t \le t_0 \tag{3.25}$$

3.3.1 A Discretização das Equações

Para discretizarmos o espaço $[0,2L] \times [0,t_0]$ definimos

$$t_j = j.k; \qquad k = t_0/m$$
$$x_i = i.h; \qquad h = 2L/2n$$
$$T_i^j = T(x_i, t_j); \qquad U_i^j = U(x_i, t_j); \qquad w_j = \tilde{w}(t_j)$$

e usamos as aproximações :

$$\left[\frac{\partial T}{\partial t}\right]_i^j = \frac{11T_i^j - 18T_i^{j-1} + 9T_i^{j-2} - 2T_i^{j-3}}{6k} = \frac{\partial T}{\partial t}(x_i, t_j) + O(k^3)$$

$$\left[\frac{\partial^2 T}{\partial x^2}\right]_i^j = \frac{-T_{i-2}^j + 16T_{i-1}^j - 30T_i^j + 16T_{i+1}^j - T_{i+2}^j}{12h^2} = \frac{\partial^2 T}{\partial x^2}(x_i, t_j) + O(h^6)$$

$$\left[\frac{\partial U}{\partial x}\right]_i^j = \frac{11U_i^j - 18U_{i-1}^j + 9U_{i-2}^j - 2U_{i-3}^j}{6h} = \frac{\partial U}{\partial x}(x_i, t_j) + O(h^3)$$

$$\left[\frac{\partial^2 T}{\partial x^2}\right]_0^j = \frac{T_0^j - 2T_1^j + T_2^j}{h^2} = \frac{\partial^2 T}{\partial x^2}(0, t_j) + O(h^3) = C\frac{dS}{dt}(t_j) - \frac{S(t_j)u_0}{1 + S(t_j)} + O(h^3)$$

$$\overline{U_i^j} = 3U_i^{j-1} - 3U_i^{j-2} + U_i^{j-3} = u(x_i, t_j) + O(k^3)$$

Escrevemos a equação (3.20) em diferenças finitas:
$i = 1$

$$C\frac{dS}{dt}(t_j) = \frac{T_0^j - 2T_1^j + T_2^j}{h^2} + \frac{S(t_j)u_0}{1 + S(t_j)}$$

$i = 2, 2n$

$$C\frac{11T_i^j - 18T_i^{j-1} + 9T_i^{j-2} - 2T_i^{j-3}}{6k} =$$
$$\frac{-T_{i-2}^j + 16T_{i-1}^j - 30T_i^j + 16T_{i+1}^j - T_{i+2}^j}{12h^2} + \frac{T_i^j}{1 + T_i^j}(3U_i^{j-1} - 3U_i^{j-2} + U_i^{j-3})$$

Temos então um sistema de $2n$ equações não-lineares:

$$\bullet \; -T_0^j + 2T_1^j - T_2^j = h^2\left(\frac{S(t_j)u_0}{1 + S(t_j)} - CS'(t_j)\right) \tag{3.26}$$

$$\bullet \frac{k}{h^2}T_{i-2}^j - 16\frac{k}{h^2}T_{i-1}^j + (22C + 30\frac{k}{h^2})T_i^j - 16\frac{k}{h^2}T_{i+1}^j + \frac{k}{h^2}T_{i+2}^j - 12k\frac{T_i^j\overline{U_i^j}}{1 + T_i^j} \tag{3.27}$$
$$= 36CT_i^{j-1} - 18CT_i^{j-2} + 4CT_i^{j-3} \qquad (i = 2, 2n)$$

Escrevemos a equação (3.21) em diferenças finitas:
$i = 1, 2$

$$\frac{U_i^j - U_{i-1}^j}{h} = -\frac{T_i^j U_i^j}{w_j(1 + T_i^j)}$$

$i = 3, 2n$

$$\frac{11U_i^j - 18U_{i-1}^j + 9U_{i-2}^j - 2U_{i-3}^j}{6h} = -\frac{T_i^j U_i^j}{w_j(1+T_i^j)}$$

Temos então as recorrências:

$$\left(1 + \frac{hT_i^j}{w_j(1+T_i^j)}\right) U_i^j = U_{i-1}^j \qquad i = 1, 2 \tag{3.28}$$

$$\left(11 + \frac{6hT_i^j}{w_j(1+T_i^j)}\right) U_i^j = 18U_{i-1}^j - 9U_{i-2}^j + 2U_{i-3}^j \qquad i = 3, 2n \tag{3.29}$$

3.3.2 O algoritmo Posseidon

Apresentamos um algoritmo que resolve as equações (3.20)-(3.25) por diferenças finitas.

$ENTRADA\{u_0, C, np, nw, m, n, c_s(.), c_w(.), t_0, L\}$

- SPC$(\overline{t}, n_q, c_q)$: interpoladora cúbica de n_q nós e coeficientes $c_q(.)$ em $t = \overline{t}$;
- u_0, C: concentração inicial e calor específico;
- m, n: inteiros, definem os parâmetros de discretização k e h;
- $c_s(.)$: vetor real dos $(np+1)$ coeficientes de $\tilde{S}(t)$, isto é,$\tilde{S}(t) = \Phi^{np} \cdot c_S$;
- $c_w(.)$: vetor real dos $(nw+1)$ coeficientes de $\tilde{w}(t)$, isto é,$\tilde{w}(t) = \Phi^{nw} \cdot c_w$;
- t_0, L: definem o domínio de discretização $0 \leq x \leq 2L; 0 \leq t \leq t_0$

INICIALIZAÇÃO

$k \leftarrow t_0/m;$ $\qquad h \leftarrow L/n;$

$T_i^{-2} \leftarrow 0;$ $\qquad U_i^{-2} \leftarrow u_0;$

$T_i^{-1} \leftarrow 0;$ $\qquad U_i^{-1} \leftarrow u_0;$

$T_i^0 \leftarrow 0;$ $\qquad U_i^0 \leftarrow u_0;$

LOOP j=1:m

$t_j \leftarrow j * k;$

$T_0^j \leftarrow SPC(t_j, np, c_s);$

$w_j \leftarrow SPC(t_j, nw, c_w);$

$U_0^j \leftarrow u_0;$ $\qquad \overline{U_i^j} \leftarrow 3U_i^{j-1} - 3U_i^{j-2} + U_i^{j-3};$

Newton$[\ -T_0^j + 2T_1^j - T_2^j = h^2\left(\frac{S(t_j)u_0}{1+S(t_j)} - CS'(t_j)\right);$

$\frac{k}{h^2}T_{i-2}^j - 16\frac{k}{h^2}T_{i-1}^j + (22C + 30\frac{k}{h^2})T_i^j - 16\frac{k}{h^2}T_{i+1}^j + \frac{k}{h^2}T_{i+2}^j - 12k\frac{T_i^j\overline{U_i^j}}{1+T_i^j} =$

$$36CT_i^{j-1} - 18CT_i^{j-2} + 4CT_i^{j-3}, i = 2:2n\ ]$$

$$\text{Loop}[\qquad \left(1 + \frac{hT_i^j}{w_j(1+T_i^j)}\right) U_i^j = U_{i-1}^j \qquad i = 1,2$$

$$\left(11 + \frac{6kT_i^j}{w_j(1+T_i^j)}\right) U_i^j = 18U_{i-1}^j - 9U_{i-2}^j + 2U_{i-3}^j \quad i = 3:2n\ ]$$

$Ul(j) \leftarrow U_n^j$;

FIM-LOOP

RETORNA$\{Ul(j), j = 1:m\}$;

FIM.

A estabilidade do algoritmo Posseidon

Nosso próximo passo será definir precisamente a fronteira de estabilidade do algoritmo *Posseidon*.

Na determinação de T_i^j, temos as $(2n)$ equações não lineares (3.26)- (3.27); asseguraremos a estabilidade da solução T_i^j pela diagonal-dominância:

$$22C + 30\frac{k}{h^2} - \frac{12k\overline{U_i^j}}{1+T_i^j} > (1+16+16+1)\frac{k}{h^2}$$

observando que

$$22C + 30\frac{k}{h^2} - \frac{12k\overline{U_i^j}}{1+T_i^j} > 22C + 30\frac{k}{h^2} - 12ku_0$$

garantimos a estabilidade da solução T_i^j se

$$22C + 30\frac{k}{h^2} - 12ku_0 \geq 35\frac{k}{h^2} > 34\frac{k}{h^2} \Leftrightarrow k \leq \frac{22C}{5/h^2 + 12u_0} \tag{3.30}$$

lembramos que k é o parâmetro de discretização no tempo .

A despeito da determinação de U_i^j pelas equações (3.28),(3.29), mostramos que a equação em diferenças

$$(11+\beta_i)\alpha_i - 18\alpha_{i-1} + 9\alpha_{i-2} - 2\alpha_{i-3} = 0 \tag{3.31}$$

produz soluções estáveis se $\beta_i > 0$.

Para tal, seguimos o procedimento clássico de localização as raízes do polinômio característico associado $p(\lambda) = (11+\beta_i)\lambda^3 - 18\lambda^2 + 9\lambda - 2$.

Ora, se $\beta_i = 0, p_0(\lambda) = (\lambda - \lambda_0)(\lambda - \lambda_1)(\lambda - \lambda_2)$, onde

$$\lambda_0 = 1, \qquad \lambda_1 = \frac{7 - \imath\sqrt{39}}{22}, \qquad \lambda_2 = \frac{7 + \imath\sqrt{39}}{22} \qquad |\lambda_1| = |\lambda_2| < 1/2 < 1 \tag{3.32}$$

Para $\beta_i \ll 1$ e negligenciando termos em ϵ^3, temos $\lambda_0^{\beta_i} = 1 + \epsilon$, onde

$$(11+\beta_i)(1+\epsilon)^3 - 18(1+\epsilon)^2 + 9(1+\epsilon) - 2 = 0 \Rightarrow (15+3\beta_i)\epsilon^2 + 3(\beta_i+2)\epsilon + \beta_i = 0 \tag{3.33}$$

e sendo ϵ_1, ϵ_2 as raízes desta equação , temos

$$\epsilon = \frac{-3(2+\beta_i) \pm 6\sqrt{1 - 2/3\beta_i - \beta_i^2/12}}{2(15+3\beta_i)} \simeq \frac{-6-3\beta_i \pm 6[1-(2/3\beta_i - \beta_i^2/12)/2]}{2(15+3\beta_i)} \tag{3.34}$$

$$= \frac{-6-3\beta_i \pm (6-2\beta_i - \beta_i^2/4)}{30+6\beta_i}$$

e então

$$\epsilon_1 \simeq \frac{-5\beta_i - \beta_i^2/4}{30+6\beta_i} < 0 \qquad \epsilon_2 \simeq \frac{-12-\beta_i+\beta_i^2/4}{30+6\beta_i} < 0 \tag{3.35}$$

Logo as três raízes de $p(\lambda)$ permanecem dentro da bola unitária se $0 < \beta_i \ll 1$, em virtude de (3.32) e (3.35).

3.3.3 O procedimento Newton.

A tarefa do algoritmo Newton é resolver o sistema $(2n \times m)$ pentadiagonal não -linear que advém da discretização da equação diferencial parcial (3.20) no domínio espaço-tempo

$$-T_0^j + 2T_1^j - T_2^j = h^2\left(\frac{S(t_j)u_0}{1+S(t_j)} - CS'(t_j)\right);$$

$$\frac{k}{h^2}T_{i-2}^j - 16\frac{k}{h^2}T_{i-1}^j + (22C + 30\frac{k}{h^2})T_i^j - 16\frac{k}{h^2}T_{i+1}^j + \frac{k}{h^2}T_{i+2}^j - 12k\frac{T_i^j\overline{U_i^j}}{1+T_i^j} =$$

$$36CT_i^{j-1} - 18CT_i^{j-2} + 4CT_i^{j-3}, i = 2:2n$$

referida no algoritmo *Posseidon*.

Precisamente, a proposta é resolver o sistema não -linear

$$F(T_j) = AT_j - G(T_j) - B_j = 0$$

onde $A_{1,j} = 2\delta_{1,j} - \delta_{2,j}$;
$A_{i,j} = \frac{k}{h^2}\delta_{i-2,j} - 16\frac{k}{h^2}\delta_{i-1,j} + (22C + 30\frac{k}{h^2})\delta_{i,j} - 16\frac{k}{h^2}\delta_{i+1,j} + \frac{k}{h^2}\delta_{i+2,j}, i > 1, j \geq 1$;

$G_{i,j} = 12h\delta_{i,j}\frac{t_j u_i^j}{w_j(1+t_j)}$;

O termo B_j é determinado pelo lado direito de (3.26) e (3.27) para uma aproximação T_j^0.

Empregando o Método de Newton, que parte de uma aproximação T_j^0, procedemos à iteração dos sistemas algébricos pentadiagonais não-lineares

$$T_0 \leftarrow T_j^0;$$
$$Jac\,[F(T_l)]\,(T_l - T_{l+1}) = F(T_l), \qquad l = 0, 1, 2...$$

e obtemos uma aproximação T_{j+1} da solução de (3.26) e (3.27) que apresenta uma discrepância frente à aproximação imediatamente anterior definida pelo parâmetro *tol* (tipicamente 1%).

3.4 O cálculo numérico de derivadas parciais.

Conforme será detalhado na seção (4.3), nosso algoritmo de otimização demanda um procedimento que calcule numericamente o vetor de derivadas parciais do funcional de concentrações $\tilde{J}(c_s)$, a ser definido em pela equação (4.1), com respeito a cada uma das componentes do vetor c_s, ou seja, com respeito a cada um dos $2(np+1)$ coeficientes que determinam a função de controle $\tilde{S}(t)$ que, sendo o valor prescrito na fronteira $x=0$ do campo de temperaturas $T(x,t)$ determina univocamente soluções $T(x,t)$ e $u(x,t)$ para as equações (3.20)-(3.25) e então permite-nos avaliar uma aproximação numérica $FJ(c_s)$ de tal funcional das concentrações $u(x,t)$, referido acima como $\tilde{J}(c_s)$.

Nossa tarefa é então calcular

$$GradJ(c_s)_j = \#\frac{\partial \tilde{J}}{\partial c_{sj}}(c_s) \qquad j=0,1,...,2np+1 \tag{3.36}$$

onde c_{sj} é a j-ésima componente do vetor de coeficientes c_s. Esclarecer a tarefa computacional que está por trás do simbolo # é a principal meta desta seção .

Para calcular a aproximação numérica subentendida por (3.36), para cada j, usamos a aproximação por diferenças finitas

$$g_k = \frac{\tilde{J}(c_s - 2(h_0/2^k)e_j) - 8\tilde{J}(c_s-(h_0/2^k)e_j) + 8\tilde{J}(c_s+(h_0/2^k)e_j) - \tilde{J}(c_s+2(h_0/2^k)e_j)}{12(h_0/2^k)} \tag{3.37}$$

$$= \frac{\partial \tilde{J}}{\partial c_{sj}}(c_s) + O((h_0/2^k)^4)$$

onde temos não o valor exato $\tilde{J}(c_s)$, mas sim sua aproximação numérica $FJ(c_s)$, a qual não substituímos em (3.37) para não gerar imprecisão , mas que ficará implícito,

e onde $e_j = (\delta_{i,j})_{i=0,...,2np+1}$, $\qquad h_0 = |c_s|/(15(2np+2))$.

Definimos então as aproximações extrapoladas [1]

$$GradJ(c_s)^k_j = \frac{16g_{k+1}-g_k}{15} = \frac{\partial \tilde{J}}{\partial c_{sj}}(c_s) + O((h_0/2^k)^6) \tag{3.38}$$

que convergirão mais rapidamente do que os g_k de (3.37) para nosso objetivo em (3.36), para cada valor de j.

3.4.1 O algoritmo GradienteJ.

Apresentamos agora o algoritmo que perfaz a tarefa definida por (3.36).

ENTRADA$\{n, m, np, Lu, Tu, u_0, C, n_w, tol, c_s\}$

INICIALIZAÇÃO

$h_0 \leftarrow |c_s|/(15(2np+2))$;

$h_p \leftarrow Tu/np$;

[1] Ver Yakowitz,Szidarovsky,An Introduction to Numerical Computations, pg106.

$c_{s1}(j) = c_{s2}(j) = c_{s3}(j) = c_{s4}(j) \leftarrow c_s(j), j = 0, ..., 2np + 1;$

LOOP $j = 0 : 2np + 1$

$h \leftarrow h_0;$

$c_{s1}(j) \leftarrow c_s(j) - 2h;$

$c_{s2}(j) \leftarrow c_s(j) - h;$

$c_{s3}(j) \leftarrow c_s(j) + h;$

$c_{s4}(j) \leftarrow c_s(j) + 2h;$

$Ul \leftarrow Posseidon(u_0, C, np, n_w, m, n, c_{s1}, c_w, t_0, Lu);$

$fj1 \leftarrow FuncionalJ(m, t_0, Ul);$

$Ul \leftarrow Posseidon(u_0, C, np, n_w, m, n, c_{s2}, c_w, t_0, Lu);$

$fj2 \leftarrow FuncionalJ(m, t_0, Ul);$

$Ul \leftarrow Posseidon(u_0, C, np, n_w, m, n, c_{s3}, c_w, t_0, Lu);$

$fj3 \leftarrow FuncionalJ(m, t_0, Ul);$

$Ul \leftarrow Posseidon(u_0, C, np, n_w, m, n, c_{s4}, c_w, t_0, Lu);$

$fj4 \leftarrow FuncionalJ(m, t_0, Ul);$

$ga \leftarrow \frac{fj1 - 8fj2 + 8fj3 - fj4}{12h};$

REPITA

$fj1 \leftarrow fj2;$

$fj4 \leftarrow fj3;$

$h \leftarrow h/2;$

$c_{s2}(j) \leftarrow c_s(j) - h;$

$c_{s3}(j) \leftarrow c_s(j) + h;$

$Ul \leftarrow Posseidon(u_0, C, np, n_w, m, n, c_{s2}, c_w, t_0, Lu);$

$fj2 \leftarrow FuncionalJ(m, t_0, Ul);$

$Ul \leftarrow Posseidon(u_0, C, np, n_w, m, n, c_{s3}, c_w, t_0, Lu);$

$fj3 \leftarrow FuncionalJ(m, t_0, Ul);$

$g_k \leftarrow \frac{fj1 - 8fj2 + 8fj3 - fj4}{12h};$

$a_k \leftarrow |ga/gk - 1|;$

$ga \leftarrow g_k;$

ATÉ $(a_k < tol)$.

$GradJ(j) \leftarrow (16g_k - ga)/15;$

$c_{sk}(j) \leftarrow c_s j, k = 1, 2, 3, 4.$

FIM-LOOP

RETORNA $\{GradJ(j), j = 0; 2np + 1\}$;

FIM.

3.4.2 O subprocedimento FuncionalJ.

A tarefa de subprocedimento *FuncionalJ* é avaliar numericamente o funcional de concentrações

$$FJ(c_s) = \#\tilde{J}(c_s) = \int_0^{t_0} u(L,t)dt \tag{3.39}$$

quando temos somente uma discretização dos valores da concentração $u(x,t)$ na secção $x = L$. Os valores discretizados $u(L, t^j)$ $(t^j = j * t_0/m, j = 0 : m)$ estão armazenados no array $Ul(.)$ e são obtidos pelo procedimento *Posseidon*.

