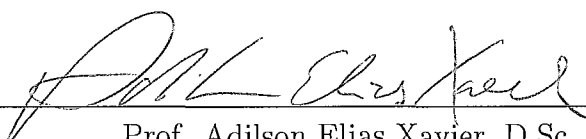


APLICAÇÃO DO MECANISMO DE EXTRAPOLAÇÃO NO
MÉTODO DE PENALIZAÇÃO HIPERBÓLICA

Leonardo Silva de Lima

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS
PROGRAMAS DE PÓS - GRADUAÇÃO DE ENGENHARIA DA
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE
EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Aprovada por:


Prof. Adilson Elias Xavier, D.Sc.


Profª. Maria Aparecida Diniz Ehrhardt, D.Sc.


Prof. Carlile Campos Lavor, D.Sc.


Prof. Marcone Jamilson Freitas Souza, D.Sc.


Prof. Luís Alfredo Vidal de Carvalho, D.Sc.

RIO DE JANEIRO, RJ - BRASIL
NOVEMBRO DE 2002

LIMA, LEONARDO SILVA DE

Aplicação do Mecanismo de Extrapolação
no Método de Penalização Hiperbólica
[Rio de Janeiro] 2002

VIII, 94 p. 29,7 cm (COPPE/UFRJ,
M.Sc., Engenharia de Sistemas e Computação,
2002)

Tese – Universidade Federal do Rio de
Janeiro, COPPE

1 - Programação Não-Linear

2 - Penalização Hiperbólica

3 - Extrapolação de Richardson-Romberg

I. COPPE/UFRJ II. Título (série)

Agradecimentos

À Deus que tem sido um amigo fiel e tem me sustentado com sua graça e poder. Glória ao Teu nome.

À minha esposa Lícia que esteve sempre ao meu lado dispensando o seu apoio, a sua dedicação e compreensão de que esse trabalho é uma vitória nossa. *Eu te amo.*

Aos meus pais José Fiel e Marilene que com tanto esforço investiram, acreditaram e também vivenciaram cada instante dessa trajetória, desde a graduação na Unicamp. Muito obrigado.

Aos meus irmãos Eduardo e Ingrid que torceram por mim e que têm sido acima de tudo amigos.

Ao meu tio Maurício e sua família pelo apoio desde a graduação. Que Deus continue abençoando vocês.

Ao meu orientador Prof. Adilson Xavier que tem sido um amigo e que conduziu esse trabalho com dedicação.

Aos amigos cearenses e cariocas que viveram o dia-a-dia comigo na COPPE.

Ao CNPQ pelo financiamento sem o qual esse trabalho seria inviável.

Aos funcionários da COPPE/Sistemas. Em especial Dona Gercina, Fred, Solange, Cláudia, Sônia, Sueli, Mercedes e tantos outros que deram suporte e sempre me atenderam com gentileza.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc)

APLICAÇÃO DO MECANISMO DE EXTRAPOLAÇÃO NO MÉTODO DE PENALIZAÇÃO HIPERBÓLICA

Leonardo Silva de Lima

Novembro/2002

Orientador : Adilson Elias Xavier

Programa : Engenharia de Sistemas e Computação

A minimização de problemas não lineares irrestritos através dos métodos de penalização exige a presença do parâmetro de penalização $\tau > 0$. Assim, a resolução dos subproblemas penalizados até que se alcance uma aproximação desejada da solução ótima, torna-se cada vez mais difícil devido aos pequenos valores que o parâmetro τ assume ao longo do processo de minimização. Isto se deve, principalmente, ao fato do mau condicionamento da matriz Hessiana, pois as curvas de nível da Hessiana são extremamente achatadas. Diante desta problemática, neste estudo utilizaremos o mecanismo de extrapolação de Richardson-Romberg com a finalidade de realizarmos boas estimativas para a solução ótima do problema com valores não críticos de τ . A extrapolação, na realidade, é um mecanismo que se utiliza de dados obtidos no passado para realizar previsões. Desta forma, a extrapolação é uma ferramenta para acelerar a convergência ao ótimo no processo de minimização, que neste trabalho é realizado utilizando-se o método de Penalização Hiperbólica, que é um método que pertence à família dos métodos de penalidade e combina características tanto dos métodos de penalidade interior como exterior. Os resultados computacionais apresentados no último capítulo corroboram com os resultados previstos teoricamente e comprovam a eficiência do método de extrapolação para acelerar a convergência em direção ao ótimo.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

APPLICATION OF THE RULE EXTRAPOLATION ON
HIPERBOLIC PENALTY METHOD

Leonardo Silva de Lima

Novembro/2002

Advisors : Adilson Elias Xavier

Department : Engineering and Computer Science

The solution for constrained nonlinear programming problems with penalty methods is usually get through the resolution of the sequence of unconstrained problems produced by penalty parameter τ . Therefore, the resolution of subproblems become increasing hard when τ approaches to zero because of ill-conditioning in the Hessian. In this paper we profit the rule of Richardson-Romberg extrapolation to get good estimatives for optimal solution of the problem when the value of the penalty term is not so critical and then speed up the convergence to the optimal in the minimization process. We make use of the Hyperbolic Penalty that belongs to penalty methods family and combine interior and exterior penalty. The computacional results in the last chapter agree with those predicted in the theory.

Índice

1	Revisão Bibliográfica	1
1.1	Introdução	1
1.2	Penalidade	3
1.3	Integração Numérica	15
2	Penalização Hiperbólica	25
2.1	Análise de Convexidade	28
2.2	Propriedades da Função Penalidade	29
2.3	Multiplicadores de Lagrange	31
2.4	Análise da Hessiana	34
2.5	Soluções do problema irrestrito penalizado	37
2.6	Algoritmo de Penalização Hiperbólica	38
3	Extrapolção em Penalização Hiperbólica	39
3.1	Desenvolvimento teórico	42
3.1.1	Existência da trajetória $x(\tau)$	43

3.1.2	Cálculo de $\frac{dx(\tau)}{d\tau}$	45
3.1.3	Existência de $\frac{d^n x(\tau)}{d\tau^n}$ e $\frac{d^n \mu(\tau)}{d\tau^n}$	46
3.1.4	Análise da solução em $\tau = 0$	48
3.2	Extrapolação Polinomial	55
3.2.1	Extrapolação Inversa	59
4	Resultados Computacionais	62
4.1	Problema Teste 1	64
4.2	Problema Teste 2	72
4.3	Problema Teste 3	79
4.4	Análise dos resultados	83
4.5	Análise de ρ	85
	Conclusões	88
	Referências Bibliográficas	92

Lista de Figuras

1.1	Curva de Nível com $\tau = 10$ e $\tau = 0.5$	14
1.2	Curvas de Nível com $\tau = 0.1$	14
1.3	Matriz de Richardson-Romberg	21
1.4	Matriz de Richardson-Romberg	23
2.1	Função Penalidade Hiperbólica	26
2.2	Primeira e Segunda Fase de P.H.	28
3.1	Estimativa do Ótimo do próximo subproblema penalizado	61

Capítulo 1

Revisão Bibliográfica

1.1 Introdução

A ciência está relacionada a previsões. Podemos fazer previsões apenas para obter modelos ou esquemas conceituais do mundo. Os modelos usados atualmente são aqueles que têm melhor sobrevivido à evolução do processo científico. Na realidade, não existem modelos novos debaixo do sol, mas os modelos desses processos têm permitido a humanidade transformar o universo. O conhecimento científico não é fruto apenas da observação, mas também de teorias, hipóteses, modelos ou esquemas conceituais que correspondem aos dados e que são confrontados diariamente com os resultados esperados e observados na prática. Existe uma contínua interação entre a hipótese e os dados observados que acabam por nos levar ao progresso e gerar mais campos de pesquisa. Ao compararmos a teoria com o experimento, devemos analisar uma quantidade de pontos tão grande quanto possível e o mais preciso possível. Desta maneira, procura-se alcançar pontos onde a conexão entre a teoria e o experimento são os mais eficientes possíveis. Neste contexto, a extrapolação é na realidade um mecanismo de predição de eventos baseados em um conjunto de dados

obtidos no passado. Essa predição está firmada numa sólida teoria e hipóteses que tornam o experimento uma boa ferramenta para avaliar a teoria desenvolvida e os resultados obtidos na prática, objetivando aproximá-los ao máximo.

A grande diversidade de modelos e problemas práticos onde a previsão do comportamento da solução com prévia antecedência pode trazer benefícios significativos, por exemplo em termos de tempo computacional, motivou-nos estudar e abordar alguns aspectos do mecanismo de extrapolação aplicado à Penalização Hiperbólica.

Fiacco [3] em 1966 e Lootsma [11] em 1968 foram os principais autores que apresentaram uma teoria que analisa o uso da extrapolação em problemas de programação matemática e sua aplicação. Apesar desses autores descreverem os possíveis ganhos proporcionados pelo uso da extrapolação, não encontramos muitos trabalhos que explorem algumas facetas importantes deste método, o que pretendemos abordar nesse trabalho.

Nesse capítulo faremos uma breve abordagem dos conceitos básicos da família dos métodos de penalidade interior e exterior e uma primeira aplicação do método de extrapolação de Richardson-Romberg no problema de integração numérica, onde foi utilizado pela primeira vez.

No capítulo 2 apresentaremos o método de Penalização Hiperbólica desenvolvido por Xavier [23] em 1982, visando solucionar problemas de programação matemática não-lineares com restrições de desigualdade. Esse método de minimização foi utilizado para a resolução dos problemas propostos com a aplicação das rotinas de extrapolação nesse trabalho. Dessa forma, apresentaremos ainda algumas propriedades interessantes do método, bem como a análise da matriz Hessiana da função irrestrita

gerada pelo método.

No capítulo 3 desenvolveremos a base teórica para a aplicação do mecanismo de extrapolação no método de Penalização Hiperbólica como uma ferramenta que tem como prioridade acelerar o processo de convergência em direção ao ótimo. Veremos também as principais características e a viabilidade teórica desse método de penalidade para o uso da extrapolação.

No capítulo 4 realizaremos os experimentos computacionais e faremos um paralelo entre a resolução do problema com a extrapolação e a resolução do mesmo sem a aplicação do mecanismo de extrapolação. Apresentaremos a resolução de um conjunto de problemas teste retirados, principalmente de Hock-Schittkwoski [7] e analisaremos a eficiência do método de extrapolação.

No capítulo 5, apresentaremos as conclusões e observações pertinentes aos resultados obtidos no capítulo 4. Conjecturamos ainda a aplicação da extrapolação em problemas mais específicos e de grande porte, como o problema do recobrimento apresentado por Xavier [19].

Nas seções seguintes faremos um estudo sobre os métodos de penalidade e sua filosofia para resolver os problemas de otimização

1.2 Penalidade

É com grande frequência que não só no campo da otimização como em todas as áreas da ciência, a resolução de uma determinada questão passa por etapas tais quais o entendimento do problema e suas características intrínsecas e a identificação

de métodos que se adequem à resolução do problema proposto. Nesse processo, na maioria dos casos, busca-se resolver o problema complexo substituindo-o pela resolução de uma seqüência de subproblemas mais simples, cujos métodos para a resolução desses estão bem definidos. Transforma-se então um problema de difícil resolução em um problema onde a resolução é conhecida, conforme narrado por Martinez [12] .

Os métodos de penalização são procedimentos que transformam problemas de otimização restritos para problemas irrestritos, os quais possuem uma ampla classe de métodos de resolução bem definidos na literatura, tais como Newton e Quase-Newton, onde podemos destacar as variantes BFGS (Broyden, Fletcher, Goldfarb, Shanno) e DFP (Davidon, Fletcher, Powell) entre outras. Assim, a aplicação dos métodos de penalidade torna viável a utilização de métodos de minimização irrestrita mais eficientes e mais robustos.

Conceitualmente, a penalização é a incorporação das restrições na função objetivo, provocando um acréscimo desta de acordo com a não satisfação das restrições. Estes ganharam popularidade devido à sua simplicidade e fácil implementação computacional.

A família dos métodos de penalidade é composta basicamente por duas classes: penalidade interior e exterior.

Na penalidade interior, agrega-se à função objetivo um termo envolvendo as restrições. Este termo tem a função de desencorajar a aproximação à fronteira estabelecida pelo conjunto viável, tendendo a infinito de acordo com a proximidade à fronteira. Cria-se assim uma barreira intransponível de modo que ao iniciar-se o

processo de minimização irrestrita com um ponto viável, toda a sequência de pontos obtidos serão viáveis. Por esse motivo, os métodos de penalidade interior são conhecidos como métodos de pontos interiores e ainda como "métodos de barreiras". É importante ressaltar que esses ganharam grande impulso após a publicação do trabalho de Karmakar [8] em 1984, sendo então alvo de novas pesquisas e implementações.

Na penalidade externa, agrega-se à função objetivo um termo que tem seu custo aumentado de acordo com a violação das restrições. Neste caso, a sequência de pontos obtidos a partir da minimização da função objetivo modificada é inviável. Portanto, a solução obtida, geralmente é um ponto que está fora da região viável, mas aproxima-se dela à medida que aumenta-se o parâmetro de penalização.