A estratégia adotada para avaliar (3.39) é usar a expressão (3.9), onde

$$y_j = Ul(j) = \#u(L, j * k), \quad k = t_0/m, j = 0 : m$$

$$d_0 = 0 = \frac{\partial u}{\partial t}(L, 0)$$

$$d_m = \frac{3Ul(m) - 4Ul(m-1) + Ul(m-2)}{2k} = \#\frac{\partial u}{\partial t}(L, t_0).$$

Assim, avaliamos a integral da cúbica que interpola $u(L,t)$ para $t = t^j$.

O subalgoritmo FuncionalJ.

ENTRADA$\{m, t_0, Ul(.)\}$

$k \leftarrow t_0/k$;

$d_m \leftarrow \frac{3Ul(m-2) - 4Ul(m-1) + Ul(m-2)}{2k}$;

$f_j \leftarrow (-d_m \cdot k/12 + Ul(0)/2 + Ul(m)/2) \cdot k$;

LOOP $i = 1 : m - 1$

$f_j \leftarrow f_j + k \cdot Ul(i)$;

FIM-LOOP

RETORNA $\{f_j\}$;

FIM.

3.5 O cálculo de projeções sobre hiperplanos.

Conforme será detalhado na subseção (4.3.3), o procedimento *Fronteira* precisa de um subprocedimento que, dado um vetor G_k e uma família de gradientes

$$GS_i \in \Re^{2(np+1)}, i = 1, ..., n_g,$$

calcule a projeção $w_k(.)$ do vetor $G_k(.)$ no espaço ortogonal ao espaço gerado por tal família.

Matematicamente, escrevemos

$$w_k = G_k - \alpha_1 GS_1 - \alpha_2 GS_2 - ... - \alpha_{n_g} GS_{n_g} \tag{3.40}$$

e então equacionamos a ortogonalidade :

$$\langle w_k, GS_i \rangle = 0, i = 1, 2, ..., n_g \tag{3.41}$$

o que nos conduz a um sistema de n_g equações

$$\begin{bmatrix} \langle GS_1, GS_1 \rangle & \langle GS_2, GS_1 \rangle & .... & \langle GS_{n_g}, GS_1 \rangle \\ \langle GS_1, GS_2 \rangle & \langle GS_2, GS_2 \rangle & .... & \langle GS_{n_g}, GS_2 \rangle \\ \langle GS_1, GS_{n_g} \rangle & \langle GS_2, GS_{n_g} \rangle & .... & \langle GS_{n_g}, GS_{n_g} \rangle \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ .... \\ \alpha_{n_g} \end{bmatrix} = \begin{bmatrix} \langle G_k, GS_1 \rangle \\ \langle G_k, GS_2 \rangle \\ ... \\ \langle G_k, GS_{n_g} \rangle \end{bmatrix} \tag{3.42}$$

A simetria deste sistema de equações sugere uma estratégia de resolução do tipo fatoração LDL^T , e será este, exatamente, o caminho a ser seguido [2] Assim, precisaremos de rotinas numéricas que implementem os respectivos algoritmos de fatoração e solução o mais eficientemente possível, principalmente em termos do tratamento eficaz (refinamento) do erro numérico que naturalmente advém neste tipo particular de aplicação , devido à imprecisão no cálculo dos produtos internos que aparecem em (3.42).

[2]Ref.: Ortega & Poole. An Introduction to numerical methods for differential equations. John Wiley & Sons, 1981.

4 O PROBLEMA DE CONTROLE DE FRONTEIRA LIVRE.

Lembrando nosso problema de controle ótimo, definido na seção (1.3),

• encontrar o vetor de coeficientes de interpolação $c_{S^*} \in \mathcal{A}^c \subset \Re^{2(np+1)}$ tal que

$$\tilde{J}(c_{S^*}) = \min_{c_S \in \mathcal{A}^c} \tilde{J}(c_S) \stackrel{def}{=} \min_{\Phi \cdot c_S \in \mathcal{A}} J(\Phi \cdot c_S) \stackrel{def}{=} \min_{\tilde{S} \in \mathcal{A}} J(\tilde{S}) = \min_{\tilde{S} \in \mathcal{A}} \int_0^{t_0} u(L,t)dt \tag{4.1}$$

onde

$$\mathcal{A} = \{S(t), 0 \leq t \leq t_o; S(t) \in C^1((0,t_0)), 0 \leq S(t) \leq N, \int_0^{t_o} S(t)dt \leq M\}$$

$$\mathcal{A}^c_{np} = \{c_S \in \Re^{2(np+1)} : \min_{0 \leq t \leq t_0} \Phi^{np}(t) \cdot c_s \geq 0; \max_{0 \leq t \leq t_0} \Phi^{np}(t) \cdot c_s \leq N; \int_0^{t_0} \Phi^{np}(t) \cdot c_s dt \leq M\} \tag{4.2}$$

e onde $u(x,t)$ e $\tilde{S}(t)$ se relacionam pelo sistema não -linear em $(0,2L) \times (0,t_0)$:

$$C\frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial x^2} + \frac{T}{1+T}u \tag{4.3}$$

$$-\tilde{w}(t)\frac{\partial u}{\partial x} = \frac{T}{1+T}u \tag{4.4}$$

$$T(x,0) = 0, \qquad u(0,t) = u_o > 0 \tag{4.5}$$

$$T(0,t) = \tilde{S}(t) \tag{4.6}$$

$$T(2L,t) = T'(2L,t) = 0, \qquad 0 \leq t \leq t_0 \tag{4.7}$$

identificamos uma tarefa de otimização restrita, onde (4.1) define a função objetiva, a ser minimizada sobre o espaço $2(np+1)$ dimensional definido por (4.2). Procuramos então por procedimentos computacionais de busca ao controle ótimo $\tilde{S}^*(t) \in \tilde{\mathcal{A}}_{np}$, bem como por resultados que assegurem a solubilidade unívoca e estável desta tarefa. Dizemos também tratar-se de um problema de controle de fronteira livre, uma vez que os controles $\tilde{S}(t) \in \tilde{\mathcal{A}}_{np}$ são condições de fronteira da temperatura $T(x,t)$, incógnita no sistema de equações parciais (4.3)-(4.7) acima.

4.1 A convexidade do espaço de controles admissíveis $\mathcal{A}^c_{np}$.

O próximo passo é demonstrar a convexidade do espaço $\mathcal{A}^c_{np} \subset \Re^{2(np+1)}$, a fim de assegurar a existência de uma solução ótima $\tilde{S}^*(t)$ e até mesmo sua determinação computacional via procedimentos de otimização restrita clássicos, que são encontrados na abundante literatura de programação não-linear hoje disponível.

Lembramos a definição do espaço $\mathcal{A}$ de controles admissíveis acima, cuja convexidade pode ser facilmente demonstrada, e também a definição de $\mathcal{A}^c_{np}$ em (4.2), espaço que agora mostraremos ser convexo.

É fácil mostrar que $c_s \in \mathcal{A}^c_{np} \Rightarrow \tilde{S}(t) = \Phi^{np}(t) \cdot c_s \in \mathcal{A}$.

Sejam $c_s^1, c_s^2 \in \mathcal{A}_{np}^c$, $\tilde{S}_1(t) = \Phi^{np}(t) \cdot c_s^1, \tilde{S}_2(t) = \Phi^{np}(t) \cdot c_s^2$.

Definimos agora $P : \Re^{2(np+1)} \times \Re \longrightarrow \Re$

$$P(c,t) = \langle c, \phi^{np}(t) \rangle = \Phi^{np}_{[0,t_0]} \cdot c, c \in \Re^{2(np+1)}, t \in [0,t_0] \tag{4.8}$$

Está claro, pela definição acima, que estamos vendo os controles $\tilde{S}(t) \in \mathcal{A}$ como funções $P(c_s,t)$ do tempo t e do respectivo vetor de coeficientes nodais e tangenciais que definem a interpoladora cúbica $\tilde{S}(t)$; e que também salientamos sua estrutura de produto interno Euclidiano.

Sejam

$$B_N^c = \{c \in \Re^{2(np+1)} : P(c,t) \leq N, 0 \leq t \leq t_0\} \tag{4.9}$$

$$B_O^c = \{c \in \Re^{2(np+1)} : P(c,t) \geq 0, 0 \leq t \leq t_0\} \tag{4.10}$$

$$B_M^c = \{c \in \Re^{2(np+1)} : \int_0^{t_0} P(c,t)dt = \left\langle \int_0^{t_0} \phi^{np}(t)dt, c \right\rangle \leq M\} \tag{4.11}$$

Mostraremos que B_N^c é um espaço convexo.

Para tal, suponhamos que B_N^c não é convexo. Então existem dois vetores $c_1, c_2 \in B_N^c$ tais que, para algum $\lambda_0, 0 \leq \lambda_0 \leq 1$ temos

$$c_{\lambda_0} = (1-\lambda_0)c_1 + \lambda_0 c_2 \notin B_N^c$$

entretanto, $\forall t \in [0,t_0]$

$$P(c_{\lambda_0},t) = \langle (1-\lambda_0)c_1 + \lambda_0 c_2, \phi^{np}(t) \rangle = (1-\lambda_0)P(c_1,t) + \lambda_0 P(c_2,t) \leq (1-\lambda_0)N + \lambda_0 N = N$$

o que contraria a hipótese que $c_{\lambda_0} \notin B_N^c$.

Segue-se que B_N^c é convexo e, analogamente, B_O^c também é convexo.

Para mostrar a convexidade de B_M^c, lembramos que, pela expressão (3.9) e considerando interpolação np-segmentada regular de intervalo h,

$$B_M^c = \{c_i : [\frac{c_0}{2} + c_1 + c_2 + ... + c_{n-1} + \frac{c_{np}}{2}]h + [c_{np+1} - c_{2*np+1}]\frac{h^2}{12} \leq M\} \tag{4.12}$$

e então a fronteira de B_M^c é o hiperplano definido por (4.12) e a convexidade segue trivialmente.

Por construção , $\mathcal{A}_{np}^c = B_N^c \bigcap B_O^c \bigcap B_M^c$ e então segue a convexidade de $\mathcal{A}_{np}^c$.

4.2 Algoritmos de Otimização Restrita.

4.2.1 O método das direções viáveis de Zoutendijk (1960).

Zoutendijk desenvolveu um método para resolver os problemas de programação mais generalizados, onde ambas a função objetiva $f(X)$ e a função de restrições $W(X)$ são não-lineares e convexas.

O algoritmo começa em um ponto X_1 na região viável para as restrições dadas. Neste método, não é necessário que o ponto inicial seja uma solução viável básica. A seguir, uma direção viável de movimento é determinada. A melhor direção possível para mover é a direção do gradiente

da função objetiva avaliado no ponto, uma vez que produz o maior decrescimento do seu valor. Por direção viável é entendido aquela ao longo da qual um pequeno passo pode ser dado sem violar qualquer uma das restrições , e em geral, o gradiente não determina uma direção viável. A direção viável escolhida é aquela que faz o menor ângulo possível θ com a direção definida pelo vetor gradiente da função objetiva. Entretanto, podem haver armadilhas. Se a restrição ativa (a restrição que forma a parte da fronteira onde está nossa corrente aproximação) é linear, tudo é satisfatório e uma das duas direções definidas por essa restrição é escolhida. Entretanto, se a restrição ativa é não -linear, é possível que este procedimento produza uma direção que saia da região viável. Uma vez dado tal passo, teríamos que dar um *salto* de volta para a região viável. Mas não há garantia que tais pares de passos não sejam executados repetitivamente, causando um zig-zag ineficiente. Para evitar estes e outros inconvenientes possíveis , torna-se fortemente óbvio que nós devamos escolher uma direção a qual mova decisivamente para o interior da região viável ao mesmo tempo que diminua o valor da função objetiva. Para esta proposta, a direção desejável d é encontrada resolvendo-se o seguinte problema

Maximizar: E

sujeito a:

$$(\nabla W_i(X_1))^T d + t_i E \leq 0, \quad \forall i : W_i(X_1) = 0, 0 \leq t_i \leq 1.$$

$$(\nabla f(X_1))^T d \geq E, \quad d^T d = 1.$$

A direção d^* que é a solução do problema colocado acima é a solução mais apropriada para usarmos. Prosseguimos nesta direção tão longe quanto possamos até que a função objetiva f começe a crescer ou então encontremos a fronteira da região viável novamente. O processo é então repetido até que o máximo valor para E seja não positivo. O processo é então terminado. Se todas as funções são convexas, o mínimo global é então encontrado. Este método frequentemente performa bem em problemas de minimização onde não temos a hipótese de convexidade.

4.2.2 O método do gradiente projetado de Rosen (1960).

Outro método para a solução de problemas de programação matemática foi desenvolvido por Rosen. Este método performa melhor quando as restrições são lineares, mas também pode ser aplicado na solução de problemas tendo restrições não-lineares. Ao atingirmos a fronteira determinada pelas restrições , ao invés de computarmos a direção que causa o maior decrescimento na função objetiva, o método simplesmente escolhe a direção na qual tal função decresce e na qual também assegura que possamos mover sem imediatamente violar alguma restrição . Para restrições lineares, esta direção é a projeção do gradiente no plano das restrições tomadas como igualdades estritas no ponto onde o gradiente está sendo calculado. Assim, a direção de cada passo subsequente pode ser determinada sem resolver um problema de programação linear, e a viabilidade é mantida, uma vez que o caminho percorrido nunca atravessa a fronteira determinada pelas restrições . O tamanho dos passos é computado analogamente ao método de Zoutendijk, e depois de cada passo o gradiente é recomputado e uma nova direção é calculada.

Para restrições não -lineares, esta direção é definida como a projeção do vetor gradiente da função objetiva f sobre a intersecção dos hiperplanos associados a cada uma das m restrições ativas. Se não existirem restrições ativas, a direção daquele gradiente é escolhida como direção de procura.

No detalhamento do método a ser aqui apresentado, será assumido que todas as funções $W_i, i = 1, 2, ..., m$ que definem a fronteira da região viável quanto $W_i(X) = 0$, são lineares. Especificamente, assumimos que o método começa com uma estimativa X_1 da solução X_*. Seja $X_k = X_1$,

1. Primeiro calculamos $\nabla f(X_k)$.

2. Seja P_k o conjunto dos hiperplanos correspondentes às restricoes ativas em X_k.

3. Encontre a projeção do gradiente $\nabla f(X_k)$ sobre a intersecção daqueles hiperplanos em P_k. (Se não houver restrições ativas, X_k é um ponto interior e P_k é vazio. Neste caso a projeção é $\nabla f(X_k)$).

4. Minimizar ao longo da direção desta projeção , tomando o cuidado para permanecer no interior da região viável.

5. Tal procedimento produz um novo ponto X_{k+1}.

6.(a) Se P_k era não -vazio no passo 2, troque X_k por X_{k+1} e retorne ao passo 1.

(b) Se P_k era vazio no passo 2, então

$$\frac{\partial f}{\partial x_i} + \sum_{i=1}^{q} \lambda_i \frac{\partial W_i}{\partial x_i} = 0, \tag{4.13}$$

onde as funções $W_1, W_2, ..., W_q$ são as funções correspondentes as hiperplanos em P_k. Se

$$\lambda_i \geq 0, i = 1, 2, ..., q, \tag{4.14}$$

X_k satisfaz as condições de Kuhn-Tucker. Então X_k é um mínimo. Se ao menos um λ_i é tal que

$$\lambda_i < 0, \tag{4.15}$$

o plano correspondente à função W_i para a qual (4.15) vale é removido de P_k, e retornamos ao passo 2.

O método não é tão eficiente em problemas com restrições não -lineares. Nestes casos, as projeções são feitas sobre hiperplanos tangentes às superfícies de restrição . Passos dados nestes hiperplanos podem muito bem conduzir para fora da região viável, e assim um procedimento de reconduza-nos de à região viável (um salto) é necessário para que então o algoritmo, possa prosseguir, e então vale a preocupação colocada na subseção anterior, no sentido de previnir um zig-zag ineficiente.

4.3 Um algoritmo de Otimização

A despeito dos procedimentos de otimização necessários para a solução de nosso problema de controle de fronteira livre, apontamos algumas tarefas fundamentais:

Dado $c_s^k \in \mathcal{A}^c$, sendo U_i^j, T_i^j as soluções discretizadas do problema do conversor catalítico, na notação da seção (3.2) e para $0 \leq x \leq 2L, 0 \leq t \leq t_0$

$$C\frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial x^2} + \frac{T}{1+T}u$$

$$-\Phi^{nw}(t) \cdot c_w \frac{\partial u}{\partial x} = \frac{T}{1+T}u$$

$T(x,0) = 0 \qquad u(0,t) = u_o > 0$

$T(0,t) = \Phi^{np}(t) \cdot c_s$

$T(2L,t) = T'(2L,t) = 0, \qquad 0 \leq t \leq t_0$

definimos

$$FJ(c_s^k) = Quadratura[U_n^j, j = 1, 2, ..., m] = \#\tilde{J}(c_s^k) = \int_0^{t_0} u(L,t)dt \tag{4.16}$$

Nossa próxima tarefa é calcular

$$GradJ(c_s^k) = \#\frac{\partial \tilde{J}}{\partial c_s}(c_s^k) \tag{4.17}$$

o gradiente da função objetiva no ponto c_s^k.

A tarefa global do algoritmo Pegasus, a ser descrito em seguida, será encontrar $c_s^{k+1} \in \mathcal{A}^c$ tal que $FJ(c_s^{k+1}) < FJ(c_s^k)$.

Nesse sentido, para $c_s^k \in int(\mathcal{A}^c)$ encontraremos c_s^{k+1} minimizando
$FJ(c_s^k - \lambda GradJ(c_s^k))$

para $\lambda \in [0, \lambda_0]$ (λ_0 tal que $c_s^k - \lambda_0 GradJ(c_s^k) \in \mathcal{A}^c$), ou seja, na direção contrária ao gradiente do funcional a minimizar, como no *Método da Descida*.

Para $c_s^k \in \partial\mathcal{A}^c$, descreveremos $\mathcal{A}^c = \{c_s \in \Re^{2(np+1)} : W^i(c_s) \leq 0, i = 1, 2, ..., 2(np+2)\}$ e definiremos os normais de superfície

$$GradS_i(c_s^k) = \#\frac{\partial W^i}{\partial c_s}(c_s^k), i = 1, 2, ..., 2(np+1) \tag{4.18}$$

e então c_s^{k+1} será encontrado segundo o Método do Gradiente Projetado de Rosen, observando que poderemos ter que considerar não um normal único, mas talvez uma família de até $2(np+1)$ normais de superfície.

4.3.1 O algoritmo Pegasus

Apresentamos um algoritmo para resolver o problema de otimização .

ENTRADA$\{c_w(.), c_s^k(.), C, u_o, np, nw, m, n\}$

- u_0, C: concentração inicial e calor específico;
- m, n: inteiros, definem os parâmetros de discretização k e h;
- $Ul(.)$: vetor dos $(m+1)$ valores discretizados da concentração na secção x=L;
- $c_s(.)$: vetor real dos $(np+1)$ coeficientes de $\tilde{S}(t)$,isto é,$\tilde{S}(t) = \Phi^{np} \cdot c_S$;
- $c_w(.)$: vetor real dos $(nw+1)$ coeficientes de $\tilde{w}(t)$,isto é,$\tilde{w}(t) = \Phi^{nw} \cdot c_w$;
- t_o, L: definem o domínio de discretização $0 \leq x \leq 2L; 0 \leq t \leq t_o$;
- lfrn,lfrm,lnl: indicam se estamos no interior ou na fronteira do espaço admissível $\mathcal{A}$;

$Ul(.) \leftarrow Posseidon(u_o, C, np, nw, m, n, c_s(.), c_w(.), t_o, L)$;
$f_j \leftarrow FuncionalJ(m, t_o, Ul(.))$;
$GradJ(.) \leftarrow GradienteJ(c_s(.), n, m, np, L, t_o, u_o, c, nw, tol)$;
$[lfrn, lfrm, lnl, \lambda_0, nrest(.)]$ ←Restrições $(np, c_s(.), GradJ(.), resn, resm)$;
SE (lfrm) ou (lfrn) ENTÃO

$c_s^{k+1} \leftarrow Fronteira(lnl, np, c_s(.), GradJ(.), N, M, nrest(.));$

SENÃO

$\lambda^* \leftarrow Minimiza(c_s(.), nl, GradJ(.), \lambda_0);$

$c_s^{k+1} \leftarrow c_s^k - \lambda^* GradJ;$

FIM-SE

FIM.

4.3.2 O procedimento Restrições .

Ao procedimento restrições cabe a tarefa fundamental de não permitir que porventura saiamos do espaço admissível (região viável na terminologia da seção anterior). Tal espaço, que define as restrições de nosso problema de otimização discreto, pode ser descrito por

$$\mathcal{A}^c = \{c_S \in \Re^{2(np+1)} : \min_{0\le t\le t_0} \Phi^{np}(t)\cdot c_s \ge 0; \max_{0\le t\le t_0} \Phi^{np}(t)\cdot c_s \le N; \int_0^{t_0} \Phi^{np}(t)\cdot c_s dt \le M\} \quad (4.19)$$

e então cada vez que tivermos uma aproximação c_s^k e uma direção d_s, procuraremos o maior $\lambda_0 \ge 0$ tal que

$$c_s^k - \lambda d_s \in \mathcal{A}^c, \qquad \forall \lambda : 0 \le \lambda \le \lambda_0;$$

e assim estamos na fronteira $\partial\mathcal{A}^c$ se $\lambda_0 = 0$, caso contrário, d_s é uma direção viável ($\lambda > 0$), e podemos procurar por um $\overline{\lambda}$ tal que $0 < \overline{\lambda} \le \lambda_0$ e

$$\tilde{J}(c_s^k - \overline{\lambda} d_s) = \min_{0<\lambda<\lambda_0} \tilde{J}(c_s^k - \lambda d_s),$$

fazendo $c_s^{k+1} = c_s^k - \overline{\lambda} d_s$ temos a garantia que $c_s^{k+1} \in \mathcal{A}^c$ e podemos seguir adiante com o algoritmo de otimização .

Nos caso $\lambda_0 = 0$, o procedimento também deverá fornecer uma lista $nrest(.)$ com todas as restrições ativas, para que, então , o procedimento Fronteira possa ser acionado, projetando o gradiente da função objetiva $\tilde{J}(c_s)$ no espaço tangente à superficie determinada por aquelas restrições . Também neste caso deverá indicar com os flags $lnl, lfrn, lfrm$ se as restrições são não -lineares, se estamos em ∂B_N^c ou em ∂B_M^c, respectivamente e na notação da seção (4.1).

O algoritmo Restrições .

ENTRADA $\{np, c_s(.), d_k(.), N, M, hp\};$

$\lambda_{max} \leftarrow |c_s|\sqrt{2.01}/(10|d_s|);$

$\lambda_i \leftarrow 0;$

$\lambda_f \leftarrow \lambda_{max};$

LOOP $i = 1 : 40$

$\lambda_m \leftarrow (\lambda_i + \lambda_f)/2$;

$c_c(j) \leftarrow c_s(j) - \lambda_m d_s(j), j = 0 : 2np + 1$;

$[lfrn, lfrm, lnl, nrest(.)] \leftarrow Restri(np, c_c, hp, N, M)$;

Se $(lfrn)$ ou $(lfrn)$ então

$\lambda_f \leftarrow \lambda_m$;

senão

$\lambda_i \leftarrow \lambda_m$;

Fim-se

FIM-LOOP.

Se $(\lambda_i > 0)$ então

$c_c(j) \leftarrow c_s(j) - \lambda_i d_s(j), j = 0 : 2np + 1$;

$[lfrn, lfrm, lnl, nrest(.)] \leftarrow Restri(np, c_c, hp, N, M)$;

Fim-se

$\lambda_0 \leftarrow \lambda_m$;

RETORNA$\{lfrn, lfrm, lnl, \lambda_0, nrest(.)\}$;

FIM.

O subalgoritmo Restri.

ENTRADA$\{np, c_c, hp, N, M\}$

$lfrm \leftarrow lfrn \leftarrow lnl \leftarrow false$;

Se $(c_c \notin B_N^c) lfrn \leftarrow true$;

Se $(c_c \notin B_M^c) lfrm \leftarrow true$;

Se $(lfrn)$ ou $(lfrm)$ então

Se ($\exists$ restrições não -lineares) $lnl \leftarrow true$;

$nrest(.) \leftarrow$ [lista das restrições violadas];

Fim-se

RETORNA$\{lfrn, lfrm, lnl, nrest(.)\}$;

FIM.

4.3.3 O procedimento Fronteira.

Se o procedimento *GradienteJ*, apresentado na seção (3.4), representa o módulo de processamento mais intenso, cabe ao procedimento *Fronteira* a tarefa mais importante e preponderante em termos da própria algoritmização da tarefa global de busca ao contrle ótimo: definir qual a direção a ser tomada quando nossa aproximação chega à fronteira do espaço admissível e a direção de maior decrescimento não é viável, garantindo ao mesmo tempo o decrescimento da função objetiva e o retorno à região admissível.

O procedimento *Fronteira* é, basicamente, a algoritmização do próprio Método do Gradiente Projetado de Rosen, mas também algoritmiza alguns procedimentos que procuram diminuir o impacto da não linearidade da função de restrições na performace e mesmo na viabilização da respectiva implementação numérica.

O procedimento inicia com uma aproximação c_k e uma direção não viável determinada pelo sentido oposto ao gradiente G_k (direção de maior decrescimento da função objetiva). O primeiro passo é determinar o conjunto de todas as n_g restrições ativas, determinar os respectivos gradientes

$$GS_i(c_k) = \#\frac{\partial W^i}{\partial c_s}(c_k), i = 1, 2, ..., n_g \tag{4.20}$$

das funções $W^i(c_k)$ que as definem e calcular a projeção w_k do vetor G_k (gradiente da função objetiva) no espaço ortogonal àquela família de gradientes (que é o plano tangente à superfície determinada pela fronteira da região admissível no ponto c_k). Este primeiro passo é executado pelo procedimento *twk* , que também contabiliza o número n_{g_n} de restrições ativas não lineares. Não apresentaremos uma descrição do algoritmo *twk*, uma vez que sua tarefa principal já foi descrita na seção (3.5).

O segundo passo, executado se existem restrições não lineares, deverá determinar uma nova direção w_k^p na qual, ao mesmo tempo, a próxima aproximação seja conduzida à região admissível e seja garantido o decrescimento da função objetiva. O que determina a inviabilidade da direção inicial $-G_k$ é a própria convexidade da fronteira admissível, que implica que qualquer plano tangente não contenha algum de seus pontos interiores, e nesse sentido a estratégia genérica do Método do Gradiente Projetado é projetar o vetor $-w_k$ na fronteira da região admissível, ou seja, a próxima aproximação c_{k+1} estaria determinada por

$$dist(c_{k+1}, c_k - w_k) = \min_{c_s \in \partial\mathcal{A}^c} dist(c_s, c_w - w_k), c_{k+1} \in \partial\mathcal{A}^c$$

conforme podemos interpretar graficamente na figura (4.1).

Nesse sentido, a estratégia adotada será usar um vetor GFi cuja direção oposta conduza rapidamente ao interior do espaço admissível, se existe apenas uma restrição ativa não linear, GFi é o próprio gradiente da função que determina tal restrição , caso contrário, uma direção mista é determinada. A estratégia, que pode ser interpretada graficamente na figura (4.2), é determinar c_{k+1} por

$$c_{k+1} = c_k - \rho_0 GFi - \alpha(\rho_0)w_k = c_k - w_k^p \in \partial\mathcal{A}^c$$

e onde a idéia é tomar tal ρ_0 de maneira que, dentre todas as direções $\rho GFi + \alpha(\rho)w_k$ nas quais a função objetiva decresce de valor, isto é, nas quais

$$\langle G_k, -\rho GFi - \alpha(\rho)w_k\rangle < 0$$

escolhemos aquela que conduz a um ponto na fronteira que mais se aproxima do ponto exterior $c_k - w_k$. Desta forma, cumprimos parcialmente a estratégia definida pelo Método de Rosen, e também asseguramos a monotonicidade das aproximações c_k, o que elimina a possibilidade de entrarmos num zig-zag ineficiente, conforme já foi alertado em (4.2.2).

O terceiro passo, executado se todas as restrições ativas são lineares e então a fronteira, localmente, é um hiperplano, deverá tão somente computar a projeção w_k do gradiente G_k no hiperplano determinado pelas restrições ativas. Assim, exceto por imprecisões de natureza numérica (que tem origem sobretudo no cálculo de projeções ortogonais), a direção determinada por w_k deverá estar totalmente dentro da região admissível em alguma vizinhança de c_k, e o algoritmo continua com $c_{k+1} = c_k - \lambda_0 w_k$, onde λ_0 é o maior λ que assegura $c_k - \lambda w_k \in \mathcal{A}^c$ e representa o tamanho do maior passo que podemos dar sem que saiamos de $\mathcal{A}^c$.

O algoritmo Fronteira

Apresentamos então uma descrição um pouco mais detalhada do procedimento acima, lembrando que os parâmetros N,M já foram definidos como os valores numéricos que definem a fronteira admissível $\mathcal{A}^c$.

ENTRADA$\{lnl, np, c_k, N, M, h_p, nrest(.)\}$

$\{w_k, n_{g_n}\} \leftarrow twk\{$ projeção de $G_k\}$;

$prod \leftarrow \langle G_k, w_k \rangle$;

SE $(prod = 0)$ encontramos o ótimo;

SE $(prod < 0)$ ENTÃO

SE (lnl) ENTÃO

$GFi \leftarrow \{$ qualquer vetor tal que $-GFi$ direção viável $\}$;

$\rho_b \leftarrow \{$maior ρ tal que $\langle G_k, -\rho GFi - \alpha(\rho) w_k \rangle < 0, \alpha(\rho)$ é tal que $c_k - \rho GFi - \alpha(\rho) w_k \in \partial \mathcal{A}^c\}$;

$\rho_0 \leftarrow \{\rho \in (0, \rho_b)$ que minimiza $dist(c_k - \rho GFi - \alpha(\rho) w_k, c_k - w_k\}$;

$\alpha_0 \leftarrow \alpha(\rho_0)$;

$c_{k+1} \leftarrow c_k - \rho_0 GFi - \alpha_0 w_k$;

SENÃO

$\alpha_0 \leftarrow \{$ maior $\alpha : c_k - \alpha w_k \in \mathcal{A}^c\}$;

$c_{k+1} \leftarrow c_k - \alpha_0 w_k$;

FIM-SE

FIM-SE

RETORNA $\{c_{k+1}\}$.

FIM.

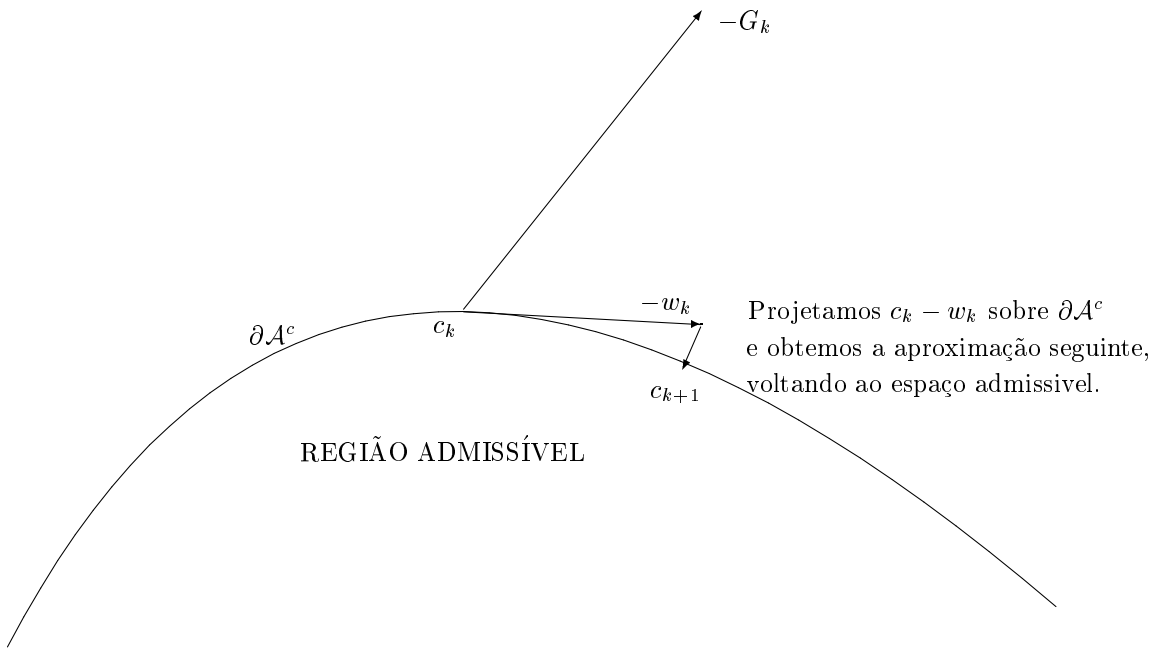

Figura 4.1: Estratégia original do Método do Gradiente Projetado.

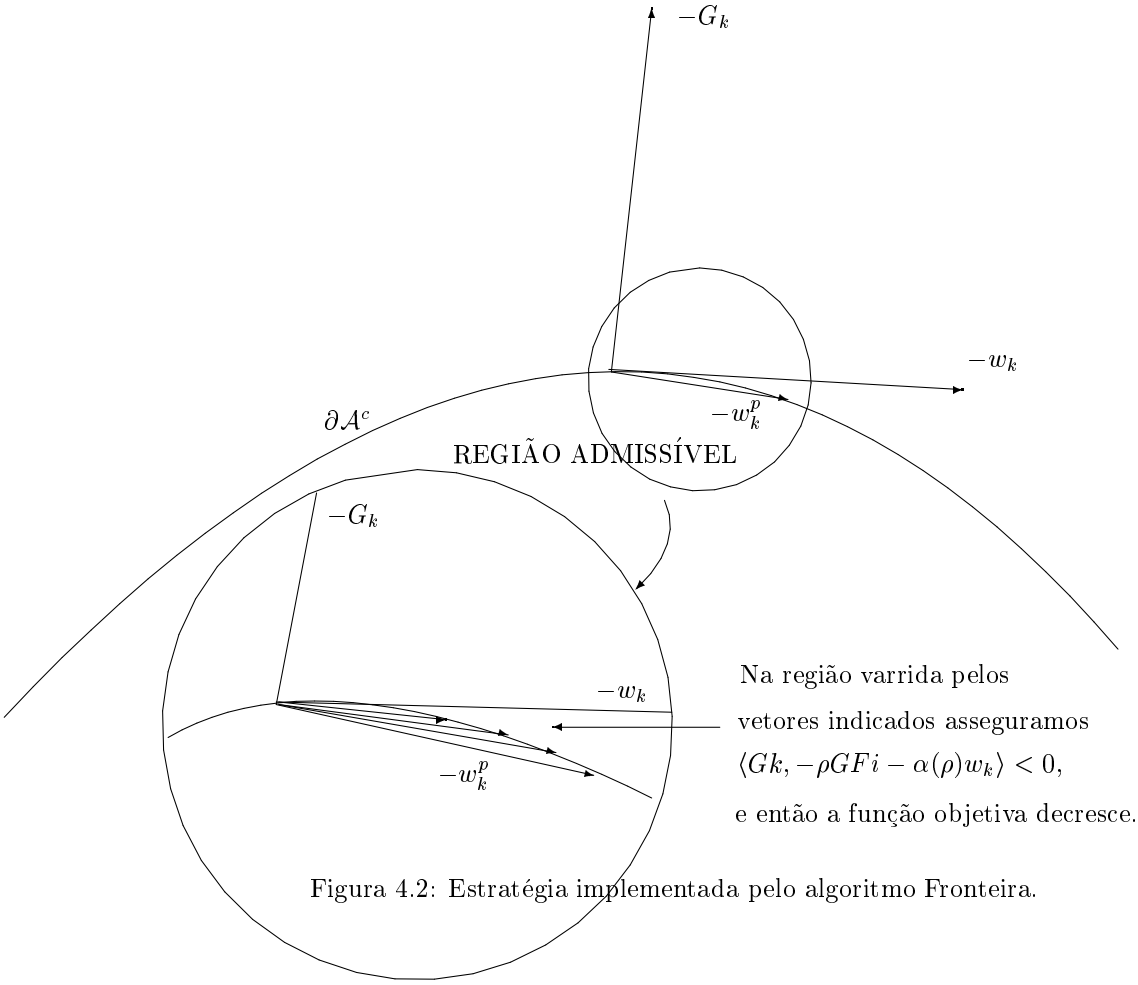

Figura 4.2: Estratégia implementada pelo algoritmo Fronteira.

5 A TAREFA COMPUTACIONAL.

5.1 A complexidade computacional.

No sentido de descrever a complexidade computacional de nosso problema de otimização , mediremos tal tarefa em termos do número de sistemas lineares algébricos pentadiagonais resolvidos pelo procedimento *Posseidon* a cada iteração do algoritmo *Pegasus*, o último descrito na seção (4.3) e o primeiro descrito na seção (3.2). Tal estratégia revela-se lícita uma vez que, com o progressivo aumento da capacidade de processamento dos computadores, parâmetros clássicos de desempenho, como o número de operações ponto flutuante, por exemplo, tornam-se cada vez menos significativos para uma compreensão comparativa de complexidade.

Conforme detalhado na subseção (3.3.2), o procedimento *Posseidon*, o responsável pela resolução dos problemas de fronteira por diferenças finitas, resolve m sistemas não lineares algébricos de dimensão $2n$, onde m é o número de passos de tamanho k no tempo e $2n$ é o número de passos de tamanho h no espaço. Os parâmetros k e h devem relacionar-se por (3.30) a fim de garantir a estabilidade numérica do procedimento. Na solução de cada sistema não linear é empregado o Método de Newton, sendo a convergência reconhecida segundo o critério da proximidade relativa de cada duas aproximações sucessivas e no máximo em 7 iterações , para manter bom desempenho. Temos então , no pior caso, que resolver $7m$ sistemas lineares pentadiagonais $(2n \times 2n)$ descritos pelas equações (3.26) e (3.27). No caso médio, para tolerância de 1% nas aproximações , precisamos de 2 ou 3 iterações .