Apesar da praticidade computacional e simplicidade, a penalização gera dificuldades numéricas decorrentes da severidade com que se penaliza o problema. Esta severidade é determinada por um parâmetro τ que multiplica a função restrição. Portanto, a função objetivo modificada na violação (penalização externa) ou no risco de violação (barreira), implica na mudança dos valores de τ , que ao assumir valores extremos torna a estrutura do problema cada vez mais desfavorável. Esta desestruturação está relacionada à instabilidade numérica do problema indicada pelo mau condicionamento da matriz Hessiana, que pode causar danos irreparáveis, como a obtenção de uma solução que não é a solução ótima. Descreveremos então nesse trabalho um mecanismo que visa amenizar esta instabilidade e promover boas previsões para a solução ótima.

Podemos ainda destacar outra classe de métodos que conjugam tanto penalização

exterior quanto interior. Queremos destacar a Penalização Hiperbólica desenvolvida por Xavier [23] que manipula dois parâmetros. Um desses parâmetros está associado à penalização exterior quando da obtenção de um ponto inviável. Depois de alcançada a viabilidade mantém-se constante a penalização exterior e varia-se o segundo parâmetro relacionado à penalização interna. Esse método foi utilizado neste trabalho para a resolução dos problemas de programação não-linear sujeitos unicamente a restrições de desigualdade, onde utilizaremos ainda o mecanismo de extrapolação como ferramenta para acelerar o processo de convergência em direção à solução ótima. Veremos com mais detalhes a função penalidade hiperbólica bem como suas principais propriedades nos capítulos posteriores.

Vejamos o método de barreira de maneira mais detalhada, em especial a barreira logarítmica, com intuito de compararmos mais adiante com a penalização hiperbólica, através do exemplo em Ariela [13].

Métodos de Barreira

O método de barreira é aplicável a problemas da forma:

$$\begin{aligned}
 & \text{Minimizar } f(x) \\
 & \text{sujeito a } g_i(x) \geq 0 \\
 & \quad i = 1, \dots, m
 \end{aligned} \tag{1.1}$$

onde o conjunto de restrições $S = \{x \mid g(x) \geq 0\}$ tem interior não-vazio. Intuitivamente, podemos dizer que o conjunto S tem interior e a partir de um ponto viável é possível alcançar um ponto da fronteira por aproximação. O método de barreira é um método de pontos viáveis, pois trabalha de forma a fazer da fronteira

uma barreira para evitar que pontos inviáveis sejam escolhidos pelo processo de busca.

A função barreira é uma função B definida no interior de S tal que :

- *i*) B é contínua;
- *ii*) $B(x) \geq 0$;
- *iii*) $B(x) \rightarrow \infty$, quando x se aproxima da fronteira de S .

Seja g_i , $i = 1, 2, \dots, m$ uma função contínua em R^n . Considere que $S = \{x : g_i(x) \geq 0, i = 1, 2, \dots, m\}$ é limitado e tem interior não-vazio. A partir de S o conjunto S_o é denotado por $S_o = \{x : g_i(x) > 0, i = 1, 2, \dots, m\}$.

As funções barreiras clássicas são:

$$B(x) = - \sum_{i=1}^m \log(g_i(x)) \quad (1.2)$$

e

$$B(x) = \sum_{i=1}^m \frac{1}{g_i(x)}. \quad (1.3)$$

As funções (1.2) e (1.3) são definidas no interior de S e deveriam satisfazer as condições *i*, *ii*, *iii*. Porém (1.2) não satisfaz à condição *ii* , já que para valores de $g_i(x) > 1$ assume valor negativo. Contudo, isto não representa um impecílio ao uso desta função, pois consegue-se facilmente construir outra função equivalente a (1.2) através de uma transformação de escala de modo que venha a satisfazer todas as condições estabelecidas.

Consideremos então a função:

$$F(x, \tau) = f(x) + \tau B(x) = f(x) - \tau \sum_{i=1}^m \log(g_i(x)) \quad (1.4)$$

onde $\tau \in R$ e $\tau > 0$.

Desta forma trocamos a resolução de (1.1) pelo problema:

$$\text{Minimizar } F(x, \tau_k) \quad (1.5)$$

$$x \in S_o \quad (1.6)$$

Na realidade o problema acima é restrito ao interior da região viável. Entretanto, podemos resolvê-lo usando uma técnica de busca irrestrita. Para encontrar a solução partimos de um ponto inicial viável e então a busca trabalha de forma a encontrar sempre pontos viáveis. Próximo a fronteira de S a função objetivo aproxima-se do infinito, logo a busca permanece na região viável e a restrição não é levada em consideração explicitamente. Do ponto de vista teórico o problema é restrito, mas do ponto de vista prático computacional é irrestrito.

Antes de analisarmos a convergência do método, devemos assumir alguma condição de regularidade. Para tal, definimos $I(x)$ como conjunto das restrições ativas no ponto x , onde $I(x) = \{i \mid g_i(x) = 0, i = 1, \dots, m\}$.

Hipótese de Regularidade: Em todo ponto $x \in R^n$, os gradientes das restrições ativas no ponto $I(x)$ são linearmente independentes.

Convergência do método

Abaixo apresentaremos um resultado conhecido da literatura [9] sobre a convergência do método das barreiras. No enunciado, está suposto a convergência da seqüência de pontos a um ponto limite. Essa hipótese será facilmente atendida se o conjunto viável S for compacto.

Teorema 1.1 *Se a hipótese de regularidade for obedecida, todo ponto limite da seqüência x_k gerada pelo método de barreira atende às condições de otimalidade de primeira ordem de Karush-Kuhn-Tucker (KKT).*

Condições de Karush-Kuhn-Tucker para (1.1):

$$\nabla f(x^*) - \sum_{i=1}^m \mu_i^* \nabla g_i(x^*) = 0 \quad (1.7)$$

$$g_i(x^*) \geq 0, \quad i = 1, \dots, m \quad (1.8)$$

$$\mu_i^* g_i(x^*) = 0, \quad i = 1, \dots, m \quad (1.9)$$

onde

$$\mu_i^* \geq 0, \quad i = 1, \dots, m \quad (1.10)$$

sendo $\mu_i^* \in R^m$ os multiplicadores ótimos de Lagrange e x^* é minimizador local regular.

Como (1.5) é um problema irrestrito, então a condição de otimalidade de 1ª ordem estabelece que o gradiente da função objetivo em um ponto de mínimo intermediário $x(\tau)$ observa a relação:

$$\nabla F(x, \tau) = \nabla f(x) - \sum_{i=1}^m \frac{\tau}{g_i(x)} \nabla g_i(x) = 0 \quad (1.11)$$

Passando (1.11) ao limite quando $\tau \rightarrow 0$ temos:

$$\lim_{\tau \rightarrow 0} [\nabla f(x(\tau)) - \sum_{i=1}^m \frac{\tau}{g_i(x(\tau))} \nabla g_i(x)] = 0. \quad (1.12)$$

Considere uma subsequência $x(\tau)$ convergente para x^* :

$$\nabla f(x^*) - \sum_{i=1}^m \lim_{\tau \rightarrow 0} \frac{\tau}{g_i(x(\tau))} \nabla g_i(x^*) = 0 \quad (1.13)$$

Comparando-se as equações (1.7) e (1.13) temos dois casos a serem analisados:

I - Se $g_i(x^*) > 0$ então $\lim_{\tau \rightarrow 0} \frac{\tau}{g_i(x(\tau))} = \lim_{\tau \rightarrow 0} \frac{\tau}{g_i(x^*)} = 0$

II - Diante desse fato, o somatório de (1.13) se restringe somente aos índices associados às restrições ativas no conjunto $I(x^*)$, em que $g_i(x^*) = 0$. Sem perda de generalidade, vamos considerar que as restrições ativas sejam as m^* primeiras. Assim, a expressão (1.13) pode ser escrita como:

$$\nabla f(x^*) - \sum_{i=1}^{m^*} \lim_{\tau \rightarrow 0} \frac{\tau}{g_i(x(\tau))} \nabla g_i(x^*) = 0 \quad (1.14)$$

Pela condição de regularidade, a expressão (1.14) tem solução única, a qual vamos denominar $\hat{\mu}$. Como, na expressão $\lim_{\tau \rightarrow 0} \frac{\tau}{g_i(x(\tau))}$ em (1.14), o denominador $g_i(x(\tau)) > 0$, já que estamos caminhando por pontos viáveis, e ainda o numerador $\tau > 0$, pode ser concluído que $\hat{\mu} \geq 0$.

Logo, verificamos que as condições de KKT (1.7), (1.8), (1.9) e (1.10) são atendidas no limite das seqüências geradas pelo método. Assim, comparando-se as equações (1.7) e (1.13) pode ser visto que o método gera valores que convergem para os multiplicadores de Lagrange. Dessa forma, os valores

$$\mu_i = \frac{\tau}{g_i(x(\tau))}$$

gerados a cada resolução de um subproblema penalizado representam estimativas dos multiplicadores de Lagrange. ■

Assim, a resolução de (1.5) com τ_k além de nos fornecer o valor $x^k(\tau_k)$ no espaço primal, ainda nos dá uma estimativa dos multiplicadores de Lagrange $\mu^k(\tau_k)$. O método de penalidade interior considerado possui a excelente propriedade de ao final da cada resolução de um subproblema penalizado, oferecer estimativas para a solução ótima primal e ainda estimativas para a solução ótima dual através dos valores μ .

Como a solução do problema se dá através da resolução de uma seqüência de subproblemas irrestritos (1.5), com valores monotonicamente decrescentes de τ , podemos afirmar que após k subproblemas resolvidos temos as seqüências de parâmetros $\tau_1 > \tau_2 > \tau_3 > \dots > \tau_k$, de valores primais $x(\tau_1), \dots, x(\tau_k)$ e ainda de valores duais $\mu(\tau_1), \dots, \mu(\tau_k)$ conhecidos. Adiantamos que essas são as seqüências base para o uso do mecanismo de extrapolação.

A seqüência $x(\tau_1), \dots, x(\tau_k)$ gera uma trajetória de pontos interiores em direção ao ótimo, chamada de Trajetória Central, onde normalmente o ponto ótimo anterior é utilizado como ponto inicial para a resolução do próximo subproblema de minimização irrestrita.

Todo o esquema parece consistentemente perfeito, entretanto a resolução de (1.5)

torna-se cada vez mais difícil numericamente. A redução do parâmetro de penalidade τ gera dificuldades numéricas crescentes à medida em que os valores de τ se aproximam de zero, face ao crescente mau condicionamento da matriz Hessiana. Em face ao fato da degenerescência da matriz Hessiana houve um crescente “abandono” dos métodos de penalidade na década de 70, pois essas dificuldades provocam a produção de resultados não confiáveis.

Observamos ainda que, conforme descreve Wright [17], se desejarmos minimizar o problema irrestrito (1.5) utilizando o método de Newton, apesar desse método apresentar comportamento favorável em problemas gerais, é conhecido como um método problemático quando aplicado a problemas irrestritos com função barreira. Isso se deve ao crescente mau condicionamento da matriz Hessiana. Em Lootsma [10], foi demonstrado que a Hessiana é mau condicionada em pontos da trajetória central para valores de τ suficientemente pequenos, e é assintoticamente singular. Wright [17] apresenta uma análise local indicando porque um passo puro de Newton num típico método de barreira de passo longo, para problemas não-lineares, pode ser inviável, mesmo quando tomados a partir de pontos aparentemente favoráveis. Isso mostra que é necessário o uso de mecanismos que realizem boas estimativas da solução ótima quando ainda não estamos muito próximos desta. Para isso, utilizaremos nesse trabalho o mecanismo de extrapolação.

Vejamos abaixo um pequeno exemplo apresentado por Martinez [12] onde o número de condição da Hessiana, associado ao parâmetro τ , cresce à medida que τ se aproxima de zero, ou seja, cresce à medida que x se aproxima de x^* .

$$\text{Min } f(x_1, x_2) = (x_1 + 1)^2 + (x_2 - 1)^2$$

$$x_1 \geq 0,$$

cuja solução é

$$x^* = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

A função barreira é dada por

$$B(x) = -\log(x_1).$$

O problema irrestrito fica :

$$\text{Min } F(x, \tau) = (x_1 + 1)^2 + (x_2 - 1)^2 - \tau \log(x_1). \quad (1.15)$$

Portanto, temos o gradiente de $F(x, \tau)$,

$$\nabla_x F(x, \tau) = \begin{pmatrix} 2(x_1 + 1) - \frac{\tau}{x_1} \\ 2(x_2 - 1) \end{pmatrix}. \quad (1.16)$$

A Hessiana é :

$$\nabla_x^2 F(x, \tau) = \begin{pmatrix} 2 + \frac{\tau}{x_1^2} & 0 \\ 0 & 2 \end{pmatrix}. \quad (1.17)$$

Os pontos estacionários com $x_1 > 0$ são da forma :

$$\tilde{x} = \begin{pmatrix} \frac{-1 + \sqrt{1 + 2\tau}}{2} \\ 1 \end{pmatrix},$$

para $\tau > 0$.