Nosso próximo passo é avaliar quantas vezes o procedimento *Posseidon* é disparado pelos procedimentos de cálculo do gradiente GradJ(.) e de minimização na direção contrária a esse gradiente para o caso de aproximações no interior do espaço convexo admissível. O procedimento *GradienteJ* executa a tarefa descrita por (4.17), da maneira descrita na seção (3.4), tendo que , no pior caso, disparar *Posseidon* $4 + 2(7 - 1) = 16$ vezes, ou seja, resolvemos para cada uma das $(2np + 2)$ componentes de GradJ(.) $16 \times (7m)$ sistemas lineares. Certamente não precisaremos das 16 chamadas para conseguir a convergência, um número médio depende muito do valor do parâmetro de tolerância *tol*, de definirá quando a aproximação já é satisfatória. Por exemplo, para $tol = .01(1\%)$, em média são necessárias 3 iterações , ou seja, $4 + 2(3 - 1) = 8$ chamadas.

Em termos práticos, a complexidade do algoritmo *Pegasus* se confunde com a complexidade do procedimento do cálculo do gradiente segundo (4.17); todos os outros procedimentos são de relevância desprezível em termos de quantidade de processamento. Podemos então contabilizar como complexidade a solução de no máximo $16(2np+2)(7m)$ sistemas lineares algébricos pentadiagonais $(2n \times 2n)$.

Para valores típicos $(m = 40, np = 12, n = 25, tol = 1\%)$, teríamos em média
$$8(26)(2 \times 40) = 16.640 \text{ sistemas lineares } (50 \times 50);$$

Para os valores $(m = 40, np = 16, n = 25, tol = 1\%)$, teríamos em média
$$6(34)(2 \times 40) = 16320 \text{ sistemas lineares } (50 \times 50);$$

Para valores um pouco maiores $(m = 40, np = 20, n = 25, tol = 1\%)$, teríamos
$$6(42)(2 \times 40) = 20.160 \text{ sistemas lineares } (50 \times 50);$$

Para uma malha grande $(m = 60, np = 40, n = 40, tol = 1\%)$, teríamos
$$6(82)(2 \times 60) = 59.040 \text{ sistemas lineares } (80 \times 80);$$

Referentemente às estimativas acima, a simulação computacional apontou, respectivamente, para os valores médios
$$14.900, 15.500, 21.000, 67.000$$

Dando uma idéia da complexidade de nossa tarefa, observamos que para gerar os resultados apresentados a seguir, precisamos resolver, a cada iteração do algoritmo *Pegasus*, entre 15.000 e 21.000 sistemas lineares pentadiagonais (50×50).

5.2 A implementação numérica.

5.2.1 As demandas de hardware e software.

Nosso algoritmo, apresentado na seção (4.3), foi inteiramente implementado em FORTRAN77 e executado em estações de trabalho SUN e DEC alfa. Uma vez delineado o algoritmo de solução , saímos a procura dos dois módulos de processamento de mais alto desempenho:

• rotina numérica em Fortran implementando a resolução de sistemas lineares algébricos pentadiagonais;

• rotina numérica em Fortran implementando a resolução de sistemas lineares algébricos simétricos por fatoração LDL^T.

Uma vez incorporados tais módulos, passamos a testar o algoritmo sob sucessivas execuções , e então ficou evidente sua baixa velocidade de convergência, característica da classe de problemas na qual nos situamos.

O equipamento DEC alfa, por ser o mais potente disponível, sitiou grande parte das execuções do programa, sobretudo aquelas que, pelo tamanho da malha da discretização ou pelo refinamento da interpolação , projetavam a resolução de mais de 20000 sistemas lineares por iteração , e uma vez que tais execuções se mostravam excessivamente lentas nas SUN's.

5.2.2 As principais dificuldades numéricas.

As dificuldades numéricas adviram em muitas situações e demandaram constante preocupação com imprecisões , o que, via de regra, se traduziu por perda de desempenho, uma vez que fizemos muitos testes e procedimentos corretivos. Como principais fontes de erro, ao menos como aquelas que representaram risco para a continuidade e mesmo para a convergência do algoritmo, podemos citar o cálculo de raízes de equações de segundo grau e a projeção vetorial sobre hiperplanos.

O cálculo de raízes ao qual nos referimos foi necessário sobretudo para o subalgoritmo de procura de máximos locais; também foi fundamental na construção dos vetores normais da superfície de restrições . Em muitas situações , pequenas imprecisões fizeram com que máximos locais trocassem de intervalo, passando a ser desconsiderados e então conduzindo nossa sequência de aproximações para fora do espaço admissível; tal também repercutiu na determinação dos normais de superfície segundo as equações da seção (3.1d).

Na projeção vetorial sobre hiperplanos, tarefa descrita na seção (3.5), dificuldades sobrevieram sobretudo devido à imprecisão no cálculo de produtos internos. Em algumas situações , o vetor de calculado tinha até mesmo uma orientação não coincidente com tais hiperplanos; no caso em que a fronteira, localmente, era linear, as aproximações do nosso algoritmo seriam conduzidas então para o exterior da região admissível ao invés de permanecer na fronteira.

5.3 Alguns resultados e evidências computacionais.

Apresentamos agora alguns resultados obtidos na simulações de nosso código em FORTRAN no qual está implementado o algoritmos *Pegasus*, descrito na subseção (4.3.1).

Para simulação , definimos 3 baterias de teste, que logo serão explicitadas; em todas elas, a função de fluxo de massa $w(t)$ foi tomada como identicamente unitária. Certamente a forma dessa função é decisiva para a determinação do controle ótimo, entretanto, não é objetivo deste trabalho estudar a dependência entre essas duas funções , muito embora o programa tenha sido elaborado para tomar como input uma função $w(t)$ positiva qualquer. Foi deixado de lado, porisso, um problema matematica e computacionalmente interessante, mas que poderia dificultar ainda mais nossa tarefa no presente momento.

A seguir, segue-se uma descrição das baterias de teste. Apresentamos também os gráficos do campo de temperaturas $T(x,t)$ e do campo de concentrações $u(x,t)$ correspondentes a aproximação inicial em cada bateria. Chamaremos de perfis as sucessivas aproximações do controle ótimo $S^*(t)$.

Nos gráficos das figuras (5.4),(5.6),(5.8) a seguir, traçamos a temperatura $T(x_j,t_k)$ para cada nodo (x_j,t_k) da malha.

Nos gráficos das figuras (5.5),(5.7),(5.9) a seguir, traçamos a concentração $u(x_j,t_k)$ para cada nodo (x_j,t_k) da malha.

Bateria 1 (executada nas estações SUN):

- $t_0 = 1, L = 1$: (definem nosso domínio espaço-tempo);
- $C = 1$:(calor específico da parede do catalizador);
- $u_0 = 30$: (concentração inicial do poluente longo do cano de descarga);
- $n = 10, m = 40$: (definem os parâmetros da discretização numérica);
- $np = 12$: (usaremos interpolação cúbica no [0,1] em 12 segmentos);
- $\mathcal{A} = \{S(t) \in C^1([0,1]) : S(t) \leq N = 89, \int_0^1 S(t)dt \leq M = 70\}$;
- $S_0(t)$, o perfil inicial da simulação , corresponde à curva $S_0(t)$ vs t abaixo (Fig 5.1):

Bateria 2 (executada no DEC alfa):

- $t_0 = 1, L = 1$: (definem nosso domínio espaço-tempo);
- $C = 1$:(calor específico da parede do catalizador);
- $u_0 = 30$: (concentração inicial do poluente ao longo do cano de descarga);
- $n = 25, m = 40$: (definem os parâmetros da discretização numérica);
- $np = 16$: (usaremos interpolação cúbica no [0,1] em 16 segmentos);
- $\mathcal{A} = \{S(t) \in C^1([0,1]) : S(t) \leq N = 89, \int_0^1 S(t)dt \leq M = 70\}$;
- $S_0(t)$, o perfil inicial da simulação , corresponde à curva $S_0(t)$ vs t abaixo (Fig 5.2):

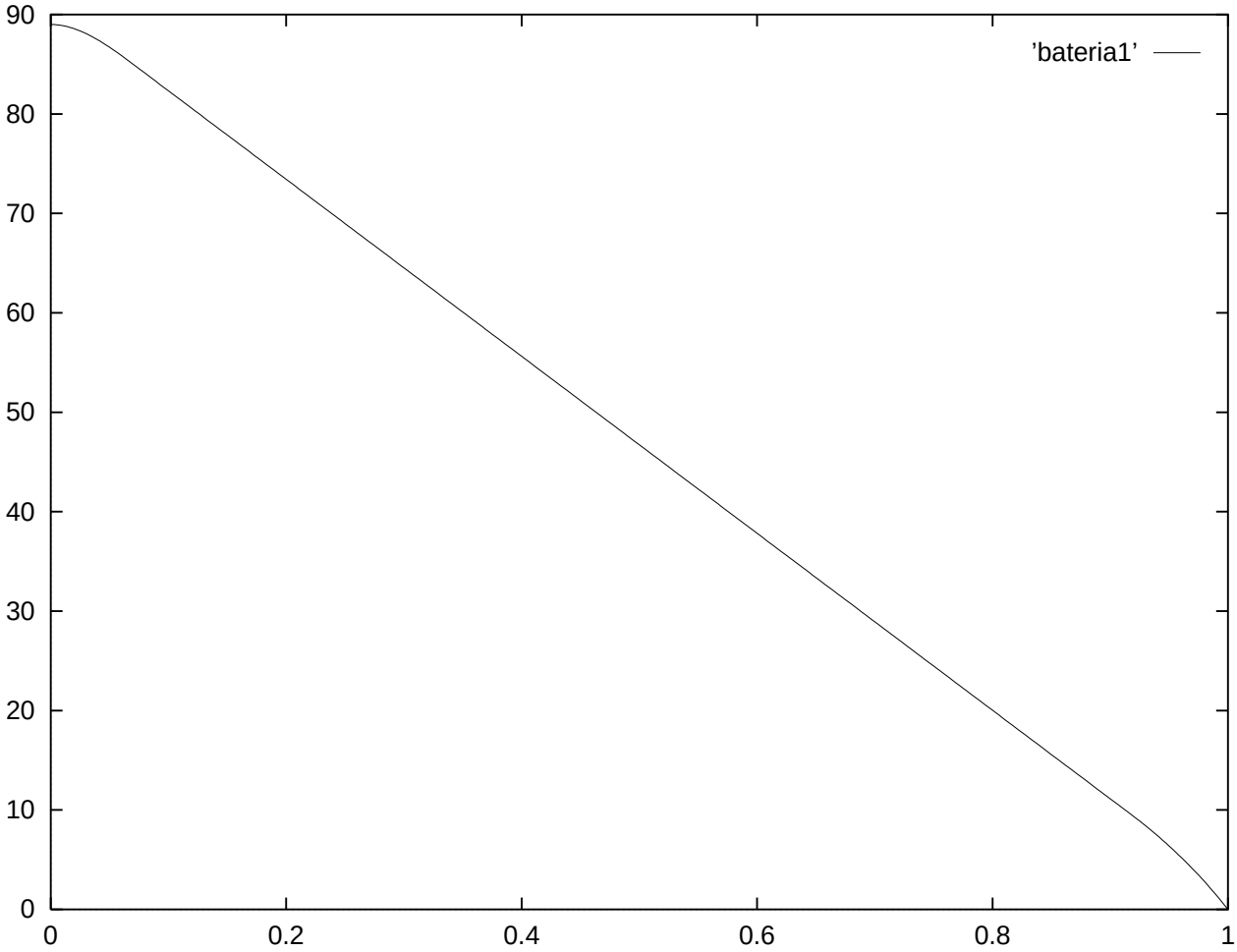

Figura 5.1: Perfil inicial para a bateria 1.

Bateria 3 (executada no DEC alfa):

- $t_0 = 1, L = 1$: (definem nosso domínio espaço-tempo);
- $C = 1$:(calor específico da parede do catalizador);
- $u_0 = 30$: (concentração inicial do poluente ao longo do cano de descarga);
- $n = 25, m = 40$: (definem os parâmetros da discretização numérica);
- $np = 20$: (usaremos interpolação cúbica no [0,1] em 20 segmentos);
- $\mathcal{A} = \{S(t) \in C^1([0,1]) : S(t) \leq N = 89, \int_0^1 S(t)dt \leq M = 70\}$;
- $S_0(t)$, o perfil inicial da simulação , corresponde à curva $S_0(t)$ vs t abaixo (Fig 5.3):

Para cada bateria de simulações definida acima, os respectivos perfis $S_k(t)$ traçados a seguir, nas figuras (5.11) - (5.13), em gráficos $S(t)$ vs t são evoluções não sucessivas do perfil inicial $S_0(t)$, e mostram, da esquerda para a direita, basicamente a evolução das aproximações por um caminho interior à fronteira essencial que deverá definir nossa solução , a fronteira correspondente à restrição energética :

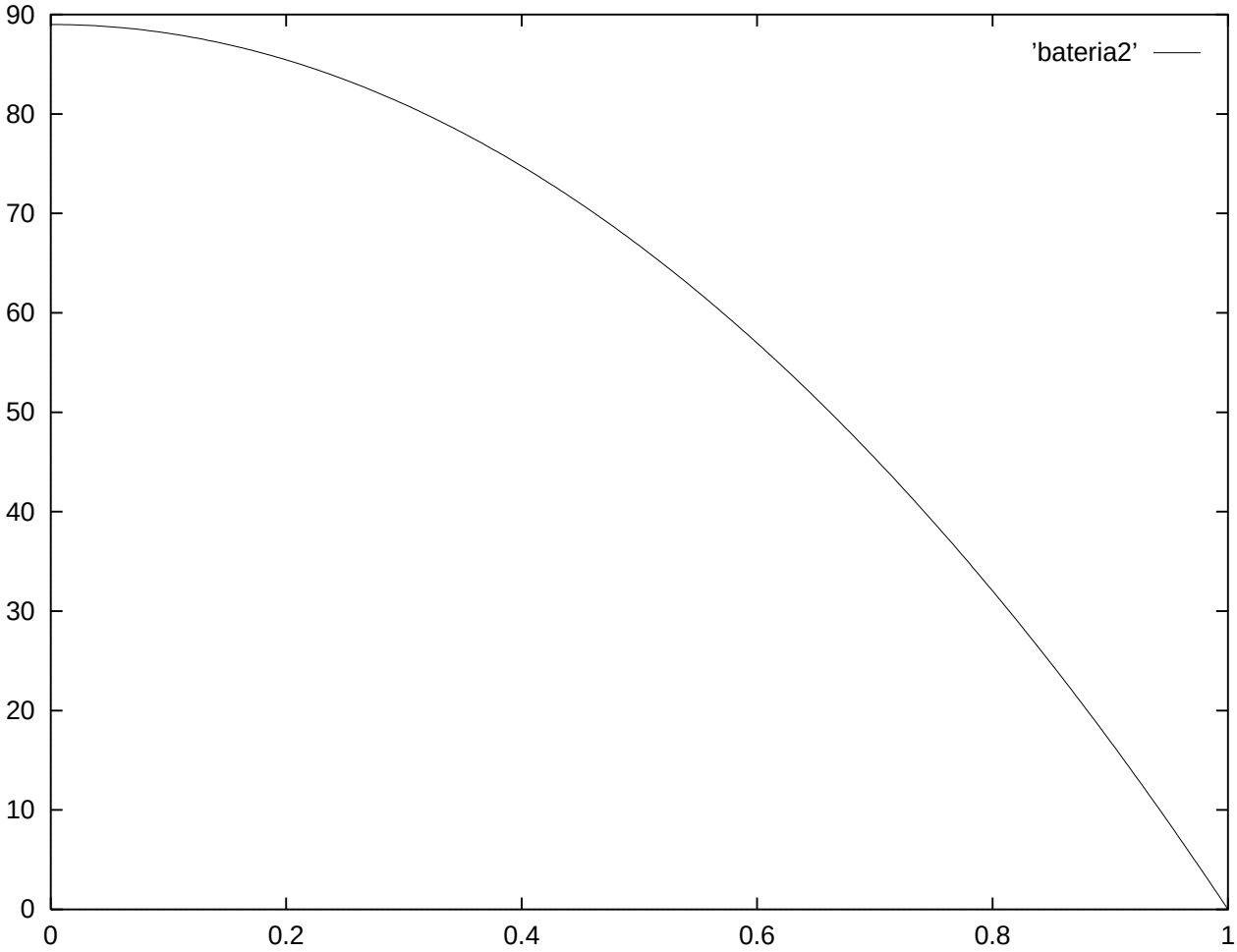

Figura 5.2: Perfil inicial para a bateria 2.

$$\int_0^1 S(t)dt \leq M = 70,$$

uma vez que os gradientes do funcional de concentrações sempre indicarão o acréscimo

dos valores da função de controle, e então, se não houvesse a restrição acima (por exemplo, se houvesse energia suficiente para manter as temperaturas no máximo valor admissível durante todo o intervalo de aquecimento), a solução de nosso problema, cujo gráfico é mostrado na figura ao lado, estaria determinada pela expressão

$$S^*(t) \equiv N = 89 \qquad , 0 \leq t \leq t_0 = 1.$$

Assim a restrição essencial do problema é a restrição energética, que determinará a área M entre a função de controle $S^*(t)$ e os eixos coordenados, conforme podemos visualizar na figura (5.10). A forma da função de controle ótimo $S^*(t)$, que dependerá também do gradiente de concentrações e que traduz-se na maneira ótima de gastar a energia disponível M, para fins de nossa ação controladora, é determinada pelo procedimento de minimização apresentado no capítulo anterior.

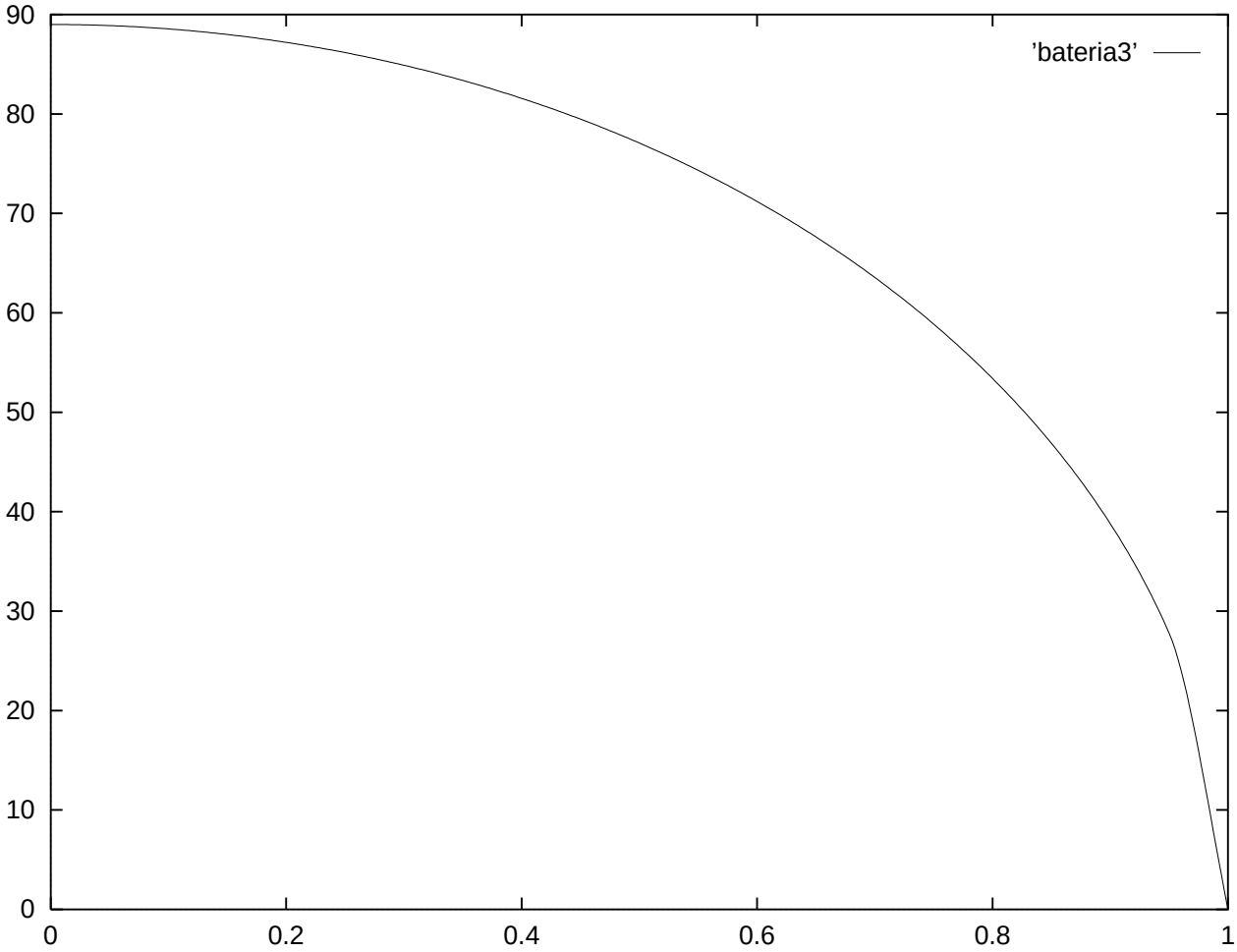

Figura 5.3: Perfil inicial para a bateria 3.

Relativamente às figuras (5.11)-(5.13) a seguir, a tabela ao lado mostra a evolução dos valores do funcional de concentrações para as aproximações traçadas.

FIGURA	PERFIL	FUNCIONAL
figura (5.11)	1	11.9860960
	2	11.9743422
	3	11.9394637
	4	11.9197847
	5	11.8978787
	6	11.8970780
figura (5.12)	1	11.9276619
	2	11.9209873
	3	11.9104787
	4	11.9042709
	5	11.8978787
	6	11.8972256
	7	11.8971140
figura (5.13)	1	15.3060081
	2	15.3028053
	3	15.3028040
	4	15.3028034

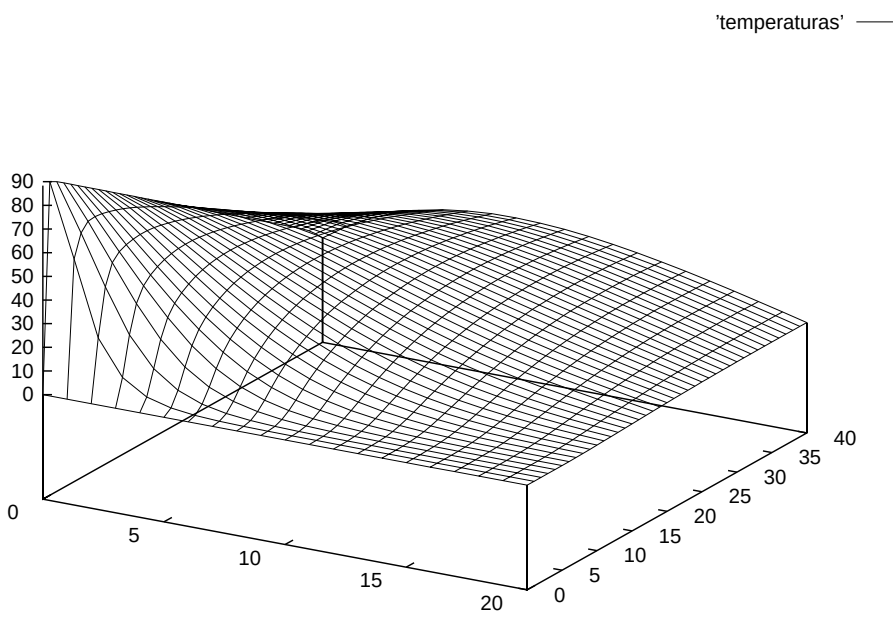

Figura 5.4: Campo de temperaturas para o perfil inicial da bateria 1.

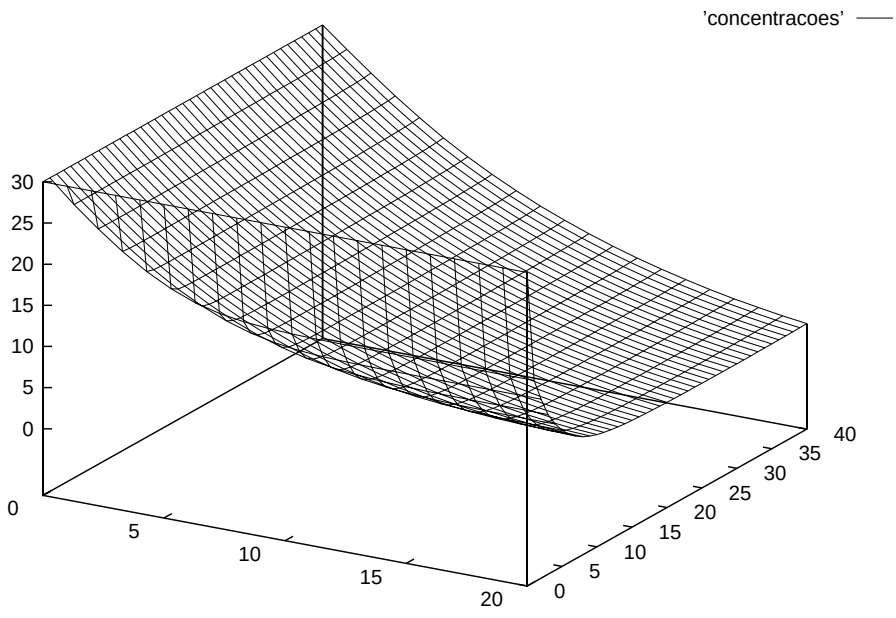

Figura 5.5: Campo de concentrações para o perfil inicial da bateria 1.

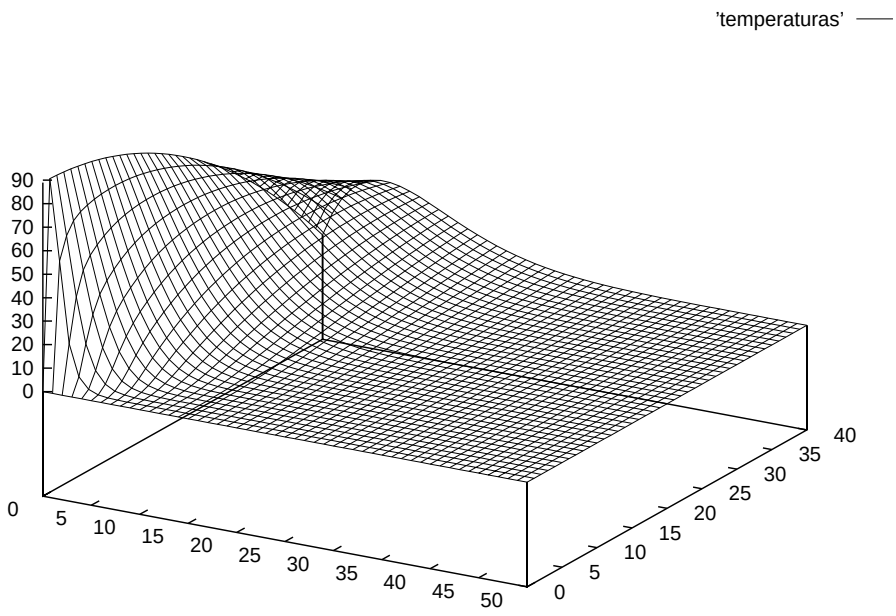

Figura 5.6: Campo de temperaturas para o perfil inicial da bateria 2.

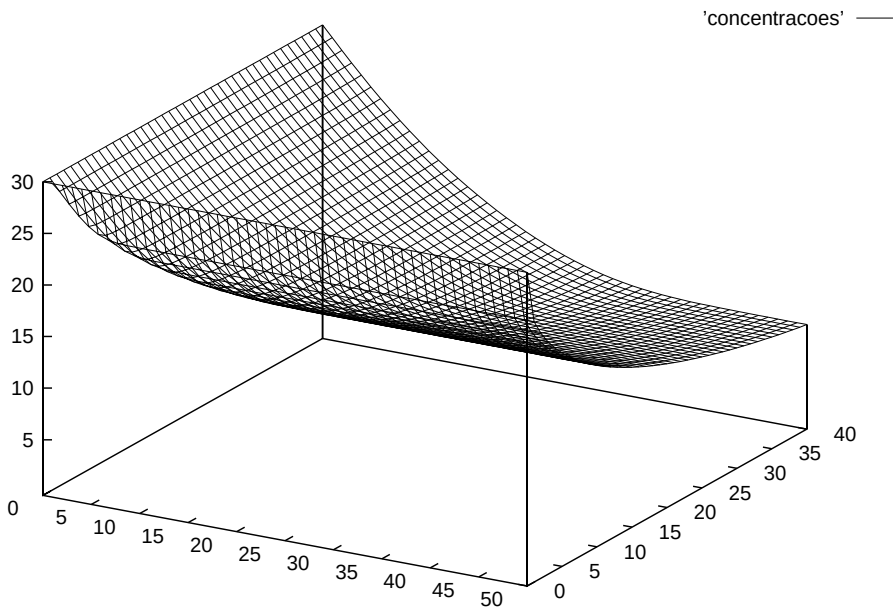

Figura 5.7: Campo de concentrações para o perfil inicial da bateria 2.

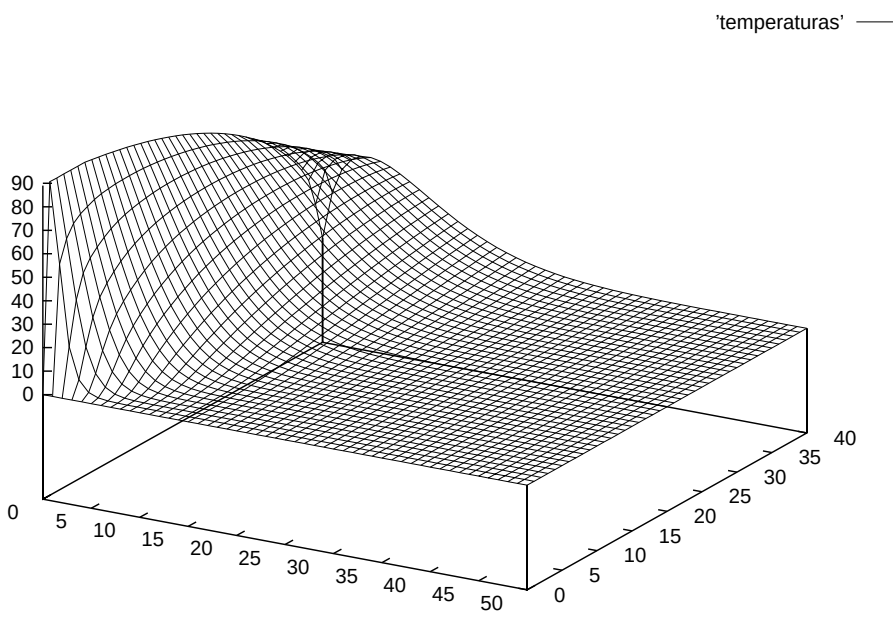

Figura 5.8: Campo de temperaturas para o perfil inicial da bateria 3.

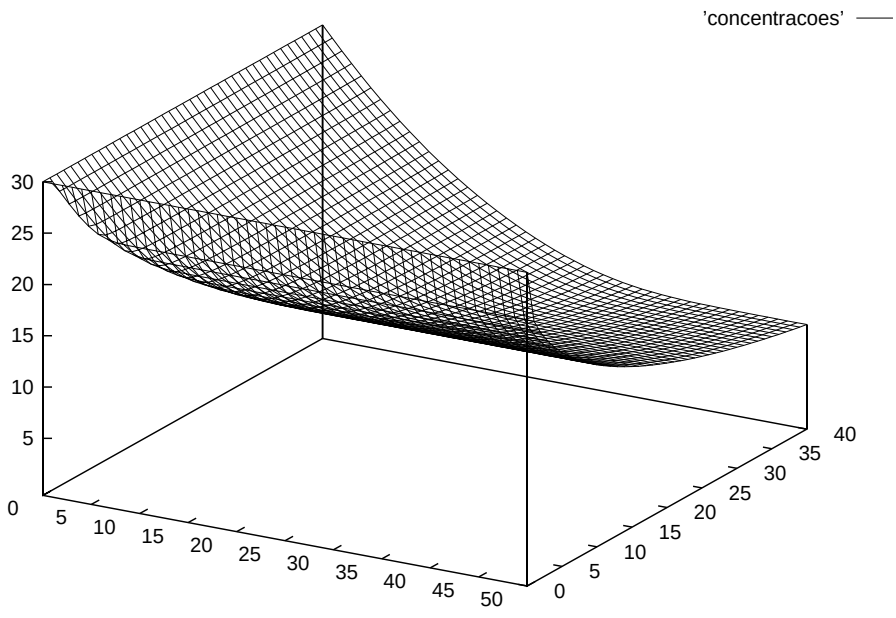

Figura 5.9: Campo de concentrações para o perfil inicial da bateria 3.

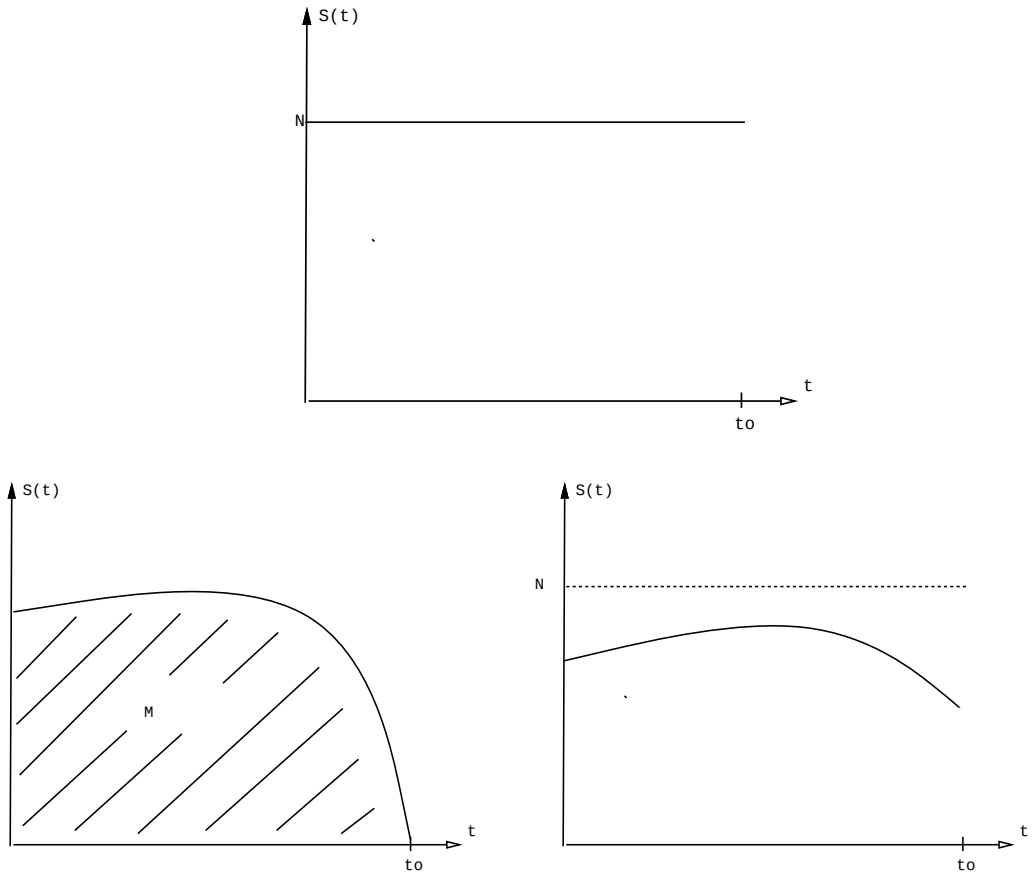

Figura 5.10: Restrições energética (esq) e mecânica (dir).

A execução do programa tem-se mostrado bastante lenta devido à grande quantidade de processamento envolvido e à baixa velocidade de convergência verificada nas simulações numéricas; o que evidencia que os resultados apresentados dão apenas uma idéia do que deva ser a solução ótima, apesar do grande esforço computacional já dispendido em nossas simulações nas workstations SUN e DEC alfa 3000.