Logo,

$$\nabla_x^2 F(x, \tau) = \begin{pmatrix} 2 + \frac{2\tau}{\tau + 1 - \sqrt{1 + 2\tau}} & 0 \\ 0 & 2 \end{pmatrix}. \quad (1.18)$$

Podemos então observar que:

$$\lim_{\tau \rightarrow 0} \frac{2\tau}{\tau + 1 - \sqrt{1 + 2\tau}} = \infty. \quad (1.19)$$

Assim, o exemplo acima ilustra que à medida que no processo de minimização $\tau \rightarrow 0$, a matriz Hessiana torna-se cada vez mais mal condicionada. Os gráficos das figuras (1.1) e (1.2) representam o comportamento das curvas de nível da função (1.15). À medida que diminuimos o parâmetro τ , a curva de nível torna-se mais achatada, indicando o mau condicionamento da matriz Hessiana. Portanto, seria conveniente realizarmos boas estimativas da solução ótima quando os valores de τ não são críticos e conseqüentemente a matriz Hessiana é bem condicionada. Estas estimativas podem ser realizadas pelo método de extrapolação de Richardson-Romberg, como veremos, desde que certas condições sobre (1.1) sejam satisfeitas.

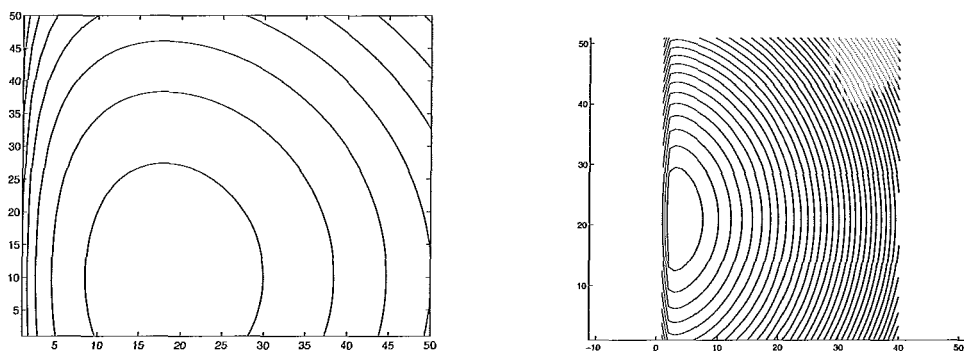


Figura 1.1: Curva de Nível com $\tau = 10$ e $\tau = 0.5$

Dessa forma, enunciaremos o teorema que mostra o crescente mau condicionamento da matriz Hessiana à medida que o parâmetro τ tende a zero, pois pode-se

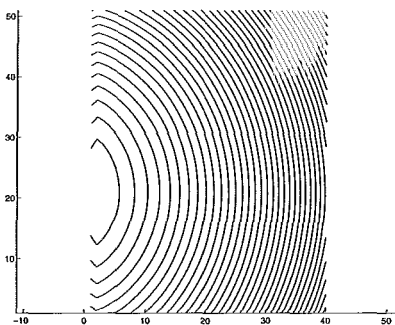


Figura 1.2: Curvas de Nível com $\tau = 0.1$

mostrar que alguns dos autovalores da matriz Hessiana da função objetivo modificada tendem ao infinito. Nessa seção faremos somente referência ao teorema enunciado por Xavier [22], pois a sua demonstração encontra-se no capítulo 2 na seção 2.4.

Teorema 1.2 *A matriz Hessiana da função objetivo modificada apresenta m^* (número de restrições ativas) autovalores infinitos quando $\tau \rightarrow 0$.*

Na seção seguinte, estudaremos o esquema de extrapolação criado por Richardson-Romberg no contexto da integração numérica. Desenvolvemos ainda uma aplicação bem simples com o objetivo de ressaltar os efeitos positivos, em termos de convergência à solução do problema, provocados pela esquema citado.

Realizamos o estudo do mecanismo de extrapolação primeiramente nos métodos de Integração Numérica, em virtude da extrapolação ter sido utilizada originalmente em conjunto com os métodos de integração. Podemos citar trabalhos importantes de aplicações do esquema de Richardson-Romberg para a resolução de problemas de integração, como problemas de integrais infinitas oscilatórias em Avram [15] e [16].

1.3 Integração Numérica

Os problemas de integração numérica do tipo

$$\int_{x=0}^{x=1} f(x)dx \tag{1.20}$$

aparecem freqüentemente em aplicações práticas. De forma geral, as técnicas de integração procuram representar a integral (1.20) através de uma soma ponderada

de valores amostrais da forma,

$$\int_{x=0}^{x=1} f(x)dx \approx \sum_i \omega_i f(x_i), \quad (1.21)$$

onde ω_i são os pesos, e x_i são os pontos amostrais e $f(x_i)$ são os valores amostrais da função. Diferentes métodos para a escolha dos pontos amostrais categorizam as várias técnicas. Assim, os pontos amostrais x_i podem ser escolhidos:

1. em intervalos regulares (Método do Trapézio);
2. fixando-se o número de pontos escolhidos para possibilitar uma alta precisão (Quadratura Gaussiana);
3. de acordo com o comportamento de $f(x_i)$ (Integração Adaptativa);
4. de maneira aleatória (Integração de Monte Carlo).

Dados então os pontos amostrais, a matemática da integração determina os pesos apropriados de tal forma que (1.21) forneça uma boa aproximação. Nesse contexto, surgem as aplicações do método de extrapolação, em particular, a extrapolação de Richardson-Romberg nos problemas de integração.

O método de Richardson-Romberg não descarta o uso das técnicas citadas acima. Pelo contrário, o método se utiliza de alguns pontos amostrais gerados por uma das técnicas citadas e gera um ponto extrapolado, que está fora do intervalo definido pelos pontos. Esse ponto extrapolado representa uma estimativa mais precisa do que as obtidas anteriormente, como mostraremos adiante. Intuitivamente, podemos dizer que carregamos algumas informações importantes sobre o comportamento da função através do método do Trapézio, por exemplo, e depois utilizamos estas

informações para gerar um ponto melhor que os obtidos anteriormente.

A obtenção do ponto extrapolado é realizada através do cálculo de um polinômio que passe pelos pontos amostrais obtidos e que representa a função original. Esse polinômio é obtido pelo uso da série de Taylor em torno de um ponto conveniente.

Uma característica destacável do método de Richardson-Romberg é a construção de um polinômio e avaliação desse em determinado ponto, sem, no entanto, precisar conhecer os coeficientes desse polinômio.

O esquema de Romberg apresenta um mecanismo recursivo que gera uma matriz triangular inferior conhecida como Matriz de Richardson-Romberg, que representa a construção de polinômios que irão gerar estimativas de ordens variadas, de acordo com o número de linhas e colunas dessa matriz.

Vejamos o polinômio:

$$q(h) = a_0 + a_1h + a_2h^2 + \dots + a_{n-1}h^{n-1} + O(h^n). \quad (1.22)$$

Definimos o coeficiente a_0 como:

$$a_0 = \int_{x=0}^{x=1} f(x)dx. \quad (1.23)$$

Consideremos então o problema de determinar a_0 como limite de uma quantidade $q(h)$ quando $h \rightarrow 0$, que pode ser calculada apenas para valores positivos de h ou mesmo apenas para um conjunto discreto de valores de h , tendo 0 como ponto de acumulação:

$$\lim_{h \rightarrow 0} q(h) = a_0 = \int_{x=0}^{x=1} f(x)dx. \quad (1.24)$$

Através do Método dos Trapézios podemos observar os valores trapezoidais abaixo que estimam o valor da integral para um número de intervalos n para um espaçamento h :

$$h \left\{ \frac{1}{2}f(0) + f(h) + f(2h) + \dots + f((n-1)h) + \frac{1}{2}f(nh) \right\}. \quad (1.25)$$

Então podemos ainda considerar a_0 como limite dos valores trapezoidais (1.25) para $h \rightarrow 0$, onde são escolhidos somente valores discretos $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots$ para h , sendo $nh = 1$. Assim, faremos uso do algoritmo de Romberg para determinar precisamente a_0 sem a necessidade de avaliar $q(h)$ para valores muito pequenos de h .

Método de Richardson-Romberg

O esquema de Richardson-Romberg foi proposto originalmente para acelerar a convergência dos métodos de integração numérica.

Descreveremos a seguir, o procedimento que mostra o processo de formação do esquema iterativo de Richardson-Romberg.

A expressão (1.24) estima o valor da integral (1.20), através do cálculo do limite de $q(h)$ quando $h \rightarrow 0$. Assim, ao invés de calcular explicitamente a integral, que pode ser de difícil resolução, determinamos $q(h)$ e calculamos o seu limite.

O procedimento para determinar $q(h)$ é válido sob a hipótese de que essa função admita uma série de potências, com $h > 0$ e $h \rightarrow 0$, pois dessa forma podemos representar $q(h)$ como:

$$q(h) = a_0 + a_1h + a_2h^2 + \dots + a_{n-1}h^{n-1} + O(h^n). \quad (1.26)$$

Os coeficientes a_0, a_1, \dots não são conhecidos, e como veremos adiante, o esquema de Richardson-Romberg torna-se mais atraente à medida que permite a avaliação da expressão (1.26) de maneira implícita, ou seja, a avaliação é realizada sem que seja necessário o cálculo desses coeficientes.

Considere ρ um número fixo, $0 < \rho < 1$. Em aplicações normalmente são utilizados os valores $\rho = \frac{1}{2}, \frac{1}{4}$ ou $\frac{1}{10}$. Vamos calcular $q(h)$ para $h = \rho^n$, $n = 0, 1, 2, \dots$. Definimos

$$q_{n,0} := q(\rho^n).$$

De (1.26) podemos obter uma expressão para $q_{n,0}$:

$$q_{n,0} \approx a_0 + a_1 \rho^n + a_2 \rho^{2n} + \dots \quad (1.27)$$

Da expressão acima, podemos ver que quando $n \rightarrow \infty$,

$$q_{n,0} - a_0 = O(\rho^n).$$

A partir da sequência $\{q_{n,0}\}$ definimos uma nova sequência $\{q_{n,1}\}$, que é de ordem 1, como uma combinação dos valores de ordem zero já calculados anteriormente.

Portanto, a expressão fica como:

$$q_{n,1} = \alpha_0 q_{n,0} + \beta_0 q_{n-1,0},$$

onde α_0 e β_0 são incógnitas e serão determinados tal que $q_{n,1}$ convirja para a_0 mais rápido.

Assumindo $a_1 \neq 0$, a expressão

$$q_{n,1} = \alpha_0 q_{n,0} + \beta_0 q_{n-1,0} \approx (\alpha_0 + \beta_0) a_0 + (\rho \alpha_0 + \beta_0) a_1 \rho^{n-1} + (\rho^2 \alpha_0 + \beta_0) a_2 \rho^{2n-2} + \dots$$

será válida, se e somente se

$$\alpha_0 + \beta_0 = 1, \quad (1.28)$$

$$\rho\alpha_0 + \beta_0 = 0. \quad (1.29)$$

Esse sistema tem como solução

$$\alpha_0 = \frac{1}{1 - \rho}, \quad (1.30)$$

$$\beta_0 = -\frac{\rho}{1 - \rho}. \quad (1.31)$$

Com essas escolhas de α_0 e β_0 obtemos,

$$q_{n,1} = \frac{q_{n,0} - \rho q_{n-1,0}}{1 - \rho}, \quad (1.32)$$

e portanto, vale a seguinte expansão assintótica:

$$q_{n,1} = a_0 + \sum_{k=2}^{\infty} \frac{1 - \rho^{1-k}}{1 - \rho} a_k \rho^{kn}, \quad n \rightarrow \infty. \quad (1.33)$$

Deve ser observado, em particular, que:

$$q_{n,1} - a_0 = O(\rho^{2n}).$$

Para obtermos estimativas de ordem mais elevada, podemos repetir o processo anterior através da formação da combinação linear

$$q_{n,2} = \alpha_1 q_{n,1} + \beta_1 q_{n-1,1}.$$

Ao eliminarmos o termo ρ^{2n} da expressão assintótica, podemos através de um cálculo similar ao realizado para $q_{n,1}$ mostrar que a equação

$$q_{n,2} = \frac{q_{n,1} - \rho^2 q_{n-1,1}}{1 - \rho^2}$$

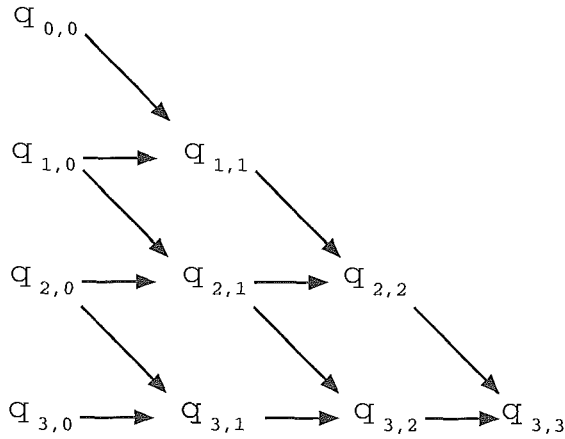


Figura 1.3: Matriz de Richardson-Romberg

satisfaz

$$q_{n,2} = a_0 + \sum_{k=3}^{\infty} \frac{(1 - \rho^{1-k})(1 - \rho^{2-k})}{(1 - \rho)(1 - \rho^2)} a_k \rho^{kn}, \quad n \rightarrow \infty.$$

Continuando a formação de $q_{n,3}$, $q_{n,4}, \dots$ podemos construir uma matriz triangular de números $q_{n,m}$ através das fórmulas:

$$q_{n,0} = q(\rho^n), \tag{1.34}$$

$$q_{n,m} = \frac{q_{n,m-1} - \rho^m q_{n-1,m-1}}{1 - \rho^m}, \tag{1.35}$$

$$n = 0, 1, 2, \dots$$

$$m = 1, 2, \dots, n.$$

A matriz da figura 1.3 é chamada Matriz de Romberg da função $q(h)$ calculada com razão ρ . As setas indicam o fluxo computacional. As entradas na primeira coluna são valores particulares de $q(h)$, e veremos mais tarde que esses valores serão gerados por um método de minimização. Os outros valores são calculados a partir de duas entradas imediatamente à esquerda por operações aritméticas triviais. Cada coluna da matriz pode ser formada tão logo seus elementos mais à esquerda

sejam calculados. O teorema a seguir conforme citado por Henrici [6] mostra-nos a validade do esquema de Romberg e pode ser provado por indução.