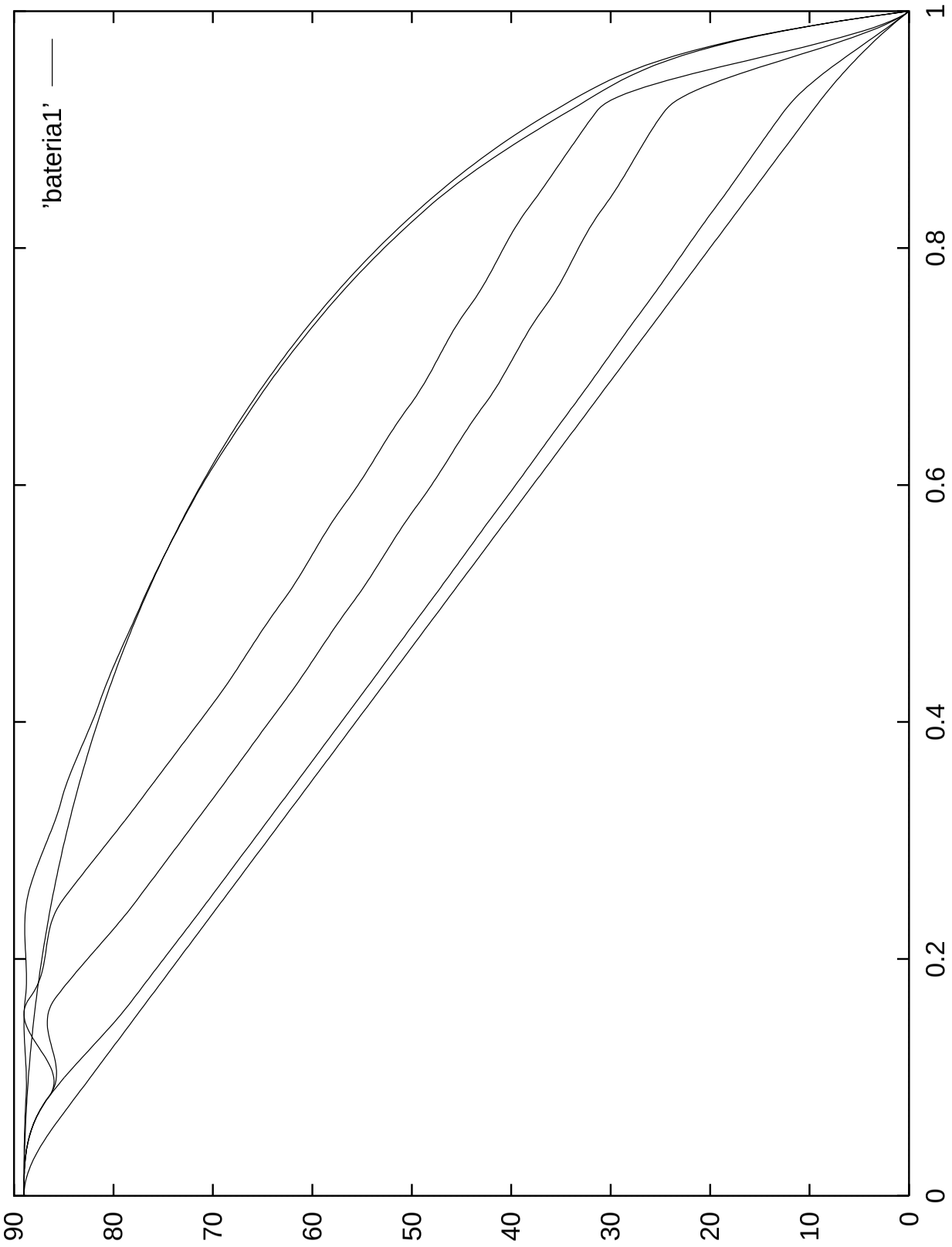

Figura 5.11: Evolução das aproximações para bateria 1.

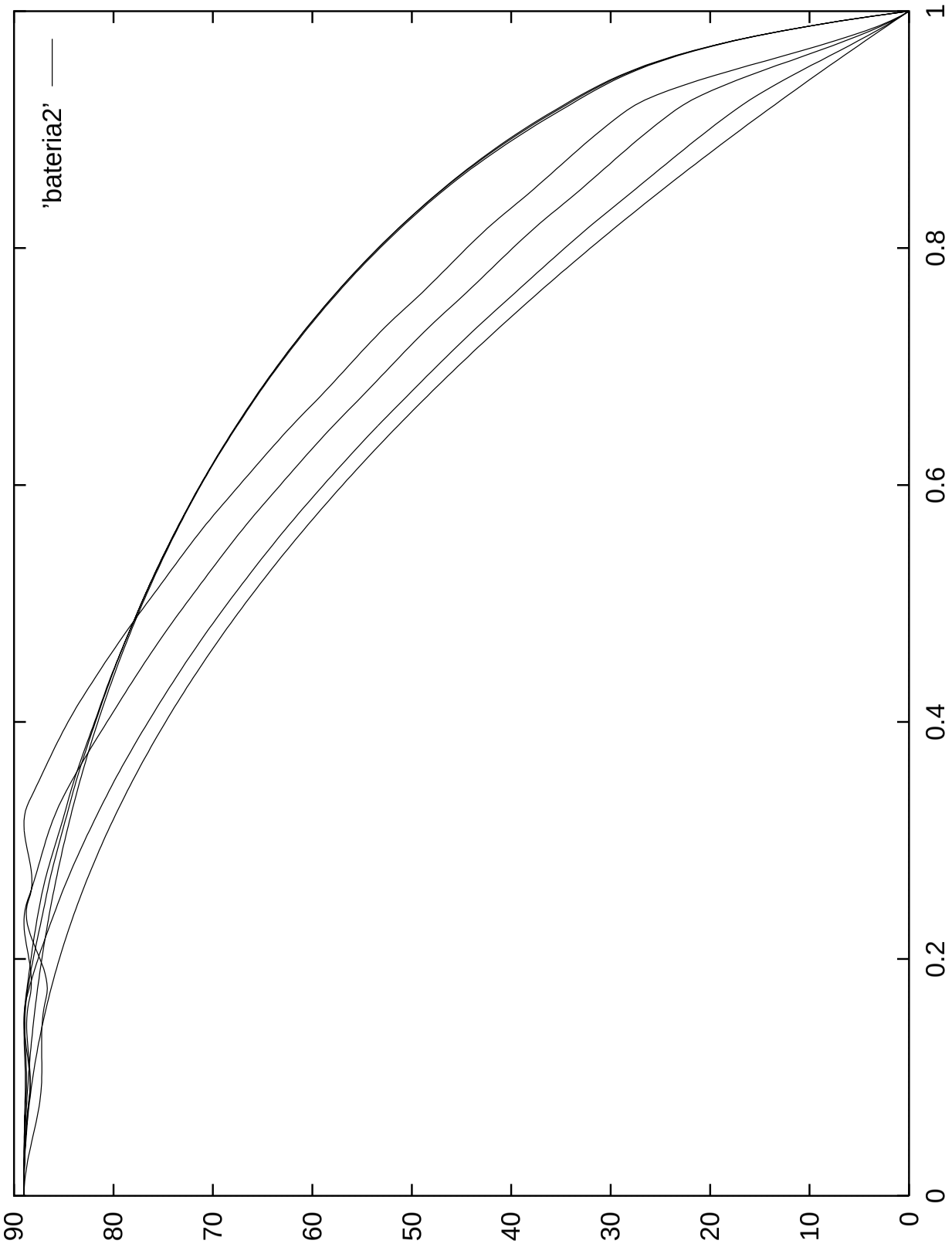

Figura 5.12: Evolução das aproximações para bateria 2.

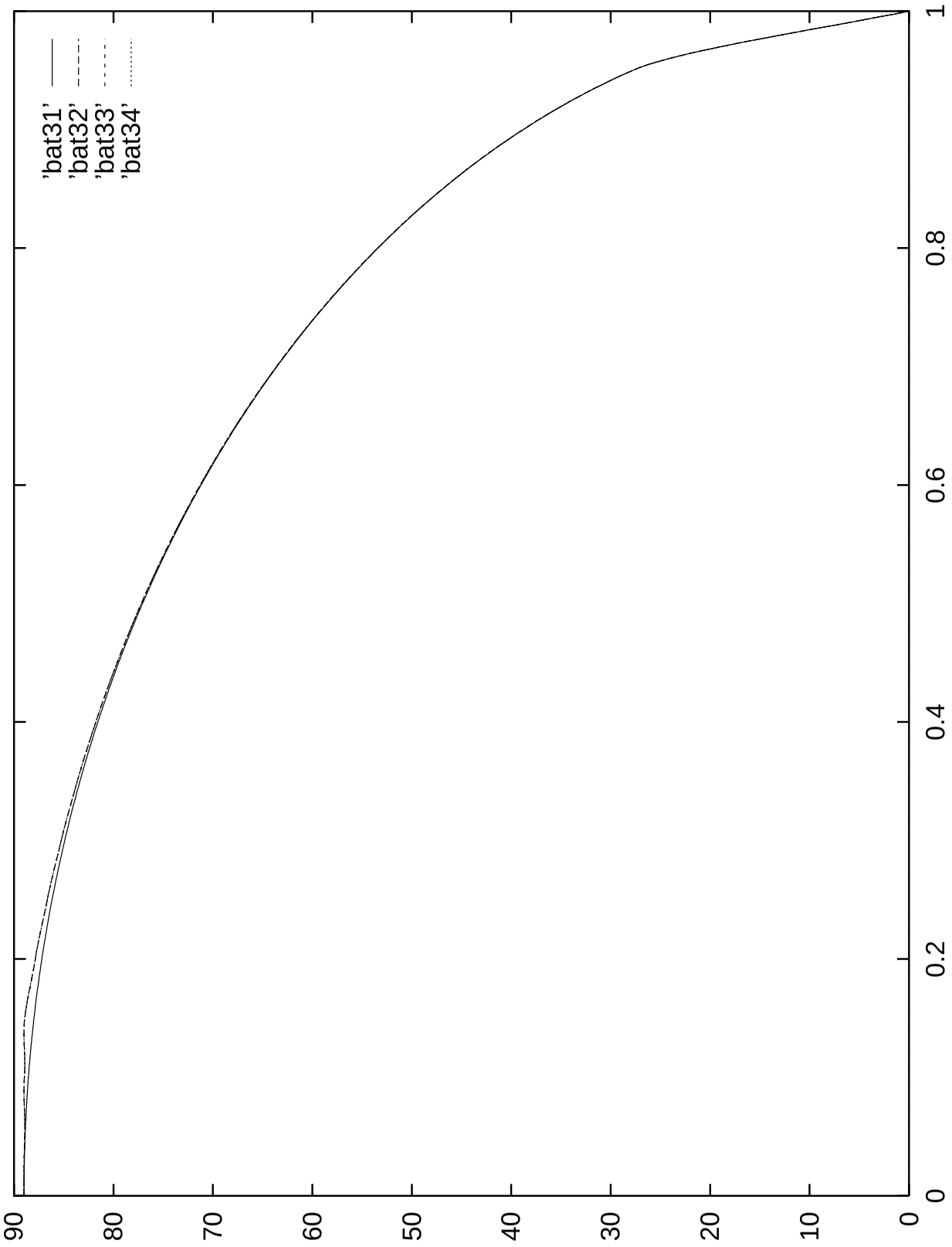

Figura 5.13: Evolução das aproximações para bateria 3.

6 CONCLUSÕES

A implementação numérica e a simulação computacional amparam as seguintes conclusões :

1. O Método do Gradiente Projetado, apesar de não termos uma expresão explícita para a função objetiva, e consequentemente para seu vetor gradiente, mostrou-se satisfatoriamente implementado; e as dificuldades que surgiram e que realmente dificultaram a progressão das simulações foram as de natureza numérica.

2. Certamente a técnica de solução dos sub-problemas com valores na fronteira influenciou muito o comportamento do programa, uma vez que a discretização por diferenças finitas é naturalmente menos precisa na fronteira da malha, e tal imprecisão obviamente afetou a determinação do gradiente da função objetiva. Assim, aperfeiçoar a técnica de solução (elementos finitos, por exemplo) certamente é uma necessidade de nosso algoritmo de busca ao controle ótimo.

3. Também o algoritmo global de solução pode ser aperfeiçoado, sempre com o objetivo de contornar eficientemente certas dificuldades oriundas da não linearidade do problema, possibilitando então uma convergência mais rápida, e obviamente sem pretensões revolucionárias dada a natureza da tarefa em questão .

4. Os perfis mostrados nas figuras (5.11)-(5.13) são fortes evidências computacionais do quê deveríamos esperar da solução do problema de otimização ; são coerentes com a solução exata e explícita do problema simplificado definido e resolvido na subseção (2.3.1), e certamente não estão muito distantes da solução que procuramos.

5. A análise do modelo feita no trabalho, ainda que limitada, evidencia a necessidade de um trabalho teórico mais profundo.

6. Por todas as conclusões apresentadas anteriormente, o presente trabalho, desde a formulação , análise, algoritmização , implementação e simulação , evidencia-se altamente bem sucedido frente à realização seus objetivos.

ANEXO A-1 APÊNDICE.

A-1.1 A teoria das Equações de Hammerstein Abstratas.

Nesta seção , desenvolveremos os pré-requisitos fundamentais para o desenvolvimento da teoria das equações integrais de Hammerstein

$$u(x) = \int_M G(x,y)f(u(y))dy, \qquad u \in Y.$$

Para maiores esclarecimentos, sugerimos a bibliografia:

H.Zeidler.Nonlinear Functional Analysis and its Applications.Springer-Verlag,ch7,VolI,(1985).

Espaços de Banach Ordenados.

Introduziremos uma relação de desigualdade para espaços de Banach que poderá ser usada analogamente à relação de desigualdade para números reais. A terminologia será escolhida de maneira a ressaltar a analogia com números reais e funções reais.

A-1.1.1 Definição

Seja X um espaço de Banach e seja K um subconjunto de X. Então K é chamado um *cone de ordem* se e somente se:

1. K é fechado, não vazio e $K \neq \{0\}$;
2. $a, b \in \Re, a, b \geq 0, x, y \in K \Rightarrow ax + by \in K$;
3. $x \in K$ e $-x \in K \Rightarrow x = 0$.

Podemos definir então

$$\begin{gathered} x \leq y \text{ se e somente se } y - x \in K; \\ x < y \text{ se e somente se } x \leq y \text{ e } x \neq y; \\ x \ll y \text{ se e somente se } y - x \in int(K); \\ x \not\leq y \text{ se e somente se } x \leq y \text{ é falso}; \\ [x,y] = \{z \in X : x \leq z \leq y\} \text{ (intervalo de ordem).} \end{gathered}$$

A condição (2) é equivalente a estabelecer que K é convexo e que se $x \in K$ e $a > 0$, então $ax \in K$. O cone de ordem K é chamado *gerador (total)* se e somente se $X = ger(K)(X = \overline{ger(K)})$. Em outras palavras, K é gerador (total) se e somente se $X = (K - K)(X = \overline{K - K})$. Observamos que $K - K = \{x - y : x, y \in K\}$. Por um espaço de Banach ordenado entenderemos um espaço de Banach associado a um cone de ordem.

Não podemos fazer confusão entre as noções de cone e cone de ordem. Um subconjunto C do espaço de Banach X é chamado um *cone* se e somente se $x \in C$ e $a > 0$ implicam $ax \in C$. Todo cone de ordem é um cone, mas não vale a recíproca.

Exemplo

Seja $X = C(\overline{M})$ para alguma região limitada $M \in \Re^N$. Definimos $C_+(\overline{M}) = \{f \in C(\overline{M}) : f(x) \geq 0 \text{ em } \overline{M}\ \}$. Então $K = C_+(\overline{M})$ é um cone de ordem em X e temos

$$f \leq g \text{ s.s.s. } f(x) \leq g(x) \text{ para todo } x \in \overline{M};$$

$$f \ll g \text{ s.s.s. } f(x) < g(x) \text{ para todo } x \in \overline{M}.$$

A-1.1.2 Proposição (Propriedades de $\leq$)

Para todos $u, x, x_n, y, y_n, z \in X$ e todos $a, b \in \Re$ temos

$$x \leq x$$
$$x \leq y \& y \leq x \Rightarrow x = y$$
$$x \leq y \& y \leq z \Rightarrow x \leq z$$

Além disso, temos

$$x \leq y \& 0 \leq a \leq b \Rightarrow ax \leq by$$
$$x \leq y \& u \leq z \Rightarrow x + u \leq y + z$$
$$x_n \leq y_n \quad \forall n \Rightarrow \lim_{n\to\infty} x_n \leq \lim_{n\to\infty} y_n.$$

Para o operador relacional $\ll$ temos

$$x \ll y \& y \ll z \Rightarrow x \ll z$$
$$x \ll y \& y \leq z \Rightarrow x \ll z$$
$$x \leq y \& y \ll z \Rightarrow x \ll z$$
$$x \ll y \& a > 0 \Rightarrow ax \ll ay.$$

Prova:

Basta usar a definição (A-1.1.1) e as propriedades de K. Por exemplo, se $x_n \to x$ e $y_n \to y$ ao $n \to \infty$, o fato que K é fechado implica que

$$x_n \leq y_n \Rightarrow y_n - x_n \in K \Rightarrow y - x \in K \Rightarrow x \leq y$$

Operadores Monótonos crescentes

A-1.1.3 Definição

(1) O cone de ordem K é chamado *normal* se e somente se existe um número $c > 0$ tal que, para todo $x, y \in X : 0 \leq x \leq y \Rightarrow \|x\| \leq c\|y\|$.

(2) O operador $T : D(T) \subseteq X \longrightarrow Y$ é chamado *monótono crescente* se e somente se é verdade para todo $x, y \in D(T)$ que $x < y$ implica $Tx \leq Ty$. O operador é chamado *estritamente* ou *fortemente* monótono crescente se e somente se o símbolo $\leq$ é trocado por $<$ ou $\ll$ respectivamente. Similarmente, definimos operadores que são estritamente ou fortemente monótonos decrescentes.

(3) O operador T é chamado positivo se e somente se $T(0) \geq 0$ e

$$x > 0 \text{ implica } Tx \geq 0 \text{ para todo } x \in D(T).$$

Como antes, o operador é estritamente ou fortemente positivo se o símbolo $\geq$ é trocado por $>$ ou $\gg$ respectivamente.

(4) O operador linear $T : D(T) \subseteq X \longrightarrow Y$ é chamado *e-positivo* se e somente se existe um elemento $e > 0$ e para todo $x \in D(T)$ existem números positivos $\alpha(x)$ e $\beta(x)$ tais que

$$\alpha(x)e \leq Tx \leq \beta(x)e.$$

No exemplo acima supomos que o cone de ordem que gera o operador $\leq$ existe.

Exemplo

(1) Seja $X = \Re, K = \Re_+$.Então para funções reais $T : D(T) \subseteq \Re \longrightarrow \Re$ os conceitos de (estritamente) monótono crescente (decrescente) acima coincidem com as definições usuais em $\Re$.

(2) Para o operador linear T, os conceitos de fortemente (estritamente) positivo são equivalentes aos de fortemente (estritamente) monótono crescente.

$$T \text{ positivo: } x \leq y \Rightarrow 0 \leq y - x \Rightarrow 0 \leq T(y - x) \Rightarrow Tx \leq Ty;$$
$$T \text{ monótono crescente: } x \leq y \Rightarrow Tx \leq Ty \Rightarrow y - x \geq 0, Ty - Tx = T(y - x) \geq 0.$$

(3) Seja $X = C(\overline{M})$ e $K = C_+(\overline{M}) = \{f \in C(\overline{M}) : f \geq 0 \text{ em } \overline{M}\}$. Uma vez que $\|f\| = \max_{x \in \overline{M}} |f(x)|$, segue de $f, g \in X$ com $0 \leq f \leq g$ que $\|f\| \leq \|g\|$,ou seja, que o cone de ordem é normal. Consideraremos agora o operador integral linear $u = Tv$ onde

$$u(t) = \int_M A(t,s)v(s)ds.$$

O núcleo $A : \overline{M} \longrightarrow \Re$ é contínuo. Então $T : X \longrightarrow X$ é um operador linear compacto. Mais do que isso, T é positivo, e então monótono crescente, se $A(t,s) \geq 0$ para todo $t, s \in \overline{M}$. Suponha ainda que temos a condição mais forte $A(t,s) > 0 \quad \forall t, s \in \overline{M}$. Fazemos $e(s) \equiv 1$. Então T é fortemente positivo e e-positivo.

Prova:

Seja $A(t,s) > 0$. De $v > 0$ em X segue $v(s_0) > 0$ para algum $s_0 \in M$. Pela continuidade de v, existe uma vizinhança $U(s_0)$ limitada de medida $m(U(s_0))$ e algum número $c > 0$ tal que $A(t,s) \geq c$ em $M \times M$ e $v(s) \geq c$ em $U(s_0)$. Então $u(t) \geq c^2 \cdot m(U(s_0))$ e

$$u(t) \leq m(M) \max_{t,s \in \overline{M}} |A(t,s)| \cdot \max_{s \in \overline{M}} |v(s)| \qquad \forall t \in M.$$

Ou seja, $0 < c_1.e \leq u \leq c_2.e$ e daí temos $u \gg 0$. m(M) é a medida de Lebesgue.

Contraexemplo. Nem todo cone de ordem é normal.

Seja $X = C^1([0,1])$ e $K = \{f \in X : f(t) \geq 0 \text{ em } [0,1] \}$ e

$$\|f\| = \max_{t \in [0,1]} |f(t)| + \max_{t \in [0,1]} |f'(t)|.$$

Agora $0 \leq f \leq g$ significa $0 \leq f(t) \leq g(t) \qquad \forall t \in [0,1]$. Uma vez que a última não contém informação sobre as derivadas, não existem constantes $c > 0$ que garantam que $0 \leq f \leq g$ implica $\|f\| \leq c\|g\|$.

A importância de cones normais para a convergência de métodos iterativos será decisiva para nossos propósitos e se tornará clara nas subseções seguintes, onde usaremos o seguinte resultado.

A-1.1.4 Proposição

Se o cone de ordem é normal então todo intervalo de ordem $[x, y]$ é limitado.

Prova:

Se $x \leq w \leq y$ então $0 \leq w - x \leq y - x$ e então $\|w - x\| \leq c\|y - x\|$.

A-1.1.5 Definição

Seja X o espaço de Banach real com cone de ordem K e seja $e > 0$. Fixamos

$$X_e = \{x \in X : \text{existe real } a > 0 \text{ tal que } -ae \leq x \leq ae\},$$
$$\|x\|_e = inf\{a > 0 : -ae \leq x \leq ae\};$$
$$K_e = K \bigcap X_e.$$

A-1.1.6 Proposição

Se K é normal, então

(1) O conjunto X_e com norma $\|.\|_e$ forma um espaço de Banach. A inclusão $X_e \subseteq X$ é contínua. Se $e \gg 0$ então $X = X_e$ e as normas sobre X e X_e são equivalentes.

(2) O conjunto K_e é um cone de ordem normal em X_e com $e \in int(K_e)$.

(3) Se para $x \in X$ existem números positivos α, β tais que $\alpha e \leq x \leq \beta e$ em X, então $x \gg 0$ em X_e.

(4) Se o operador linear $T : X \longrightarrow X$ é e-positivo e se $T(X) \subseteq X_e$, então $T : X \longrightarrow X_e$ é fortemente positivo.

É digno de se mencionar que K_e tem um ponto interior mesmo quando K não tem, e que podemos obter operadores fortemente positivos.

Prova de (1):

(I) Pela proposição (A-1.1.2),$\|.\|$ é um norma. Em particular, $\|x\|_e = 0$ imediatamente implica que $-ae \leq x \leq ae$ para todo $a > 0$, fazemos $a \to 0$ e segue $x = 0$.

(II) $\|x\|_e = 1$ implica $x \in [-e, e]$, isto é,$\|x\| < r$ com r fixo, pela proposição (A-1.1.4). Portanto ,$\|x\| \leq r\|x\|_e$ para todo $x \in X_e$, e a inclusão é contínua.

(III) X_e é um espaço de Banach. Na verdade, toda sequência de Cauchy (x_n) em X_e é também uma sequência de Cauchy em X, e então $x_n \to x$ em X ao $n \to \infty$. Agora $\|x_n - x_m\|_e < \epsilon$

para $n, m \geq n(\epsilon)$ implica $-\epsilon e \leq x_n - x_m \leq \epsilon e$. Ao $m \to \infty$, obtemos $-\epsilon e \leq x_n - x \leq \epsilon e$ para $n > n(\epsilon)$. Portanto $x_n \to x$ em X_e.

(IV) Seja $e \gg 0$ ($e \in int(K)$). Então existe uma bola $U(0,R)$ de raio R tal que
$$e + \overline{U(0,R)} \subseteq K.$$

Se $x \in X$ com $x \neq 0$, então

$$e \pm Rx/\|x\| \in K, \text{ isto é, } e \pm Rx/\|x\| \geq 0$$

e por conseguinte $-\|x\|e/R \leq x \leq \|x\|e/R$. Isto significa que $x \in X_e$ e $\|x\|_e < \|x\|/R$. Mas isto, com (II) implica que $X = X_e$ e as normas $\|.\|$ e $\|.\|_e$ são equivalentes.

Prova de (2):

Mostramos que K_e é fechado. Seja (x_n) uma sequência em K_e com $x_n \to x$ em X_e ao $n \to \infty$. Como $K_e \subseteq K$, a sequência também pertence a K. Por (II), $x_n \to x$ em X ao $n \to \infty$, e então $x \in K$ e daí temos $x \in X \bigcap K = K_e$.

K_e é normal, pois $K_e \subseteq K, x \leq y$ em X_e implica $x \leq y$ em X. E então $\|x\|_e \leq \|y\|_e$.

Finalmente, $e \in int(K_e)$, uma vez que $e + [-e, e] \in K$, onde $[-e, e]$ é a bola unitária em X_e.

Prova de (3):

O intervalo de ordem $[-\alpha e, \alpha e]$ é uma das vizinhanças da origem em X_e. Como $x + y \in K_e \quad \forall y \in [-\alpha e, \alpha e]$, obtemos $x \in int(K_e)$.

Prova de (4):

(4) é uma consequência imediata de (3).

Aplicação a desigualdades integrais.

Como uma primeira aplicação da teoria dos espaços de Banach ordenados, enunciaremos uma proposição que será importantíssima na investigação de processos evolutivos.

A-1.1.7 Proposição

Seja $A : X \longrightarrow X$ um operador contínuo, linear e positivo sobre o espaço de Banach X e com raio espectral $r(A) < 1$. Sejam $x, y, g \in X$.Então
$$x \leq g + Ax, \qquad y = g + Ay \Rightarrow x \leq y.$$

Prova:

Seja $Bx = g + Ax$. Então $x \leq Bx$ implica $x \leq Bx \leq B^2x \leq ... \leq B^nx$. Como $r(A) = r(B) < 1$, a série de Neumann converge, então $\|A^n\| \to 0$ ao $n \to \infty$. O que implica

$$x \leq B^n x = \sum_{k=0}^{n-1} A^k g + A^n x \to (I - A)^{-1} g = y.$$

segue então $x \leq y$.

Supersoluções ,subsoluções , métodos iterativos.

Consideraremos a equação operacional

$$x = Tx \tag{A-1.1}$$

junto com o correspondente método iterativo

$$u_{n+1} = Tu_n, \qquad v_{n+1} = Tv_n. \tag{A-1.2}$$

As seguintes definições são básicas na teora de existência para equações operacionais em espaços de Banach ordenados.

A-1.1.8 Definição

A função x é chamada uma *supersolução* ,supersolução *estrita*, supersolução *forte* de (A-1.1) se e somente se $x \geq Tx, x \gg Tx$, respectivamente. Analogamente definimos *subsoluções* (estritas,fortes). Observamos ainda que supersoluções (subsoluções) não são soluções .

A-1.1.9 Teorema (Métodos Iterativos Monótonos)

Suponha que $T : [u_0, v_0] \subseteq X \longrightarrow X$ é um operador compacto monótono crescente sobre o espaço de Banach real X com cone de ordem normal X_+. Então as seguintes afirmativas valem:

(A) Convergência do método iterativo. Se u_0 é um subsolução de (A-1.1) e se v_0 é uma supersolução de (A-1.1) com $u_0 \leq v_0$, a sequência iterativa (v_n) em (A-1.2) converge a um ponto fixo de T, em termos concretos, ao maior ponto fixo v de T em $[u_0, v_0]$, e (u_n) converge ao menor ponto fixo u de T em $[u_0, v_0]$.

Mais do que isso, temos as estimativas de erro

$$u_n \leq u \leq v \leq v_n \qquad \forall n = 0, 1, 2, ...$$

(B) Estabilidade da solução . Se u_0 é uma subsolução estrita de (A-1.1) (v_0 é uma supersolução estrita), se T é fortemente monótono crescente, e se a derivada $T'(u)(T'(v))$ existe e é um operador fortemente positivo, então o raio espectral $r(T'(u)) \leq 1 (r(T'(v)) \leq 1)$, isto é $u(v)$ é um ponto fixo não expansivo de T.

Pontos fixos não expansivos são também chamados *fracamente estáveis*. É o esperado então que o método iterativo (A-1.2) possa ser usado apenas para construir soluções não expansivas de (A-1.1).

Prova do Teorema

(A) Como $u_0 \leq Tu_0, Tv_0 \leq v_0$ e $u_0 \leq v_0$ juntas implicam que $u_0 \leq u_1 \leq v_0$ e similarmente que

$$u_0 \leq u_1 \leq ... \leq u_n \leq v_n \leq ... \leq v_1 \leq v_0 \text{ para todo } n,$$

segue que (u_n) é monótona não decrescente limitada superiormente e então existe $u^* \in [u_0, v_0]$ tal que $u_n \to u^*$. (v_n) é monótona não crescente limitada inferiormente e então existe $v^* \in [u_0, v_0]$ tal que $v_n \to v^*$, e também $u_n \leq u^* \leq v^* \leq v_n \qquad \forall n.$

(B) Defina $u_{n+1} = Tu_n$ e tome $u = \lim_{n\to\infty} u_n$, que existe pela parte (A). Mostraremos por contradição que $r(T'(u)) \leq 1$ se $u_0 \leq Tu_0$.

Então , assumimos que $r(T'(u)) > 1$. O operador T é fortemente monótono crescente, e segue que $u_0 < Tu_0 = u_1 < Tu_1 = u_2 < ... < u$, o que implica que $Tu_0 \ll Tu$. Pelas propriedades de $\ll$,$u_0 < u_1$ e $u_1 \ll u$ implicam que $u_0 \ll u$ (crucial).

O operador $T'(u)$ é fortemente positivo, então existe $h > 0$ tal que $T(u-h) < u-h$; como a norma de h pode ser escolhida arbitrariamente pequena, podemos supor

$$u_0 < u - h.$$

Agora, u_0 é uma subsolução e $u-h$ é uma supersolução de (A-1.1), pela parte (A) existe um ponto fixo $w \in [u_0, u-h]$ da equação (A-1.1), o que gera um contradição com o fato que u é a menor solução de (A-1.1).

Aplicação (Equação Integral)

Seja M uma região limitada do $\Re^N$. Seja $X = C(\overline{M}), X_+ = C_+(\overline{M})$ e considere a equação integral

$$u(x) = \int_M G(x,y) f(y, u(y)) dy \qquad \forall x \in \overline{M} \tag{A-1.3}$$

com núcleo contínuo e não negativo $G : \overline{M} \times \overline{M} \longrightarrow \Re$ e função $f : \overline{M} \times \Re \longrightarrow \Re$ contínua e monótona crescente em u. Escrevemos a equação integral na forma $u = Tu, u \in X$. Então o operador $T : X \longrightarrow X$ é compacto e monótono crescente.

Definimos subsoluções e supersoluções trocando $=$ por $\leq$ e $\geq$, respectivamente, na equação integral. O teorema (A-1.1.9) implica que se $u_0 \in X$ é uma subsolução e $v_0 \in X$ é uma supersolução com $u_0 \leq v_0$ em $\overline{M}$, então para $n \to \infty$, o método iterativo

$$u_{n+1}(x) = \int_M G(x,y) f(y, u_n(y)) dy \tag{A-1.4}$$

converge uniformemente em $\overline{M}$ a uma solução $u \in X$ da equação integral, com $u_0 \leq u \leq v_0$ em $\overline{M}$. O limite $u \in X$ é a menor solução de (A-1.3). Pelo contrário, se o método iterativo parte de v_0, então obtemos a maior solução de (A-1.3) com $u_0 \leq u \leq v_0$.

A-1.1.10 Lema do Cone

Seja X um espaço de Banach real, com cone de ordem X_+ contendo um ponto interior. Seja $u \gg 0$. Então para toda $v \not\geq 0$ existe um número unicamente determinado $\alpha_u(v) > 0$ tal que

1. $0 \leq \alpha \leq \alpha_u(v)$ implica $u + \alpha v \geq 0$
2. $\alpha > \alpha_u(v)$ implica $u + \alpha v \not\geq 0$.

Uma consequência importante, que nós devemos usar frequentemente, é que

$$u + \alpha v \gg 0 \text{ e } \alpha > 0 \text{ implicam } \alpha < \alpha_u(v).$$

Prova:

(1) Construção de $\alpha_u(v)$. Considere o raio $\rho = \{u + \alpha v : \alpha \geq 0\}$. Para $\alpha \geq 0$ pequeno, temos $u + \alpha v \in X_+$ e para $\alpha \geq \alpha_0$ grande temos $u + \alpha v \notin X_+$. Caso contrário, se $u + nv \in X_+$ para $n \in N$ grande e $(u/n) + v \in X_+$ obteríamos a contradição $v \in X_+$ fazendo $n \to \infty$.

Então $u + \alpha_u(v)v$ é o ponto de intersecção unicamente determinado entre o raio ρ e a fronteira ∂X_+ do cone X_+.

(2) Continuidade de α_u. Para $\epsilon > 0$ existe um $\delta > 0$ tal que $\|v - w\| < \delta$ implica

$$u + (\alpha_u(v) - \epsilon)w \in int(X_+)$$
$$u + (\alpha_u(v) + \epsilon)w \not\in \quad int(X_+)$$

Assim $\|\alpha_u(w) - \alpha_u(v)\| < \epsilon$.

O Teorema Principal para Operadores de Tipo Monótono

Consideraremos operadores T para os quais uma das três relações seguintes vale para todos $x, y \in D(T)$:

$$Tx \leq Ty \Rightarrow x \leq y \ ; \tag{a}$$
$$Tx < Ty \Rightarrow x < y; \tag{b}$$
$$Tx < Ty \Rightarrow x \ll y. \tag{c}$$

A-1.1.11 Definição

Sejam X e Y espaços de Banach reais ordenados e seja $T : D(T) \subseteq X \longrightarrow Y$ um operador contínuo.

(1) T é de *tipo monótono, estritamente monótono, fortemente monótono* se e somente se valem (a),(b) e (c) respectivamente.

(2) T é convexo se e somente se $D(T)$ é convexo e para todos $x, y \in D(T)$ com $x < y$ e todo $t \in (0, 1)$,

$$T(tx + (1-t)y) \leq tTx + (1-t)Ty.$$

T é concavo se e somente se $-T$ é convexo.

(3) T é *sublinear* se e somente se $T(0) \geq 0$ e $\quad \forall x \in D(T) - 0, \quad \forall t \in (0, 1)$
$$tTx \leq T(tx)$$

T é *superlinear* se e somente se $-T$ é sublinear.

Analogamente, definimos operadores estritamente (fortemente) convexos (sublineares).

Agora consideraremos a equação operacional

$$y = Tx. \tag{A-1.5}$$

A-1.1.12 Proposição

Se $T : D(T) \subseteq X \longrightarrow Y$ é um operador de tipo monótono, então as seguintes afirmativas são corretas:

(1) Unicidade. Para $y \in Y$ fixo, a equação (A-1.5) tem no máximo uma solução x.

(2) Estimativa de erro. Se x é uma solução de (A-1.5), e se $Tu \leq y$ e $Tv \geq y$, então $u \leq x \leq v$.

(3) Caracterização . As seguintes afirmações são equivalentes

1. T é de tipo monótono;
2. T é injetivo, e T^{-1} é monótono crescente.

A equivalência vale para operadores estritamente (fortemente) monótonos.

Prova de (1)

Se $Tu = y$ e $Tv = y$. $Tu \leq Tv$ e $Tv \leq Tu \Rightarrow u \leq v$ e $v \leq u \Rightarrow u = v$.

Prova de (2)

Se $Tu \leq Tx \leq Tv$ então $u \leq x \leq v$.

Prova de (3)

Comparar definições (A-1.1.11) e (A-1.1.3).

Muito embora os últimos resultados sejam consequências triviais das respectivas definições , operadores de tipo monótono são de grande importância em análise numérica devido à facilidade de avaliar-se as estimativas de erro que eles fornecem. Em muitos sistemas não lineares parabólicos, elípticos e hiperbólicos, o operador diferencial de segunda ordem que conduz à solução , bem como os sistemas lineares que advém da discretização de suas equações diferenciais, são de tipo ou característica monótona.

O Teorema Principal, a seguir, mostra que operadores de tipo monótono têm uma importante propriedade, estabelece que a existência e unicidade de soluções e estimativas de erro podem ser determinadas do conhecimento de super e subsoluções , que são frequentemente fáceis de obter. Consideraremos a equação

$$Au + Hu = 0 \tag{A-1.6}$$

A-1.1.13 Teorema Principal para Operadores de Tipo Monótono

Suponha que as seguintes hipóteses sejam satisfeitas

(i) X e Y são espaços de Banach reais ordenados;

(ii) O operador $A : D(A) \subseteq X \longrightarrow Y$ é linear e $A^{-1} : Y \longrightarrow X$ existe e é um operador compacto;

(iii) O operador $H : X \longrightarrow Y$ é contínuo;

(iv) Existe uma região G em X com $0 \in G$ e existem elementos $v_1, v_2 \in G \bigcap D(A)$ satisfazendo

$$Av_1 + \tau Hv_1 \leq 0, Av_2 + \tau Hv_2 \geq 0 \quad \forall \tau \in [0,1] \tag{A-1.7}$$

onde $[v_1, v_2] \subseteq G$ e $[v_1, v_2]$ é limitado em X.

(v) Para todo $\tau \in [0,1]$, os operadores $A + \tau H$ são de tipo monótono em $\overline{G} \bigcap D(A)$.

Então temos as seguintes conclusões :

(a) Existência e unicidade. A equação (A-1.6) tem exatamente uma solução u em $\overline{G} \bigcap D(A)$ e $v_1 \leq u \leq v_2$.

(b) Estimativa de erro. Se pudermos encontrar elementos $u_1, u_2 \in \overline{G} \bigcap D(A)$ que satisfaçam

$$Au_1 + Hu_1 \leq 0, \quad Au_2 + Hu_2 \geq 0 \tag{A-1.8}$$

então é garantido que temos $u_1 \leq u \leq u_2$.

Prova:

(I) Existência. Consideraremos, em lugar de (A-1.6),

$$u = \tau Tu, \qquad Tu = -A^{-1}H(u), 0 \leq \tau \leq 1. \tag{A-1.9}$$

Por (ii) e (iii) o operador $T : X \longrightarrow X$ é compacto. Seja u uma solução da equação acima em $\overline{G}$, com $\tau \in [0,1]$. Entao $Au + \tau H(u) = 0, u \in \overline{G} \bigcap D(A)$, e $A + \tau H$ é de tipo monótono em $\overline{G} \bigcap D(A)$. Assim (A-1.7) implica que $v_1 \leq u \leq v_2$, e então $u \in [v_1, v_2]$.

Seja $G = X$. Pelo Princípio de Leray-Schauder [1], (A-1.9) tem solução para $\tau = 1$, isto é, (A-1.6) tem uma solução .