Teorema 1.3 *Para valores fixos $m = 0, 1, 2, \dots$, a seguinte expansão assintótica é verdadeira:*

$$q_{n,m} = a_0 + \sum_{k=m+1}^{\infty} \prod_{j=1}^m \frac{(1 - \rho^{j-k})(1 - \rho^{j-k})}{(1 - \rho^j)(1 - \rho^j)} a_k \rho^{kn}, \quad n \rightarrow \infty. \quad (1.36)$$

Corolário : Se todo $a_k \neq 0$ em (1.26), então cada coluna da Matriz de Romberg converge para a_0 , mais rápido que a coluna anterior.

O esforço computacional exigido para calcular colunas adicionais em (1.3) pode ser negligenciado se comparado com o esforço exigido para se calcular a primeira coluna. Portanto, este corolário é bastante útil para nossos propósitos numéricos. Contudo, a convergência de cada coluna é ainda linear; o erro tende a zero como γ^n , onde γ é uma quantidade fixa para cada coluna (apesar desse valor ficar menor a cada coluna sucessiva).

No exemplo abaixo ilustraremos o uso da extrapolação de Richardson-Romberg no âmbito de integração numérica. Para a obtenção dos valores da primeira coluna utilizamos o método do Trapézio dividindo a área abaixo da curva em 2, 4, 8, 16 e 32 partes, para valor de $\rho = 1/2$.

Considere o problema :

$$\int_0^1 \sin(x) dx \quad (1.37)$$

Como a função $\sin(x)$ pode ser obtida através de uma série de potências, utilizaremos o método de Richardson-Romberg descrito anteriormente para obtermos uma

estimativa do valor exato da integral. Em primeiro lugar, calculamos os valores particulares correspondentes a $q_{n,0}$ através do método do Trapézio. Esses valores estão representados na primeira coluna da matriz de Richardson-Romberg na figura (1.4). A partir dos valores dessa coluna, aplicamos o método iterativo de extrapolação e obtivemos os outros valores descritos nas colunas da matriz representada na figura (1.4).

```

0.450080520
0.457300940 0.46452136
0.459098970 0.46089700 0.459688880
0.459548040 0.45999711 0.459697147 0.459698328
0.459660280 0.45977252 0.459697657 0.459697729 0.459697689
0.459688340 0.45971640 0.459697693 0.459697699 0.459697696 0.459697697

```

Figura 1.4: Matriz de Richardson-Romberg

O resultado exato da integral (1.37) é 0.459697694 . Através da análise da matriz 1.4 obtemos estimativas de ordem linear, quadrática até a estimativa de quinta ordem. Esta estimativa de mais alta ordem é representada por $a_0 = 0.459697697$ como a melhor estimativa para o valor exato da integral, o que nos fornece um erro relativo de aproximadamente $7.0e - 9$.

A solução obtida por extrapolação é uma estimativa de quinta ordem, já que o ponto foi retirado da última coluna, pois o corolário nos garante que as melhores estimativas são dadas pelos polinômios de maior grau.

Se tivéssemos utilizado o método do Trapézio em sua forma ortodoxa, iríamos obter o resultado 0.459697692 que apresenta mesma precisão do valor obtido por Richardson-Romberg, somente para $\rho = \frac{1}{2^{12}}$, que é 128 vezes menor que a

estimativa realizada com a extrapolação quando $\rho = \frac{1}{2^5}$. Portanto, com um valor relativamente alto $\rho = \frac{1}{2^5}$, conseguimos realizar uma boa estimativa para (1.37).

Esse pequeno exemplo ilustra que o esquema de Richardson-Romberg é um mecanismo eficiente para realizar estimativas baseadas em informações passadas obtidas em um método numérico iterativo.

Capítulo 2

Penalização Hiperbólica

Neste capítulo faremos a apresentação do método de Penalização Hiperbólica, proposto por Xavier [23] para resolver o problema geral de programação não-linear sujeito unicamente a restrições de desigualdade:

$$\begin{aligned} & \text{Minimizar } f(x) \\ & \text{sujeito a } g_i(x) \geq 0, \\ & \quad i = 1, \dots, m \end{aligned} \tag{2.1}$$

onde $x \in R^n$.

O método proposto pertence à família dos métodos de penalidades e transforma o problema (2.1) em problemas irrestritos. Ou seja, a solução é obtida através da resolução de uma sequência de subproblemas sem restrições , cujos valores obtidos convergem para o ótimo do problema original.

A Penalização Hiperbólica é objeto importante no presente estudo de extrapolação, já que este método foi utilizado como base para a obtenção dos mínimos que formam a primeira coluna da matriz de Richardson-Romberg.

A função penalidade que dá origem ao método é a seguinte :

$$P(y, \lambda, \tau) = -\lambda y + \sqrt{\lambda^2 y^2 + \tau^2}, \quad (2.2)$$

onde $\lambda = \frac{1}{2} \tan(\frac{\alpha}{2})$, $\alpha \in (0, \frac{\pi}{2})$ e $\tau \geq 0$.

O gráfico da função penalidade (2.2) para valores fixos de λ e τ é dado pela figura 2.1 abaixo :

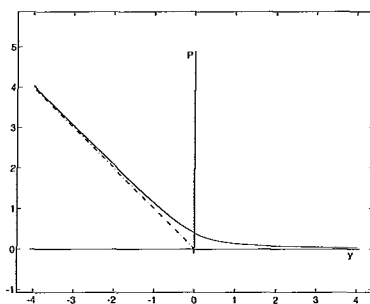


Figura 2.1: Função Penalidade Hiperbólica

Considerando y como o valor de uma restrição, ou seja, $y = g_i(x)$, passamos ao valor associado a essa restrição:

$$P(g_i(x), \lambda, \tau) = -\lambda g_i(x) + \sqrt{\lambda^2 g_i(x)^2 + \tau^2}. \quad (2.3)$$

Desta forma, transformando o problema original (2.1) num problema irrestrito considerando a função de penalidade definida em (2.3) para cada uma das m restrições, temos o problema penalizado como:

$$\text{minimizar } F(x, \lambda, \tau) = f(x) + \sum_{i=1}^m P(g_i(x), \lambda, \tau). \quad (2.4)$$

Analisando a função penalidade (2.3) podemos ver inicialmente que:

Em $\tau = 0$ e $g_i(x) \geq 0$: $P(g_i(x), \lambda, \tau) = 0$, o que significa que em $\tau = 0$, a penalização é nula.

Em $g_i(x) \geq 0 : P(g_i(x), \lambda, \tau) \leq \tau$, o que mostra que para $y = g(x) \geq 0$, ou seja, para pontos viáveis, a penalização passa a ser limitada superiormente pelo parâmetro τ . Então, à medida que os valores de τ decrescem monotonicamente para zero, a penalização é reduzida gradativamente para zero para pontos viáveis.

O método de Penalização Hiperbólica conjuga penalização externa com penalização interna, através da manipulação de dois parâmetros distintos λ e τ . O parâmetro λ está relacionado à variação do ângulo α e representa a medida da penalização exterior; o parâmetro τ está relacionado ao comprimento da ordenada na origem, ou seja, está associado a uma medida da penalidade interna.

Dessa forma, o método de Penalização Hiperbólica é composto por duas fases distintas. A primeira fase consiste na obtenção de um ponto viável através do aumento do parâmetro de penalização exterior λ , conforme representado pelo gráfico (2.2) na figura 2a.

Na segunda fase mantém-se constante a penalização exterior e gradativamente diminuímos o parâmetro de penalização interior τ , conforme representado pelo gráfico (2.2) na figura 2b.

Como o método é dividido em duas etapas distintas, os parâmetros são manipulados separadamente em cada uma das fases e, por esse motivo, a existência dos dois parâmetros não provoca grandes dificuldades no processo de resolução do problema.

Como característica importante do método, podemos destacar a liberdade de escolha de um ponto inicial qualquer, sem exigência de viabilidade, pois essa é

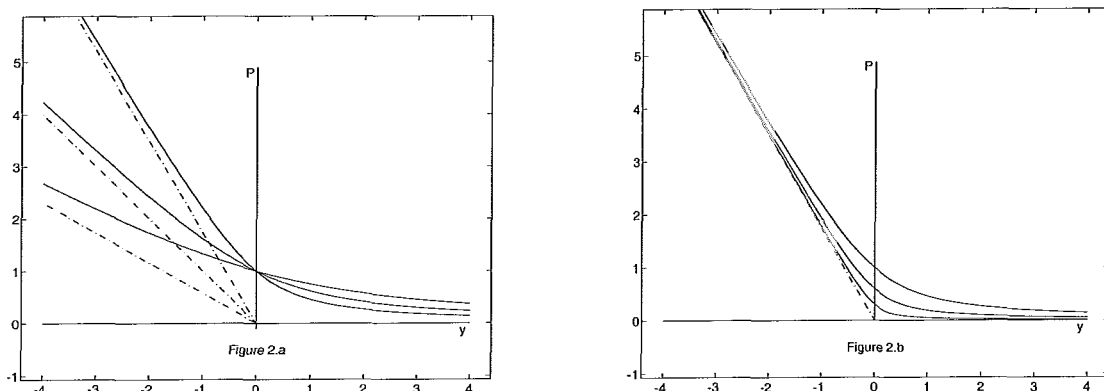


Figura 2.2: Primeira e Segunda Fase de P.H.

tratada na primeira fase do método. A propriedade mais marcante do método é, entretanto, a contínua diferenciabilidade da função penalidade proposta (2.3).

2.1 Análise de Convexidade

É especialmente importante a definição de conjuntos convexos já que, freqüentemente, aparecem na prática problemas de otimização com domínio convexo. Desta forma, analisaremos a convexidade da função penalidade e do problema irrestrito (2.4).

Uma importante classe dos problemas de programação matemática é a classe dos problemas de programação convexa. Essa classe corresponde aos problemas (2.1) onde a função objetivo é convexa e as funções $g_i(x)$ são côncavas ($-g_i(x)$ são convexas).

Podemos observar que a função $P(y, \lambda, \tau)$ é convexa em y . Pode-se adicionalmente provar que a função penalidade hiperbólica em $y = g_i(x)$ côncava

ainda é uma função convexa em x .

Portanto, supondo 2.1 um problema de programação convexa, como a função penalidade é convexa em x , o problema irrestrito (2.4) será convexo. Assim, podemos dizer que a função penalidade conserva a natureza convexa do problema original (2.1).

Os problemas convexos apresentam características importantes como, qualquer mínimo local é mínimo global, ademais, se a função objetivo for estritamente convexa teremos certamente unicidade de solução.

Veremos mais adiante, que o mecanismo de extrapolação pressupõe a convexidade do problema.

Estudaremos agora, as principais propriedades da função penalidade hiperbólica, pois o método de Penalização Hiperbólica está consolidado para toda a classe de funções que obedecem as propriedades descritas na seção seguinte. Essas propriedades indicam as características de continuidade e diferenciabilidade da função penalidade hiperbólica e sobretudo o comportamento da função à medida que os parâmetros λ e τ sofrem variação.

2.2 Propriedades da Função Penalidade

O método de Penalização Hiperbólica está consolidado não só para a função apresentada em (2.2), mas para as funções de penalização que apresentem as propriedades abaixo :

P01 - $P(y, \lambda, \tau)$ é uma função contínua, bem como, continuamente diferenciável em y para valores de $0 < \alpha < \pi/2$, $\tau > 0$.

P02 - $P(y, \lambda, \tau)$ é assintoticamente tangente às retas $r_1(y) = -y \cdot \tan(\alpha)$ e $r_2(t) = 0$ para $\lambda > 0$.

P03 - $\lim_{y \rightarrow +\infty} P(y, \lambda, \tau) = 0$ para $\lambda \geq 0$ e $0 < \alpha < \pi/2$.
 $\lim_{y \rightarrow -\infty} P(y, \lambda, \tau) = +\infty$ para $\tau \geq 0$, $0 < \alpha < \pi/2$.

P04 - $P(0, \lambda, \tau) = \tau$, para $\tau \geq 0$ e $0 < \alpha < \pi/2$.

P05 - $P(y, \lambda_{k+1}, \tau) < P(y, \lambda_k, \tau)$, para $y > 0$,

$$P(y, \lambda_{k+1}, \tau) = P(y, \lambda_k, \tau), \text{ para } y = 0,$$

$$P(y, \lambda_{k+1}, \tau) > P(y, \lambda_k, \tau), \text{ para } y < 0,$$

onde $\tau > 0$ e $0 < \alpha^k < \alpha^{k+1} < \pi/2$.

P06 - $P(y, \lambda_{k+1}, \tau)$ é uma função convexa e decrescente em y para $\tau > 0$ e $0 < \alpha < \pi/2$ e é uma função convexa e não crescente em y para $\tau = 0$ e $0 < \alpha < \pi/2$.

P07 - $P(y, \lambda, \tau_{k+1}) < P(y, \lambda, \tau_k)$ para $\forall y$, $0 < \alpha < \pi/2$ e $0 \leq \tau_{k+1} < \tau_k$.

P08 - $\lim_{\tau \rightarrow 0} P(y, \lambda, \tau) = \left\{ \begin{array}{ll} 0 & : y \geq 0 \\ -y \tan(\alpha) & : y < 0 \end{array} \right\}$, onde $0 < \alpha < \pi/2$.

P09 - $P(y, \lambda, \tau) \geq -y \tan(\alpha)$ para $\forall y$, $0 < \alpha < \pi/2$.

P10 - $|P(y, \lambda, \tau_{k+1}) - P(t, \lambda, \tau_{k+1})| < |P(y, \lambda, \tau_{k+1}) - P(t, \lambda, \tau_{k+1})|$

P11 - $P(g(x), \lambda, \tau)$ é uma função convexa em x se $g(x)$ for uma função côncava.

P12 - $Max_y P(y, \lambda, \tau_o) - P(y, \lambda, \tau_1) = \tau_o - \tau_1$ e ocorre em $y = 0$ para $0 < \alpha < \pi/2$ e $0 \leq \tau_1 \leq \tau_o$.

P13 - A derivada da função penalidade em relação a y , ou seja, $P'_y(y, \lambda, \tau)$ é uma função decrescente com τ para pontos $y > 0$ (e é uma função crescente com τ para pontos $y < 0$).

A função (2.2) evidentemente satisfaz as 13 propriedades enunciadas acima.

2.3 Multiplicadores de Lagrange

A partir da resolução do problema de otimização através do método de Penalização Hiperbólica podemos obter estimativas dos multiplicadores de Lagrange, conforme Xavier [20]. Essas estimativas dos multiplicadores pode ser bastante útil para acelerar a convergência, pois a partir de uma base de multiplicadores, se pudermos aplicar o mecanismo de extrapolação obteremos uma boa estimativa dos multiplicadores, estimativa esta, bem mais próxima da solução ótima. Assim, na resolução do problema pode-se utilizar conjuntamente o mecanismo de extrapolação nas variáveis do espaço primal e do espaço dual. Conjectura-se que esta estratégia venha acelerar a convergência de um método que faça uso dos multiplicadores de Lagrange, como o método Lagrangeano Hiperbólico em Xavier [20].

Similarmente ao apresentado para o método de Barreira Logarítmica no capítulo 1, o método de Penalização Hiperbólica também gera um conjunto de valores que convergem para os Multiplicadores de Lagrange.

No processo iterativo de resolução de cada subproblema (2.4), realizam-se estimativas dos Multiplicadores de Lagrange, pois ao minimizar $F(x, \lambda_k, \tau_k)$ obtemos

o ponto $x(\tau_k)$ que satisfaz:

$$\nabla f(x(\tau_k)) + \sum_{i=1}^m P'(g_i(x(\tau_k)), \lambda_k, \tau_k) \nabla g_i(x(\tau_k)) = 0, \quad (2.5)$$

logo, comparando-se (2.5) com (1.7) obtemos paralelamente:

$$\mu_i(\tau_k) = -P'(g_i(x), \lambda, \tau_k), \quad i = 1, \dots, m. \quad (2.6)$$

Conseqüentemente,

$$\mu_i(\tau_k) = \lambda_k \left[1 - \frac{\lambda_k g_i(x^k)}{\sqrt{\lambda_k^2 g_i^2(x(\tau_k)) + \tau_k^2}} \right], \quad (2.7)$$

ou ainda podemos escrever como:

$$\mu_i(\tau_k) = \lambda_k \left[1 - \frac{1}{\sqrt{1 + \frac{\tau_k^2}{\lambda_k^2 g_i^2(x(\tau_k))}}} \right]. \quad (2.8)$$

Podemos verificar que $0 \leq \mu_i \leq \lambda_k$. Em seguida, provaremos que $\mu_i(\tau_k)$ converge para μ_i^* .

Teorema 2.1 *Se a seqüência x^k gerada pelo método converge para x^* , uma solução regular para o problema (2.1), então os valores μ_i^k definidos por $\mu_i(\tau_k) = -P'(g_i(x), \lambda, \tau_k)$ convergem para os multiplicadores de Lagrange correspondentes à solução ótima, μ_i^* .*

Demonstração: Se x^k é por hipótese o minimizador do problema de irrestrito (2.4), então:

$$\nabla_x F(x^k, \lambda_k, \tau_k) = \nabla f(x^k) + \sum_{i=1}^m P'(g_i(x^k), \lambda_k, \tau_k) \nabla g_i(x^k) = 0. \quad (2.9)$$

Tomando o limite da seqüência x^k que por hipótese converge para x^* , temos:

$$\nabla f(x^*) + \lim_{k \rightarrow \infty} \sum_{i=1}^m P'(g_i(x^k), \lambda_k, \tau_k) \nabla g_i(x^*) = 0. \quad (2.10)$$

Como por hipótese x^* é um ponto regular, pelas condições estabelecidas por Karush-Kuhn-Tucker existe multiplicador de Lagrange $\mu^* \geq 0$ tal que:

$$\nabla f(x^*) - \sum_{i=1}^m \mu_i^* \nabla g_i(x^*) = 0. \quad (2.11)$$

Subtraindo-se as equações (2.10) e (2.11) temos:

$$\sum_{i=1}^m \left(\lim_{k \rightarrow \infty} P'(g_i(x^k), \lambda_k, \tau_k) + \mu_i^* \right) \nabla g_i(x^*) = 0. \quad (2.12)$$

Pelas condições necessárias de Karush-Kuhn-Tucker, temos $\mu_i^* = 0$ para todo $i \notin I(x^*)$. Quando $k \rightarrow \infty$, temos $\tau_k \rightarrow 0$ e λ_k constante na segunda fase do método de Penalização Hiperbólica. Portanto, verificamos pela expressão analítica da penalidade hiperbólica (2.3), que $P'(g_i(x^*), \lambda_k, 0) = 0$ para todo índice i em que $g_i(x^*) > 0$.

Desta forma a equação (2.12) se reduz a:

$$\sum_{i \in I(x^*)} \left(\mu_i^* + \lim_{k \rightarrow \infty} P'(g_i(x^k), \lambda_k, \tau_k) \right) \nabla g_i(x^*) = 0. \quad (2.13)$$

Como o ponto x^* é regular, os gradientes $\nabla g_i(x^*)$ são linearmente independentes e os coeficientes que os multiplicam devem ser nulos para que a igualdade da expressão (2.13) seja verdadeira. Isto implica que:

$$\mu_i^* = - \lim_{k \rightarrow \infty} P'(g_i(x^k), \lambda_k, \tau_k) = \lim_{k \rightarrow \infty} \mu_i^k. \quad \blacksquare \quad (2.14)$$

Com a obtenção de uma seqüência de multiplicadores obtidos por estimativas podemos formar uma base de Multiplicadores de Lagrange, que tendem ao multiplicador ótimo. Estudaremos desta forma, o estabelecimento das condições teóricas que viabilizem a utilização dessa base de multiplicadores para traçar, pela expansão

em série de Taylor, um polinômio que passe pelos multiplicadores gerados. Esses polinômios podem ser obtidos pelo mecanismo de extrapolação não somente para as variáveis primais, como também para as variáveis duais.

2.4 Análise da Hessiana

A questão da matriz Hessiana é um ponto crítico, já que nos métodos de penalidades interior e exterior a variação do parâmetro de penalidade provoca instabilidade à medida que o processo de minimização avança. Essa instabilidade é medida pelo mau condicionamento da Hessiana, que implica na degenerescência do problema. É bem verdade que, nesses métodos, a questão da degenerescência é inevitável, como demonstrado no exemplo citado no capítulo anterior, pois à medida que o parâmetro de controle tende a infinito, m^* valores próprios desta matriz tendem a infinito, sendo m^* igual ao número de restrições ativas no ponto ótimo do problema.

Esse mau condicionamento da Hessiana, como é sabido, pode dificultar, ou até mesmo impedir, o sucesso da minimização sem restrições. Como registrado por Avriel [2], os mais eficientes métodos para minimização sem restrições, que pertencem ao grupo dos métodos tipo Newton, Gradientes Conjugados e Métrica Variável, ironicamente são os mais vulneráveis à principal desvantagem apresentada pelos métodos das penalidades interiores e exteriores.

Dessa forma, o processo de minimização precisa ser interrompido antes que o condicionamento da matriz Hessiana seja problemático e gere uma solução insatisfatória e longe do ótimo, pois, como observa Luenberger [9], a exploração efetiva desses métodos requer que sejam inventados esquemas especiais que eliminem o

efeito provocado pelos autovalores infinitos da matriz Hessiana.

Semelhantemente, o método de Penalização Hiperbólica, como mostraremos logo a seguir, também apresenta a matriz Hessiana da função objetivo modificada com m^* autovalores infinitos conforme descrito no teorema abaixo citado por Xavier [22]. Sendo assim, o método proposto enfrenta as mesmas dificuldades descritas anteriormente para os outros métodos.

Teorema 2.2 *A matriz Hessiana da função objetivo modificada apresenta m^* (número de restrições ativas) autovalores infinitos quando $\tau \rightarrow 0$.*

Demonstração:

A matriz Hessiana, para o nosso caso, é obtida derivando-se a expressão (3.1) em relação a x :

$$\nabla_x^2 F(x, \lambda, \tau) = \nabla_x^2 f(x) + \sum_i P'(g_i(x), \lambda, \tau) \nabla_x^2 g_i(x) + \sum_i P''(g_i(x), \lambda, \tau) \nabla g_i(x) \nabla^t g_i(x).$$

onde $\nabla_x^2 f(x)$ e $\nabla_x^2 g_i(x)$ representam respectivamente as Hessianas de $f(x)$ e de $g_i(x)$. A última parcela pode ser substituída pelo produto entre as matrizes $\nabla g(x) R \nabla^t g(x)$, onde $\nabla g(x)$ é uma matriz com n linhas e m colunas e R é uma matriz diagonal de ordem m cujo elemento da diagonal principal na posição i (r_{ii}) é dado por:

$$P''(g_i(x), \lambda, \tau) = \frac{\tau^2}{(\lambda^2 g_i^2(x) + \tau^2)^{\frac{3}{2}}}. \quad (2.15)$$

Na seqüência de minimizações sem restrições, vamos ver o que acontece com este

termo quando $\tau \rightarrow 0$ e a restrição é ativa, ou seja, $g_i(x(\tau)) \rightarrow 0$.

$$\lim_{\tau \rightarrow 0} P''(g_i(x(\tau)), \lambda, \tau) = \lim_{\tau \rightarrow 0} \frac{\tau^2}{(\lambda^2 g_i^2(x(\tau)) + \tau^2)^{\frac{3}{2}}} \quad (2.16)$$

$$= \lim_{\tau \rightarrow 0} \frac{\frac{1}{(\frac{\lambda^2 g_i^2(x(\tau))}{\tau^2} + 1)}}{\lambda g_i(x(\tau)) (1 + \frac{\tau^2}{\lambda^2 g_i^2(x(\tau))})^{\frac{1}{2}}}. \quad (2.17)$$

Vejamos novamente a expressão dos multiplicadores definida em (2.7):

$$\mu_i(\tau_k) = \lambda_k \left[1 - \frac{\lambda_k g_i(x^k)}{\sqrt{\lambda_k^2 g_i^2(x(\tau_k)) + \tau_k^2}} \right]. \quad (2.18)$$

Observamos que como os multiplicadores $\mu_i^* > 0$, para todo índice $i \in I(x)$, podemos verificar que $\lim_{\tau \rightarrow 0} \frac{\tau^2}{g_i^2(x)} = s_i^{*2}$ tem necessariamente valor não nulo. Caso contrário, o multiplicador se anularia.