Seja $G \neq X$. Escolha uma região limitada G_1 tal que $[v_1, v_2] \subseteq G_1 \subseteq G$. Como (A-1.9) não pode ter soluções na fronteira ∂G_1,(A-1.9) não tem solução para $\tau = 1$, novamente pelo Princípio de Leray-Schauder.

(II) Unicidade e afirmação (b) seguem da proposição (A-1.1.12).

O Teorema Principal para equações de Hammerstein Abstratas

Estudaremos agora a chamada equação de Hammerstein Abstrata

$$u = KF(u) \tag{A-1.10}$$

paralelamente ao seu correspondente método iterativo

$$u_{n+1} = KF(u_n), \qquad v_{n+1} = KF(v_n) \qquad n = 0, 1, 2, ... \tag{A-1.11}$$

sob variadas hipóteses sobre o operador F. Mais uma vez, trocando $=$ por $\leq$ e $\geq$ em (A-1.10) definimos sub e supersoluções , respectivamente. Lembramos que F é positivo se e somente se $u \geq 0$ implica $F(u) \geq 0$. Formularemos três hipóteses

(H1)Y e Z são espaços de Banach reais ordenados. O cone de ordem Y_+ sobre Y é normal e com interior não vazio (o que implica que ele é gerador e total).

(H2) O operador $F : Y \longrightarrow Z$ é contínuo e o operador $K : Z \longrightarrow Y$ é linear, compacto e positivo.

(H3) $K : Z \longrightarrow Y$ é fortemente positivo.

As hipóteses (H1)-(H3) são encontradas naturalmente em muitas aplicações , e o problema definido por (A-1.10) pode ter as seguintes origens

• A equação integral de Hammerstein

$$u(x) = \int_M G(x,y) f(y, u(y)) dy$$

• O problema elíptico de valores na fronteira

$$M : Lu = f(x,u), \qquad \partial M : Bu = g$$

• O problema inicial de valores na fronteira para a eq. parabólica

$$u_t + Lu = f(x,t,u)$$

[1] H.Zeidler,Nonlinear Functional Analysis and its applications VolI,pg245,556

onde f é não linear e contínua.

Todos os problemas acima podem ser transformados equivalentemente em (A-1.10), onde a função real f gera o operador F. A essência é que todas as propriedades da aplicação $u \to f(x,t,u)$ de continuidade, diferenciabilidade, linearidade assintótica e a relação de ordem $\leq$ correspondam exatamente às mesmas propriedades da aplicação $u \to F(u)$. Garantimos (H3) para problemas de valores na fronteira escolhendo $Y = X_e$ (ver definição (A-1.1.5)), $Z = X$, onde $X = C(\overline{M})$ e $Le = 1, Be = 0$.

A-1.1.14 Teorema Geral de Existência

Se as hipóteses (H1)-(H2) são satisfeitas, temos os seguintes resultados

(a) Convergência do método iterativo para F monótono crescente em $[u_0, v_0]$. Se u_0 é uma subsolução e v_0 uma supersolução de (A-1.10) com $u_0 < v_0$, então (u_n) em (A-1.11) converge para a menor solução u de (A-1.10) em $[u_0, v_0]$, e (v_n) converge para a maior solução v de (A-1.10) em $[u_0, v_0]$. Além disso, temos as estimativas de erro

$$u_n \leq u \leq v \leq v_n \qquad n = 0, 1, 2, ...$$

(b) Existência de solução para F monótono decrescente. Se existir um $z \in Z$ tal que $F(u) \leq z$ para todo $u \in Y$, então (A-1.10) tem uma solução .

Prova do Teorema

Prova de (a)

(a) é uma consequência imediata do teorema (A-1.1.9).

Prova de (b)

Aplicamos o Teorema do Ponto Fixo de Schauder para a aplicação $KF : M \longrightarrow M$ onde

$$M = \{u \in Y : KF(Kz) \leq u \leq Kz\}.$$

M é não vazio, pois $F(Kz) \leq z$ e então $KF(Kz) \leq Kz$. Obviamente M é convexo e fechado. Como Y_+ é normal, M é fechado pela proposição (A-1.1.4). Finalmente, mostraremos que $KF(M) \subseteq M$. Suponha que $u \in M$. Então $F(u) \leq z$ implica $KF(u) \leq Kz$, e $u < Kz$ implica $F(u) \geq F(Kz)$, e então $KF(u) \geq KF(Kz)$. Segue então que $KF(u) \in M$.

Finalmente, o Teorema do Ponto Fixo de Schauder assegura uma solução $u = KF(u)$.

A-1.1.15 Corolário (Unicidade)

Suponha que (H1)-(H3) são satisfeitas. Temos 2 casos

(a) Estritamente sublinear e monótono estritamente crescente F. Neste caso, a equação (A-1.10) tem no máximo uma solução $u > 0$. Se $v > 0$ é uma subsolução estrita de (A-1.10), então não existe uma solução u de (A-1.10) com $0 < u \leq v$.

(b) Estritamente superlinear e monótono estritamente crescente F. Neste caso, a equação (A-1.10) não tem duas soluções distintas positivas u e v. Se $v > 0$ é uma supersolução estrita de (A-1.10), então não existe solução u de (A-1.10) com $0 < u \leq v$ (unicidade fraca).

Prova do Corolário

Prova de (a)

Seja $u = KF(u)$ e $v = KF(v)$ com $u \neq v, u, v > 0$. Podemos assumir que $v \not\leq u$. Como $KF(0) > 0$ e KF é monótona fortemente crescente, segue $u \gg 0$. Pelo Lema do Cone (A-1.1.10) existe um real $t > 0$ tal que $(u - tv) \in \partial Y_+$.

Como $v \not\leq u, t < 1$. O operador KF é fortemente sublinear, e então

$$u = KF(u) \geq KF(tv) \gg tKF(v) \geq tv$$

ou seja, $(u - tv) \in int(Y_+)$. Isto é uma contradição .

Se v é uma subsolução estrita, isto é $v < KF(v)$, onde $0 < u \leq v$, a conclusão segue similarmente.

Prova de (b)

O argumento é similar ao caso (a). Seja $u = KF(u), v = KF(v), 0 < u < v$. As hipóteses feitas sobre KF implicam que $0 < u \ll v$. Segue, pelo Lema do Cone, que existe um real $t > 0 : (tv - u) \in \partial Y_+$. Entretanto, como $u \ll v$ segue $t < 1$ e então

$$u = KF(u) \leq KF(tv) \ll tKF(v) \leq tv,$$

ou seja, $(tv - u) \in int(Y_+)$. Isto é uma contradição .

Se v é uma supersolução estrita, argumentamos similarmente.

Aplicações a Equações Integrais de Hammerstein.

Consideraremos a equação operacional abstrata (A-1.10) e o método iterativo

$$u_{n+1} = KF(u_n), \quad u_0 \in Y, n = 0, 1, 2... \tag{A-1.12}$$

Como protótipo de (A-1.10), consideraremos a equação integral de Hammerstein

$$u(x) = \int_M G(x, y) f(y, u(y)) dy \quad \forall x \in \overline{M} \tag{A-1.13}$$

com o método iterativo

$$u_{n+1}(x) = \int_M G(x, y) f(y, u_n(y)) dy \tag{A-1.14}$$

Para esta equação integral queremos obter um teorema de equivalência que permita-nos aplicar todos os resultados abstratos de (A-1.10) no problema concreto (A-1.13), o que conduzirá a uma abundância de resultados e aplicações .

Definiremos $Y = Z = C(\overline{M}), Y_+ = C_+(\overline{M})$ e o operador $K : Z \to Y$ por

$$(Kv))(x) = \int_M G(x, y) v(y) dy \tag{A-1.15}$$

e o operador $F : Y \longrightarrow Z$ por $z = F(u)$ e

$$z(x) = f(x, u(x)).$$

F é chamado de *operador de Nemyckii.*

A-1.1.16 Proposição (Teorema de Equivalência)

Suponha que M é uma região limitada em $\Re^N$ com $N \geq 1$, que o núcleo $G : \overline{M} \times \overline{M} \longrightarrow \Re$ é contínuo e não negativo e que $f : \overline{M} \times \Re \longrightarrow \Re$ é contínua. Então temos os seguintes resultados.

(1) O operador $K : Z \longrightarrow Y$ é linear, compacto e positivo. Se G é positivo, então K é fortemente positivo.

(2) O operador $F : Y \longrightarrow Z$ é contínuo. F é positivo se $f(x,u) \geq 0$ para todo $u \geq 0$ e $x \in \overline{M}$. Similarmente, F é monótono crescente or decrescente, convexo ou côncavo, e sublinear ou superlinear, se a função real $u \mapsto f(x,u)$ tem a correspondente propriedade em $\Re$ para todo $x \in \overline{M}$. Mais do que isso, $F(u) \geq cu$ se $f(x,u(x)) \geq cu(x)$ para todo $x \in \overline{M}$.

(3) F é continuamente diferenciável em Y se a derivada parcial f_u é contínua em $\overline{M} \times \Re$. Se colocarmos $y = F'(u)h$, então $y(x) = f_u(x,u(x))h(x)$ para todo $x \in \overline{M}$.

(4) A derivada $F'(\infty)$ existe se existir um número real α tal que

$$\frac{f(x,u))}{u} \to \alpha \quad ao \quad |u| \to \infty \tag{A-1.16}$$

e a passagem ao limite é uniforme com respeito a todo $x \in \overline{M}$. Se isto vale somente para $u \to \infty$, então a derivada positiva $F'_+(\infty)$ existe. Se colocarmos $y = F'(\infty)h$ ou $y = F'(\infty)h$, respectivamente, então $y(x) = \alpha h(x) \quad \forall x \in \overline{M}$.

(5) A função $u \in y$ é uma sub ou supersolução de (A-1.10) se e somente se o símbolo $=$ em (A-1.13) é trocado por $\leq$ ou $\geq$, respectivamente.

(6) O cone de ordem Y_+ é normal e tem interior não vazio e então , em nosso contexto, ele é gerador e total.

A prova segue imediatamente das correspondentes definições .

A-1.1.17 Lema A

Existe uma solução única $T_\beta \in X$ de (2.5) para

$$f(T(x,t)) = f_\beta(T(x,t)) = \beta \frac{u_0}{C} \frac{T}{1+T}(x,t) \tag{A-1.17}$$

e $0 < \beta \leq 1, u_0 \in \Re$.

Prova:

(I) É imediato, pela representação integral, que existem sub e supersoluções triviais $T_N(x,t)$ e $T_c(x,t)$ tais que

$$C\frac{\partial T_N}{\partial t} = \frac{\partial^2 T_N}{\partial x^2} \qquad \text{e} \qquad C\frac{\partial T_c}{\partial t} = \frac{\partial^2 T_c}{\partial x^2} + u_0$$

substituem (2.1) e então $T_N < T_\beta < T_c$ em $M \quad \forall \beta : 0 < \beta \leq 1$, uma vez que

$$0 < f_\beta(T(x,t)) < \frac{u_0}{C}$$

para toda função $T(x,t) \in X$.

(II) Mostraremos que f_β é uma função monótona estritamente crescente e estritamente sublinear para todo $\beta : 0 < \beta \leq 1$.

Sejam duas funções quaisquer $T_{\alpha_1}, T_{\alpha_2} \in X, T_{\alpha_1} < T_{\alpha_2}$ (isto é $T_{\alpha_1}(x,t) \leq T_{\alpha_2}(x,t)$ em M e $T_{\alpha_1} \neq T_{\alpha_2}$). Temos

$$f_\beta(T_{\alpha_2}) - f_\beta(T_{\alpha_1}) = \beta\frac{u_0}{C}\left[\frac{T_{\alpha_2}}{1+T_{\alpha_2}} - \frac{T_{\alpha_1}}{1+T_{\alpha_1}}\right] = \beta\frac{u_0}{C}\frac{T_{\alpha_2} - T_{\alpha_1}}{(1+T_{\alpha_2})(1+T_{\alpha_1})} > 0 \qquad \text{(A-1.18)}$$

ou seja, $f_\beta(T_{\alpha_1}) < f_\beta(T_{\alpha_2})$.

Seja $\lambda \in \Re, 0 < \lambda < 1$. Observamos que

$$1 + \lambda T < 1 + T \Rightarrow \lambda \cdot \frac{1}{1+\lambda T} > \lambda \cdot \frac{1}{1+T}$$

e então $f_\beta(\lambda T) > \lambda f_\beta(T) \forall \lambda \in (0,1)$.

(III) No sentido de satisfazer as hipóteses de (A-1.1.14) e (A-1.1.15), já observamos que o núcleo de Green definido em (2.10) verifica

$$G_0(x,t,y) > 0 \qquad \forall x, y \in (0,2L), \forall t \in (0,t_0)$$

e então , escrevendo (2.14) na forma

$$T = K_0 F_0(T), \qquad T \in X = C(M) \qquad \text{(A-1.19)}$$

o operador K_0 é fortemente positivo como aplicação de $X = C(M) \longrightarrow X_e$.

Amparados no teorema (A-1.1.14) e no corolário (A-1.1.15) (apêndice), asseguramos existência e unicidade de soluções em (2.5) para

$f(T(x,t)) = f_\beta(T(x,t)), 0 < \beta \leq 1$.

A-1.1.18 Lema B

Sendo a função $f(T) : D(f) = X = C(\overline{M}) \longrightarrow \Re$ definida por

$$f(T(x,t)) = \frac{u_0}{C}\frac{T}{1+T}(x,t)\exp\left[-\int_0^x \frac{T}{1+T}(s,t)ds\right]$$

podemos mostrar que $f(T)$ é estritamente sublinear (e então $F(T)$ em

$$T = KF(T) \qquad \text{(A-1.20)}$$

é um operador estritamente sublinear).

Prova

Primeiramente, observamos que, para $0 < \lambda < 1$

$$1 + \lambda T < 1 + T \Rightarrow \frac{1}{1+\lambda T} > \frac{1}{1+T}$$

então

$$\frac{\lambda T}{1+\lambda T}\exp\left[-\int_0^x \frac{\lambda T}{1+\lambda T}(s,t)ds\right] = \frac{\lambda T}{1+\lambda T}\exp\left[-\int_0^x \left(1-\frac{1}{1+\lambda T}\right)ds\right] =$$
$$\frac{\lambda T}{1+\lambda T}\exp(-x)\exp\left[\int_0^x \frac{1}{1+\lambda T}ds\right] > \frac{\lambda T}{1+\lambda T}\exp(-x)\exp\left[\int_0^x \frac{1}{1+T}ds\right] =$$
$$\lambda\cdot\frac{T}{1+T}\exp\left[-\int_0^x \frac{T}{1+T}ds\right]$$

segue, $\forall T \in D(f), 0<\lambda<1$

$$\lambda f(T) < f(\lambda T)$$

e f é estritamente sublinear.

A-1.1.19 Lema C

Sendo $f(T)$ definida conforme hipótese do Lema B, asseguramos sua monotonicidade estrita para $L \leq 1/2$. Este resultado, combinado com o Lema B, garante existência e unicidade de soluções da equação (2.16) (e consequentemente da equação (2.14)) como aplicação do teorema (A-1.1.14) e do corolário (A-1.1.15) .

Prova do Lema C

(I) Observamos que, derivando $f(T)$ com respeito a variável real T, temos

$$\frac{\partial f(T)}{\partial T} = \frac{u_0}{C}\frac{T}{1+T}(x,t)\exp\left[-\int_0^x \frac{T}{1+T}(s,t)ds\right]\left[\frac{1}{T(1+T)} - \int_0^x \frac{1}{(1+T)^2}ds\right] \qquad \text{(A-1.21)}$$

e que a maneira mais adequada de definir o cone de ordem X_+ sobre X é

$$X_+ = \{T \in X : T(x,t) > 0 \quad \forall x,t \in M\} \qquad \text{(A-1.22)}$$

concluímos que então precisamos assegurar que

$$D(x,t) = \frac{1}{T(1+T)}(x,t) - \int_0^x \frac{1}{(1+T)^2}ds > 0 \quad \forall x,t \in M, \forall T \in X \bigcap X_s \qquad \text{(A-1.23)}$$

e então $f(T)$ é monótona estritamente crescente em $X \bigcap X_s$.

A conclusão de monotonicidade estrita para qualquer $L > 0$ seria bastante natural no contexto dos problemas de reação -difusão e do problema do calor, onde os operadores que conduzem à solução traduzem em sípropriedades associadas a decaimentos suaves e uniformes, como por exemplo, o desaquecimento de uma placa metálica ou a difusão de um poluente em meio fluído, exemplos protótipos de problemas de natureza parabólica, que são caracterizáveis pela validade do Princípio do Máximo em alguma de suas variantes. Entretanto, apesar de todas essas evidências, somente conseguimos assegurar (A-1.23) para valores de L pequenos. A chave para a extensão de nossos resultados certamente será uma majoração mais refinada do termo integral

$$\int_0^x \frac{1}{(1+T)^2}(s,t)ds$$

para $0 < x < 2L$. Observamos que

$$\int_0^x \frac{1}{(1+T)^2}(s,t)ds \leq x\cdot\max_{0<s<x}\frac{1}{(1+T)^2}(s,t) = x\cdot\frac{1}{(1+T)^2}(x,t) \leq \frac{2L}{(1+T)^2}(x,t)$$

e então vale (A-1.23) se

$$2L < \frac{1 + T(x,t)}{T(x,t)}$$

e então garantimos (A-1.23) se $2L \leq 1$.

BIBLIOGRAFIA

[But 82] A.G.Butkovskiy. Green's Functions and Transfer Functions Handbook. John Wiley & Sons, 1982.

[Cav 80] S.Oh,J.Cavendish & L.Hegedus.Mathematical modeling of catalytic converter lightoff: Single-pellet studies,AIChE J.,26(1980),p935.

[Cav 85] S.H.Oh and J.C.Cavendish. Mathematical modeling of catalytic converter lightoff. PartII: Model verification by engine-dynamometer experiments; PartIII: Prediction of vehicle exhaust emissions and parametric analysis,AIChE J.,31(1985),pp.935-942; pp.943-949.

[Cra 76] J.Crank. The Mathematics of Difusion. Oxford Press, 1976.

[Fou 81] L.R.Foulds,Optimization Techniques.Springer-Verlag,1981,pg 346.

[Fri 64] A.Friedman. Partial differential equations of parabolic type. Prentice-Hall, 1964.

[Fri 91] A.Friedman. Mathematics in Industrial Problems, Ch7, Springer-Verlag,1991.

[Fri 94] A.Friedman,W.Littman. Industrial Mathematics. A Course in Solving Real World Problems, SIAM 1994.

[Kir 70] D.Kirk, Optimal Control Theory. Prentice-Hall (1970).PartIV.

[Mar 75] G.Marchuk. Methods of Numerical Mathematics. Springer -Verlag,1975,pg 71.

[Smo 83] J.Smoller. Shock Waves and reaction-difusion equations. Springer-Verlag, 1983.

[Wil 67] Wilde and Beightler. Foundations of Optimization. Prentice-Hall,1967.

[Zei1 85] Zeidler,Nonlinear Functional Analysis and its Applications I.Springer-Verlag, 1985,ch2.

[Zei3 85] Zeidler,Nonlinear Functional Analysis and its Applications III.Springer-Verlag, 1985,ch37,38.