Portanto,

$$\lim_{\tau \rightarrow 0} \frac{g_i^2(x)}{\tau^2} = \frac{1}{s_i^{*2}} = \omega_i^{*2} < \infty. \quad (2.19)$$

Assim, como para as restrições ativas $\lim_{\tau \rightarrow 0} g_i(x(\tau)) = 0$, temos:

$$\lim_{\tau \rightarrow 0} P''(g_i(x), \lambda, \tau) = \frac{\frac{1}{\lambda^2 \omega_i^{*2} + 1}}{\lambda g_i(x(\tau)) (1 + \frac{1}{\lambda^2 \omega_i^{*2}})^{\frac{1}{2}}} = \infty. \quad (2.20)$$

Deste modo, vimos que a matriz $R(x(\tau))$ possui m^* valores próprios que tendem a infinito quando $\tau \rightarrow 0$. Isto leva a matriz Hessiana $\nabla^2 F(x(\tau))$ possuir idêntica característica. ■

2.5 Soluções do problema irrestrito penalizado

O problema irrestrito (2.4) é resolvido através da resolução de uma seqüência de subproblemas penalizados.

Cada subproblema resolvido gera uma solução $x(\lambda_k, \tau_k)$ e uma estimativa do multiplicador de Lagrange $\mu(\lambda_k, \tau_k)$. Na segunda fase do algoritmo os pontos dependem exclusivamente de τ_k , já que λ_k é mantido constante. Então por simplicidade de notação, faremos referência aos sucessivos pontos de mínimo como $x(\tau_k), x(\tau_{k+1}), \dots$ e aos multiplicadores como $\mu(\tau_k), \mu(\tau_{k+1}), \dots$. Assim como os métodos de Barreira, a Penalização Hiperbólica, apesar de retardar um pouco mais a degenerescência da Hessiana, mantêm vivas as dificuldades numéricas à medida que o parâmetro $\tau \rightarrow 0$, como vimos na seção anterior.

Desta forma, justificamos o estudo e uso do mecanismo de extrapolação para, a partir de um conjunto de sucessivos minimizadores $x(\tau_1), \dots, x(\tau_k)$, realizarmos boas estimativas para o ótimo x^* e evitarmos assim a resolução do subproblema irrestrito quando τ está bem próximo de zero.

Além disso, iremos ainda, a partir de um conjunto de multiplicadores de Lagrange $\mu(\tau_1), \dots, \mu(\tau_k)$, estudar a aplicação da extrapolação sobre o conjunto de multiplicadores para realizarmos estimativas do ótimo μ^* .

Veremos no capítulo 3 o embasamento teórico que viabiliza a aplicação deste mecanismo nos problemas de otimização.

2.6 Algoritmo de Penalização Hiperbólica

Uma alternativa de implementação do método de Penalização Hiperbólica segue os seguintes passos:

- 1) Faça $k = 0$, $\lambda_1 = \lambda_0$, $\tau_1 = \tau_0$, sendo $0 < \alpha < \pi/2$ e $\lambda_0 > 0$. Tome ponto inicial x^0 .
- 2) Teste se x^k é viável. Se for, vá para o passo 6.
- 3) Resolva problema de minimização sem restrições da função: $F(x, \lambda_k, \tau_k) = f(x) + \sum_{i=1}^m P(g_i(x), \lambda_k, \tau_k)$ a partir de um ponto inicial x^{k-1} achando o ponto ótimo x^k .

4) Teste se x^k é viável. Se esse é viável vá para o passo 6.

5) Faça a Fase 1 :

$$\lambda_{k+1} = \rho\lambda_k + (1 - \rho) \pi/2, \quad 0 < \rho < 1$$

e $\tau_{k+1} = \tau_k$. Vá para o passo 2.

6) Regra de Parada : Verificar se x^k é aceitável como solução. Se for aceitável, vá para o passo 8.

7) Faça a Fase 2:

$$\tau_{k+1} = \rho\tau_k, \quad 0 < \rho < 1$$

$$\lambda_{k+1} = \lambda_k.$$

Vá para o passo 2.

8) A solução é x^k . Fim.

Capítulo 3

Extrapolação em Penalização Hiperbólica

A idéia básica do mecanismo de extrapolação está perfeitamente contida na acepção principal deste termo: determinar ou prever eventos ou comportamentos futuros calcados em episódios ou fatos passados. Numericamente, a extrapolação passa pelo cálculo de um polinômio que passe exatamente por k pontos dados. Dessa forma, o passado é explicado perfeitamente pelo polinômio. A extrapolação fundamenta-se na crença da perseverança futura do comportamento observado no passado. Nesse sentido, a extrapolação equivale à interpolação, entretanto utilizaremos o polinômio encontrado através da extrapolação para fazer estimativas de pontos que estejam fora do intervalo definido pelos k pontos fornecidos inicialmente.

Como descrito anteriormente, o mecanismo de extrapolação de Richardson-Romberg utiliza-se de k ($k \geq 2$) pontos, para gerar polinômios que passem exatamente por esses pontos. Assim, para k pontos conhecidos, podemos gerar $k - 1$ polinômios $p_1, p_2, p_3, \dots, p_{k-1}$ com grau, respectivamente, $1, 2, \dots, k - 1$ passando, respectivamente, pelos últimos $2, \dots, k$ pontos. Adiantamos que esses

polinômios são da forma $x = x(\tau)$.

Ao gerarmos vários polinômios, a motivação é de que possamos escolher o polinômio que melhor modele a trajetória central em direção ao ponto ótimo x^* , já que avaliando-se os polinômios obtêm-se estimativas da solução ótima. Adiantamos que essa escolha é determinada de acordo com a avaliação da função irrestrita no ponto \hat{x}^* gerado pelo polinômio. O ponto que gera menor valor de função objetivo é tomado como estimativa do ótimo e conseqüentemente o polinômio que empiricamente o gerou é tomado como o mais adequado naquele subproblema.

Neste trabalho, os k pontos iniciais são gerados através da resolução do problema irrestrito (2.4) pelo método de Penalização Hiperbólica. Portanto, para $k = 2$, podemos gerar somente uma extrapolação linear, onde os dois mínimos são gerados pela minimização de (2.4) variando-se o parâmetro τ . Dessa forma, construímos a matriz de extrapolações de Richardson-Romberg, semelhante a apresentada na figura (1.3), onde a cada obtenção de um novo minimizador, atualiza-se a matriz acrescentando-se uma nova linha.

É fato que a cada iteração usamos τ_k conhecido e obtemos através do método de minimização irrestrita $x(\tau_k) = x^k$, que é o ótimo na iteração k . Geralmente, esse ponto é usado como ponto inicial para a resolução do próximo subproblema penalizado que gera o ponto x^{k+1} . Com a aplicação do mecanismo de extrapolação, podemos ainda realizar boas estimativas para o ótimo da iteração seguinte $k + 1$. Essa estimativa, é realizada utilizando-se a mesma trajetória (polinômio) obtida para estimar x^* . Esse mecanismo é denominado neste trabalho de *extrapolação inversa* porque corresponde a uma seqüência de cálculos na matriz de Richardson-Romberg

na ordem inversa. A expectativa é de que \hat{x}^{k+1} esteja bem próximo ao ótimo desse subproblema, o que deve provocar, em princípio, uma redução no número de passos para atingir o ponto ótimo do $(k + 1)$ -ésimo subproblema (x^{k+1}).

Portanto, no presente estudo, analisaremos a extrapolação de Richardson-Romberg, sob dois aspectos práticos:

- Mecanismo para estimar o ponto ótimo do problema original (\hat{x}^*);
- Mecanismo para estimar o ponto de ótimo do próximo subproblema penalizado na iteração $k + 1$ (\hat{x}^{k+1});

As duas estratégias descritas acima de extrapolação serão conjugadas neste trabalho, firmadas em uma base teórica que garanta o uso do mecanismo sob certas condições. Abaixo, são descritas as condições a serem impostas sobre o problema (1.1) para o uso da extrapolação.

C1 - f e $-g_i$, $i = 1, \dots, m$ são funções convexas e continuamente diferenciáveis;

C2 - $F(x, \lambda, \tau)$ é estritamente convexa para $\lambda \geq 0$;

C3 - x^* é ponto regular;

C4 - x^* é solução não-degenerada.

Desta forma, para os problemas de otimização que satisfazem às condições acima, podemos garantir o uso do mecanismo de extrapolação de Richardson-Romberg.

Vale ressaltar que, conforme descrito por Fiacco [3], para que **C2** seja válida basta que a função f ou qualquer uma das funções $g_i(x)$, $i = 1, \dots, m$

seja estritamente convexa, ou então, que haja n ou mais restrições afins com gradientes formando base para o R^n .

Nas seções seguintes, desenvolveremos todo o embasamento teórico que garanta o sucesso do método de extrapolação em Penalização Hiperbólica.

3.1 Desenvolvimento teórico

A teoria garante o sucesso do método de extrapolação se as seguintes propriedades são satisfeitas:

I1 - podemos escrever x como um polinômio em função de τ ;

I2 - existe a derivada de primeira ordem de x em função de τ , $\frac{dx(\tau)}{d\tau}$, e ela é finita;

I3 - existe a derivada de ordem n de $x(\tau)$ em função de τ , $\frac{d^n x(\tau)}{d\tau^n}$;

I4 - no ponto $\tau = 0$ todas as propriedades anteriores [I1],[I2] e [I3] são válidas.

Será mostrado a seguir, que essas propriedades são válidas no presente caso.

Dado um conjunto de parâmetros τ_1, \dots, τ_k , considerando-se somente a segunda fase do método de Penalização Hiperbólica, onde λ é constante, obtemos uma seqüência de minimizadores $x(\tau_1), x(\tau_2), \dots, x(\tau_k)$. Essa seqüência será a base para a aplicação do esquema de extrapolação.

A idéia do mecanismo de Richardson-Romberg é escrever um polinômio $x(\tau)$ através da expansão em série de Taylor de x em torno de $x^* = x(0)$. Como

a expansão em série exige que a trajetória seja continuamente diferenciável nesse ponto, precisamos verificar todas as propriedades descritas acima.

Mostraremos na seção seguinte, através do uso do Teorema da Função Implícita (TFI), que é possível escrever x como uma função de τ , ou seja, é possível obter a trajetória $x(\tau)$.

3.1.1 Existência da trajetória $x(\tau)$

Ao longo da minimização irrestrita, para uma sequência de valores de τ , obtemos um conjunto de minimizadores $x(\tau)$, que delineam uma trajetória.

Inicialmente, podemos pensar em modelar esta trajetória através de um polinômio, utilizando a sequência de minimizadores já obtida ao longo do processo de minimização.

Porém, para construir este polinômio, precisamos garantir que x pode ser escrito em função do parâmetro de penalização interior τ e garantir a existência das derivadas $\frac{dx(\tau)}{d\tau}$ no ponto limite $\tau = 0$.

Dessa forma, o Teorema da Função Implícita (TFI) é ponto central nesse trabalho para todo o desenvolvimento da base teórica que garante a viabilidade do mecanismo de extrapolação, pois sob certas condições, o teorema garante que podemos escrever x como função de τ .

Vejamos o enunciado do teorema extraído de Luenberger [9]:

Teorema 3.1 Teorema da Função Implícita

Considere o ponto $x^0 = (x_1^0, x_2^0, x_3^0, \dots, x_n^0)$ como um ponto em E^n satisfazendo as seguintes propriedades:

i) As funções $h_i \in C^p$, $i = 1, \dots, m$ em alguma vizinhança de x^0 , para algum $p \geq 1$.

ii) $h_i(x^0) = 0$, $i = 1, \dots, m$.

iii) A matriz Jacobiana de ordem m

$$J = \begin{bmatrix} \frac{\partial h_1(x^0)}{\partial x_1} & \dots & \frac{\partial h_1(x^0)}{\partial x_m} \\ \vdots & & \vdots \\ \frac{\partial h_m(x^0)}{\partial x_1} & \dots & \frac{\partial h_m(x^0)}{\partial x_m} \end{bmatrix},$$

é não singular.

Então existe uma vizinhança de $\hat{x}^0 = (x_{m+1}^0, x_{m+2}^0, \dots, x_n^0) \in E^{n-m}$ tal que, para $\hat{x} = (x_{m+1}, x_{m+2}, \dots, x_n)$ nesta vizinhança, existem funções $\phi_i(\hat{x})$, $i = 1, \dots, m$ tal que

i) $\phi_i \in C^p$;

ii) $x_i^0 = \phi_i(\hat{x}^0)$, $i = 1, \dots, m$;

iii) $h_i(\phi_1(\hat{x}), \dots, \phi_m(\hat{x}), \hat{x}) = 0$, $i = 1, \dots, m$.

Agora, analisaremos o resultado do uso do TFI para o nosso caso.

Assumimos $F(x, \lambda, \tau)$ convexa e contínua no intervalo definido de acordo com

as condições **C1** e **C2** . Logo, $F(x, \lambda, \tau)$ possui ponto de mínimo (\tilde{x}, τ) que satisfaz a expressão:

$$\nabla_x F(\tilde{x}, \tau) = \nabla f(\tilde{x}) + \sum_{i=1}^m P'(g_i(\tilde{x}), \lambda, \tau) \cdot \nabla g_i(\tilde{x}) = 0. \quad (3.1)$$

Chamemos $G(x, \tau) = \nabla_x F(x, \tau)$. Em face das condições de continuidade de **C1**, $G(x, \tau)$ é uma função contínua com derivadas contínuas e é anulada em (\tilde{x}, τ) . Como $F(x, \lambda, \tau)$ é uma função estritamente convexa, por **C2**, então ∇G é estritamente definida positiva e portanto é não singular.

Desse modo, todas as hipóteses do Teorema da Função Implícita são válidas em nosso estudo, portanto poderemos, como formalizado a seguir, utilizar o resultado do teorema enunciado, escrevendo x como função de τ , ou seja, $x = x(\tau)$. Além disso, poderemos garantir que $x(\tau)$ é diferenciável.

3.1.2 Cálculo de $\frac{dx(\tau)}{d\tau}$

Já provamos na seção anterior que o valores de x descrevem uma trajetória em função do parâmetro τ . Nesta seção nosso objetivo é obter uma expressão analítica para a derivada de primeira ordem de x , $\frac{dx(\tau)}{d\tau}$, que será útil em nossos propósitos de mostrar que existe um polinômio que descreve a trajetória $x(\tau)$.

Consideremos ainda o sistema descrito pela expressão (3.1). Derivando-a em relação à τ obtemos:

$$\left\{ \nabla^2 f(x) + \sum_{i=1}^m P''(g_i(x), \lambda, \tau) \cdot \nabla g_i(x) \cdot \nabla^t g_i(x) + \sum_{i=1}^m P'(g_i(x), \lambda, \tau) \cdot \nabla^2 g_i(x) \right\} \frac{dx(\tau)}{d\tau} = \sum_{i=1}^m \left\{ \frac{\partial^2 P(x, \lambda, \tau)}{\partial x \partial \tau} \cdot \nabla g_i(x) \right\}$$

Como $F(x, \lambda, \tau)$ é estritamente convexa pela condição **C2**, então podemos encontrar uma expressão para $\frac{dx(\tau)}{d\tau}$,

$$\frac{dx(\tau)}{d\tau} = Q^{-1}T, \quad (3.2)$$

onde

$$Q = \nabla^2 f(x) + \sum_{i=1}^m P''(g_i(x), \lambda, \tau) \cdot \nabla g_i(x) \cdot \nabla^t g_i(x) + \sum_{i=1}^m P'(g_i(x), \lambda, \tau) \cdot \nabla^2 g_i(x)$$

e

$$T = \sum_{i=1}^m \frac{\partial^2 P(g_i(x), \lambda, \tau)}{\partial x \partial \tau} \cdot \nabla g_i(x).$$

Dessa forma, mostramos que a derivada de primeira ordem de x nos fornece um valor finito. Com essa informação podemos representar a trajetória com um polinômio linear para $\tau > 0$. Porém, a expansão em série de Taylor de mais alta ordem de $x(\tau)$ exige igualmente derivadas de ordem mais elevada. Mostraremos então nas seções seguintes, a existência das derivadas de ordem mais elevada em $\tau > 0$. Como afinal estamos interessados no limite quando $\tau \rightarrow 0$, $x(0) = x^*$, devemos igualmente mostrar a existência de todas essas derivadas no ponto $\tau = 0$.

3.1.3 Existência de $\frac{d^n x(\tau)}{d\tau^n}$ e $\frac{d^n \mu(\tau)}{d\tau^n}$

Nessa seção discutiremos a existência das derivadas de $x(\tau)$ e $\mu(\tau)$ em relação ao parâmetro τ . Esse resultado é de fundamental importância para nossa pretensão, que é representar $x(\tau)$ pela série de Taylor usando termos de ordem mais elevada, aplicando raciocínio análogo para o multiplicador de Lagrange $\mu(\tau)$.

Teorema 3.2 *Sob as condições pré-estabelecidas, $x(\tau)$ e $\mu(\tau)$ têm todas as derivadas com respeito a τ para qualquer $\tau > 0$.*

Demonstração:

Seguindo a mesma seqüência apresentada em Lootsma [11], consideremos que o ponto de mínimo $x(\tau)$ anula o gradiente da função objetivo modificada $\nabla F(x, \lambda, \tau)$.

Portanto, $x(\tau)$ anula o sistema abaixo:

$$\nabla f(x(\tau)) + \sum_{i=1}^m P'(g_i(x(\tau)), \lambda, \tau) \cdot \nabla g_i(x(\tau)) = 0. \quad (3.3)$$

Como por hipótese, $F(x, \lambda, \tau)$ é estritamente convexa, isto implica que a matriz Hessiana de $F(x, \lambda, \tau)$ seja definida positiva e pelo teorema da função implícita podemos afirmar que o vetor $x(\tau)$ é continuamente diferenciável em $\lambda > 0$, como feito para $dx/d\tau$ na seção 3.1.2.

Analisaremos agora a diferenciabilidade dos Multiplicadores de Lagrange que são dados por:

$$\mu_i(\tau_k) = \lambda_k \left[1 - \frac{\lambda_k g_i(x(\tau_k))}{\sqrt{(\lambda_k g_i(x(\tau_k)))^2 + \tau_k^2}} \right]. \quad (3.4)$$

Diferenciando a expressão acima em relação a τ temos:

$$\frac{d\mu_i(\tau)}{d\tau} = \frac{\partial \mu_i(\tau)}{\partial \tau} + \frac{\partial \mu_i(\tau)}{\partial g_i(x(\tau))} \sum_{j=1}^n \frac{\partial g_i(x_j(\tau))}{\partial x_j(\tau)} \frac{\partial x_j(\tau)}{\partial \tau}. \quad (3.5)$$

Podemos verificar através da diferenciação da função penalidade (2.3), que os termos $\frac{\partial \mu_i(\tau)}{\partial \tau}$ e $\frac{\partial \mu_i(\tau)}{\partial g_i(x(\tau))}$ são continuamente diferenciáveis. Pela condição **C1**, os termos $\frac{\partial g_i(x_j(\tau))}{\partial x_j(\tau)}$ também são continuamente diferenciáveis. Como provamos acima, o

termo $\frac{\partial x_j(\tau)}{\partial \tau}$ é continuamente diferenciável . Portanto, podemos concluir que o Multiplicador de Lagrange $\mu(\tau)$ também será continuamente diferenciável em relação a τ . ■

A seguir, mostraremos que no ponto $\tau = 0$, o problema apresenta as mesmas características de diferenciabilidade apresentadas acima para $\tau > 0$.

3.1.4 Análise da solução em $\tau = 0$

Nos métodos de penalização interior, a teoria nos mostra que o ótimo é alcançado em $\tau = 0$, onde obtemos x^* e μ^* . Precisamos, então, garantir a finitude e a existência das derivadas de $x(\tau)$ e $\mu(\tau)$ no ponto $\tau = 0$, pois é exatamente nesse ponto que deveremos fazer o desenvolvimento de Taylor.

Teorema 3.3 *Sob as condições pré-estabelecidas $x(\tau)$ e $\mu(\tau)$ possuem derivada primeira em relação a τ no ponto $\tau = 0$.*

Demonstração:

Considere a função Lagrangeana :

$$L(x, \mu) = f(x) - \sum_{i=1}^m \mu_i g_i(x). \quad (3.6)$$

Derivando a expressão em relação a τ temos :

$$\frac{d}{d\tau} \begin{vmatrix} \nabla_x L \\ \nabla_\mu L \end{vmatrix} = \frac{d}{d\tau} \begin{vmatrix} \nabla_x L \\ -g \end{vmatrix} = \begin{vmatrix} H(x, \mu) & -\nabla_x g \\ -\nabla_x^t g & 0 \end{vmatrix} \begin{vmatrix} \frac{dx}{d\tau} \\ \frac{d\mu}{d\tau} \end{vmatrix}, \quad (3.7)$$

onde $H(x, \mu)$ é a Hessiana de $L(x, \mu)$ e é dada por :

$$H(x, \mu) = \nabla^2 f(x) - \sum_{i=1}^m \mu_i \nabla^2 g_i(x).$$

Todas as funções em (3.7) são avaliadas no ponto $(x(\tau), \mu(\tau))$, cuja existência já foi provada para $\tau > 0$.

Antes de tomarmos o limite da expressão (3.7) para $\tau \rightarrow 0$, observemos as considerações :

A) Para toda restrição ativa, ou seja, para todo i tal que $g_i(x^*) = 0$, mostramos na seção 2.4 que :

$$\lim_{\tau \rightarrow 0} \frac{\tau}{g_i(x(\tau))} = s_i^* > 0. \quad (3.8)$$

Logo,

$$\lim_{\tau \rightarrow 0} \frac{g_i(x(\tau))}{\tau} = \omega_i^* < \infty. \quad (3.9)$$

Podemos ainda, a partir da expressão (3.9) obter

$$\lim_{\tau \rightarrow 0} dg_i(x(\tau))/d\tau = \omega_i^*. \quad (3.10)$$

Por outro lado, como $x(\tau)$ é ponto de mínimo para qualquer valor de τ , necessariamente devemos ter a identidade:

$$\nabla_x L(x(\tau), \mu(\tau)) = \nabla_x f(x(\tau)) - \sum_{i=1}^m \mu_i(\tau) \nabla_x g_i(x(\tau)) \equiv 0, \quad (3.11)$$

já que $\mu_i(\tau) = -P'(g_i(x), \lambda, \tau)$.

Assim,

$$\frac{d}{d\tau} [\nabla_x L(x(\tau), \mu(\tau))] = 0. \quad (3.12)$$

Considerando as restrições não ativas ($g_i(x(\tau)) > 0$), no ponto de mínimo, podemos calcular $\frac{d\mu_i(\tau)}{d\tau}$ a partir da expressão (2.8):

$$\frac{d\mu_i(\tau)}{d\tau} = \frac{-\phi(\tau)\lambda^2 dg_i(x(\tau))/d\tau + \lambda^2 g_i(x(\tau))\phi'(\tau)}{\phi^2(\tau)}, \quad (3.13)$$

onde $\phi(\tau) = \sqrt{\lambda^2 g_i^2(x(\tau)) + \tau^2}$.

Portanto, como $g_i(x(\tau)) > 0$, temos:

$$\lim_{\tau \rightarrow 0} \frac{d\mu_i(\tau)}{d\tau} = 0. \quad (3.14)$$

Finalmente, a partir da expressão acima, para todo i tal que $g_i(x^*) = 0$, podemos reduzir o sistema de equações (3.7). Seja m^* o número de restrições ativas, vamos supor, sem perda de generalidade, que estas sejam organizadas seqüencialmente g_1, \dots, g_{m^*} . Definimos ainda $dv/d\tau = |d\mu_1/d\tau, \dots, d\mu_{m^*}/d\tau|$.

Considerando os resultados anteriores (3.10) e (3.12) e tomando o limite quando $\tau \rightarrow 0$, a equação (3.7) se transforma em :

$$\begin{vmatrix} 0 \\ \omega_1^* \\ \vdots \\ \omega_{m^*}^* \end{vmatrix} = \begin{vmatrix} H(x^*, \mu^*) & -\nabla g_1(x^*) & \dots & -\nabla g_{m^*}(x^*) \\ -\nabla_x g_1(x^*) & & & \\ \vdots & & 0 & \\ -\nabla_x g_{m^*}(x^*) & & & \end{vmatrix} \begin{vmatrix} \frac{dx(0)}{d\tau} \\ \vdots \\ \frac{dv(0)}{d\tau} \end{vmatrix} \quad (3.15)$$

Vejam as condições para que $\frac{dx(0)}{d\tau}$ seja finito:

I - A matriz representada na equação (3.15) deve ser inversível;

II - Os valores ω_i^* , $i = 1, 2, \dots, m^*$ devem ser finitos, ou seja, os multiplicadores

μ_i^* , $i = 1, 2, \dots, m$ devem ser diferentes de zero.

Pela condição **C3**, os gradientes das restrições ativas são linearmente independentes. A matriz $H(x^*, \mu^*)$ é definida positiva, pois os multiplicadores de Lagrange são positivos e observando a condição **C2**, podemos dizer que, ou $f(x)$ é estritamente convexa, ou alguma restrição $g_i(x)$ é estritamente côncava. Isso mostra que a matriz em (3.15) é inversível. Assim, a condição (I) é satisfeita.

A partir da condição **C4**, podemos afirmar que os multiplicadores μ_i^* são positivos, o que implica, conforme mostramos em (2.19), que ω_i^* seja finito. Assim, a condição (II) acima está satisfeita. ■

Desta forma, conseguimos mostrar a existência e finitude da derivada de primeira ordem $\frac{dx(\tau)}{d\tau}$ no ponto $\tau = 0$. Em seguida, demonstraremos que as derivadas de ordem superior de $g_i(x(\tau))$ em relação a τ também existem no ponto $\tau = 0$. Esse é um resultado preliminar para demonstrar o teorema seguinte, que garante as derivadas de $(x(\tau), \mu(\tau))$ no ponto $\tau = 0$.

Lema 3.4 *A existência da derivada de ordem $(n - 1)$ de $\mu(\tau)$ com respeito a τ no ponto $\tau = 0$, para todo i tal que $\mu_i(0) > 0$, implica na existência da derivada de ordem n de $g_i(x(\tau))$ com respeito a τ no ponto $\tau = 0$.*

Demonstração:

A demonstração do Lema foi baseada na demonstração de Xavier [22]. Consideremos novamente a expressão obtida para os multiplicadores de Lagrange:

$$\mu_i(\tau_k) = \lambda_k \left[1 - \frac{\lambda_k g_i(x(\tau_k))}{\sqrt{(\lambda_k g_i(x(\tau_k)))^2 + \tau_k^2}} \right]. \quad (3.16)$$

Por simplicidade de notação, consideraremos $g(x(\tau)) = g(\tau)$. Definamos

$z_i(\tau) = \frac{g_i(\tau)}{\tau}$. Como já demonstramos, $\mu_i^* > 0$, implica que $\lim_{\tau \rightarrow 0} z_i(\tau) < \infty$.

A existência da derivada de ordem n de $\mu_i(\tau)$ em relação a τ no ponto $\tau = 0$ implica na existência da derivada de ordem n de $z_i(\tau)$ em relação a τ .

O desenvolvimento que faremos a seguir é rigorosamente igual ao realizado por Fiacco [3].

Considere a regra de Leibnitz para calcular a derivada de ordem $(n - 1)$ de $z_i(\tau)$ com respeito a τ .

$$\frac{d^{n-1}z(\tau)}{d\tau^{n-1}} = \sum_{k=0}^{n-1} \binom{n-1}{k} \frac{d^{n-k-1}g(\tau)}{d\tau^{n-k-1}} \cdot \frac{d^k \tau^{-1}}{d\tau^k}. \quad (3.17)$$

Pelo Teorema 3.2 todas as derivadas de $g(\tau)$ existem para $\tau > 0$. Mas desde que $d^k(\tau^{-1})/d\tau^k = (-1)^k k! (\tau^{-1})^{k+1}$ a equação (3.17) é modificada para:

$$\frac{d^{n-1}z(\tau)}{d\tau^{n-1}} = \frac{(n-1)!}{\tau^n} \sum_{k=0}^{n-1} \frac{(-1)^k}{(n-k-1)!} \frac{d^{n-k-1}g(\tau)}{d\tau^{n-k-1}} \tau^{n-k-1}. \quad (3.18)$$

Consideremos agora o desenvolvimento por Taylor de $g(0)$ a partir de τ :

$$g(0) = g(\tau) + \frac{dg(\tau)}{d\tau}(-\tau) + \frac{d^2g(\tau)}{d\tau^2} \frac{(-\tau)^2}{2} + \dots + \frac{d^{n-1}g(\tau)}{d\tau^{n-1}} \frac{(-\tau)^{n-1}}{(n-1)!} + O(\tau^n) \quad (3.19)$$

Como a restrição é ativa, temos $g(0) = 0$. De outro lado, observando-se o somatório do lado direito da equação (3.18) e comparando-o com o desenvolvimento de Taylor realizado na equação (3.19), podemos simplificar a equação (3.18):

$$\frac{(n-1)!}{\tau^n} \cdot \frac{\tau^n}{n!} \cdot \frac{d^n g(\nu)}{d\tau^n}, \quad \text{onde } 0 < \nu < \tau. \quad (3.20)$$

Tomando o limite quando $\tau \rightarrow 0$, obtemos:

$$\frac{d^{n-1}z(0)}{d\tau^{n-1}} = \frac{1}{n} \cdot \frac{d^n g(0)}{d\tau^n} \quad \blacksquare \quad (3.21)$$

Como os multiplicadores μ_i só dependem das variáveis $z_i(\tau)$, o teorema está provado.

Para completar o embasamento teórico que dá suporte à extrapolação, vejamos o teorema a seguir :

Teorema 3.5 *Sob as condições pré-estabelecidas, as derivadas de $(x(\tau), \mu(\tau))$ com respeito a τ existem no ponto $\tau = 0$ e são finitas.*

Demonstração:

Consideremos o ponto de mínimo $x(\tau)$ que anula o sistema de equações abaixo:

$$\nabla f(x(\tau)) + \sum_{i=1}^m P'(g_i(x(\tau)), \lambda, \tau) \cdot \nabla g_i(x(\tau)) = 0. \quad (3.22)$$

Podemos reescrever o sistema acima da seguinte maneira:

$$\nabla_x f(\tau) - \sum_{i=1}^m \mu_i(\tau) \cdot \nabla_x g_i(\tau) = 0. \quad (3.23)$$

Diferenciando a equação (3.23) em relação a τ temos:

$$H(\tau)x' - \sum_{i=1}^m \mu_i'(\tau) \nabla_x g_i(\tau) = 0, \quad (3.24)$$

onde $x' = dx(\tau)/d\tau$, $\mu_i'(\tau) = d\mu_i(\tau)/d\tau$ e $H(\tau)$ é a Hessiana do Lagrangeano, ou seja, $H(\tau) = \nabla^2 f(\tau) - \sum_{i=1}^m \mu_i \nabla^2 g_i(\tau)$.

Considere ainda a relação de identidade:

$$\nabla_x^t g(\tau)x' = g'(\tau). \quad (3.25)$$

Diferenciando as equações (3.24) e (3.25) com relação a τ obtemos:

$$H(\tau)x'' - \sum_{i=1}^m \mu_i''(\tau)\nabla_x g_i(\tau) = -H'(\tau)x' + \sum_{i=1}^m \mu_i'(\tau)\nabla_x g_i'(\tau) \quad (3.26)$$

$$-\nabla_x^t g(\tau)x'' = -g''(\tau) + (\nabla_x^t g'(\tau))x' \quad (3.27)$$

Novamente, é possível reduzir o sistema de equações acima e considerar somente as restrições ativas, pois $\lim_{\tau \rightarrow 0} d^n \mu_i(0)/d\tau^n = 0$, para todo i em que $g_i(0) > 0$.

Observando os termos à direita das expressões (3.26) e (3.27), podemos dizer que cada um dos termos existe pelas seguintes razões : pela condição **C1**, as funções $f(x)$ e $g(x)$ possuem diferenciabilidade de todas as ordens; o Lema 3.4 garante que a existência das derivadas de ordem $n - 1$ para $\mu_i(\tau)$ no ponto $\tau = 0$ implica na existência da derivada de ordem n de $g_i(\tau)$; o Teorema 3.3 garante a existência de $dx(\tau)/d\tau$ e $d\mu(\tau)/d\tau$ no ponto $\tau = 0$.

Para calcular x'' e μ'' , temos que resolver o mesmo sistema (3.15), que como visto tem solução. Dessa forma, obtemos as derivadas segundas de x e μ em relação a τ quando $\tau = 0$.

No que concerne às derivadas de mais alta ordem, procedendo exatamente da maneira usada para as derivadas segundas, o sistema (3.26) e (3.27) sempre aparecerá. Destarte, obteremos todas as derivadas de x e μ em relação a τ quando

$\tau = 0$. ■

Todo o embasamento teórico desenvolvido nas seções anteriores garantem o uso do desenvolvimento de Taylor de $x(\tau)$ e $\mu(\tau)$ no ponto $\tau = 0$. Esse desenvolvimento, viabiliza a aplicação do esquema de extrapolação no problema irrestrito gerado pelo método de Penalização Hiperbólica.

Na seção seguinte descreveremos a idéia do mecanismo de extrapolação polinomial bem como sua fórmula iterativa para gerar estimativas.

3.2 Extrapolação Polinomial

Todo o embasamento teórico realizado nas seções anteriores permite-nos afirmar que, dada uma seqüência de minimizadores da forma $x(\tau)$, é possível encontrar um polinômio que modele essa trajetória rumo ao ponto ótimo.

Essa trajetória é modelada por um polinômio, que é obtido através da expansão em série de Taylor de $x(\tau)$. Por conveniência, realizaremos esta expansão em torno do ponto limite do processo iterativo de minimização irrestrita, que é em $\tau = 0$. Dessa forma, a expansão é realizada em torno de $x(0)$, que na realidade é o ótimo x^* .

A nossa motivação é de que, com a trajetória, possamos realizar estimativas da solução ótima avaliando o polinômio no ponto $\tau = 0$. Grande parte da nossa motivação deve-se também à obtenção desse polinômio sem grandes esforços computacionais e à utilização do mesmo sem precisar calcular os seus coeficientes de

maneira explícita, como veremos a seguir.

Suponha que $F(x, \lambda, \tau)$ tenha sido minimizado para os parâmetros $\tau_1 > \tau_2 > \dots > \tau_k$, obtendo os minimizadores $x(\tau_1), x(\tau_2), \dots, x(\tau_k)$ e os multiplicadores $\mu(\tau_1), \mu(\tau_2), \dots, \mu(\tau_k)$. Já vimos que $(x(\tau), \mu(\tau))$ é n vezes continuamente diferenciável para $\tau \geq 0$. Portanto, o polinômio que passa por $x(\tau_1), \dots, x(\tau_k)$ é dado pelo conjunto de equações

$$x(\tau_i) = \sum_{j=0}^{k-1} a_j (\tau_i)^j, \quad i = 1, \dots, k, \quad (3.28)$$

onde os a_j são vetores com n componentes, ou seja, a mesma dimensão de x . Podemos observar em (3.28) que $x(0) = x^*$ é aproximado por a_0 .

A teoria desenvolvida mostra que um outro polinômio pode ser construído para realizar estimativas do Multiplicador de Lagrange ótimo dada uma seqüência $\mu_1, \mu_2, \dots, \mu_k$. Esse polinômio, com raciocínio análogo ao caso das variáveis primais, pode ser escrito como:

$$\mu(\tau_i) = \sum_{j=0}^{k-1} b_j (\tau_i)^j, \quad i = 1, \dots, k, \quad (3.29)$$

onde os b_j são vetores com m componentes, ou seja, a mesma dimensão de μ .

A expansão por Taylor de $x(\tau)$ e $\mu(\tau)$, em torno de $\tau = 0$, é dada, respectivamente, por:

$$x(\tau) = a_0 + a_1\tau + a_2\tau^2 + \dots + O(\tau^k), \quad (3.30)$$

$$\mu(\tau) = b_0 + b_1\tau + b_2\tau^2 + \dots + O(\tau^k), \quad (3.31)$$

onde os coeficientes $\{a_j\}$ e $\{b_j\}$ são desconhecidos.

O método de extrapolação de Richardson-Romberg adota a regra de variação do parâmetro τ através da relação $\tau_{k+1} = \rho\tau_k$, tal que $0 < \rho < 1$. Esta regra torna possível o desenvolvimento de um simples esquema iterativo para computar estimativas sem, no entanto, calcular explicitamente os coeficientes a_0, a_1, \dots, a_k e b_0, b_1, \dots, b_k .

Consideremos $x_{i,j}$, para $i = 1, \dots, k$ e $j = 1, \dots, i - 1$. O índice j representa a j -ésima estimativa de $x(0)$ após i mínimos alcançados.

Veja que $x_{3,1}$ representa uma extrapolação de primeira ordem, vide matriz da figura 1.3, após alcançados 3 mínimos. Assim, todos os elementos que estão nesta mesma coluna representam extrapolações lineares. À medida que incorporam-se novos mínimos obtidos pelo processo de minimização, realizam-se novas extrapolações e conseqüentemente extrapolações mais precisas, já que teoricamente sabe-se que quanto mais para baixo e para a direita o fluxo computacional, melhores são as estimativas realizadas.

O esquema iterativo é dado segundo a mesma formulação proposta em (1.34) e (1.35). Repetiremos aqui o esquema iterativo, porém com as adaptações pertinentes às estimativas realizadas do ponto ótimo.

Desta forma temos :

$$x_{i,0} = x(\tau_0 \cdot \rho^{i-1}), \quad i = 1, \dots, k, \quad (3.32)$$

onde τ_0 é o valor inicial de τ e

$$\begin{aligned} x_{i,j} &= \frac{x_{i,j-1} - \rho^j x_{i-1,j-1}}{1 - \rho^j}, \\ i &= 2, \dots, k - 1, \end{aligned} \quad (3.33)$$

$$j = 1, 2, \dots, i - 1.$$

Então, em princípio, a melhor estimativa de $x(0)$ é dada por :

$$x(0) = x_{k,k-1} = a_0. \quad (3.34)$$

A mesma matriz de Richardson-Romberg apresentada na seção de Integração Numérica pode ser construída aqui. Na realidade, para um vetor $X \in R^n$ são realizadas n extrapolações, uma para cada componente individual de X . Desta forma, representamos, conforme descreve a teoria, x_1 como uma função de τ , x_2 como uma função de τ , e assim por diante até x_n .

De acordo com o esquema desenvolvido por Richardson-Romberg e descrito acima, na resolução dos problemas de otimização, precisamos ao menos ter dois pontos de mínimo obtidos do processo de minimização irrestrita para, a partir daí, aplicarmos o mecanismo de extrapolação. Com dois pontos, podemos apenas realizar uma extrapolação linear. Com a obtenção de um terceiro ponto obtido na minimização irrestrita, podemos realizar uma estimativa linear e ainda uma estimativa quadrática e assim por diante. Com poucos pontos gerados, a partir do processo de minimização, esperamos obter boas estimativas para o ótimo, como no exemplo apresentado no problema de integração numérica.

De outro lado, podemos observar que a solução de (2.4) passa pela resolução de uma seqüência de subproblemas, onde é necessário um ponto inicial para realizar o processo de minimização para τ_{k+1} . Geralmente estes pontos iniciais são tomados como a solução do subproblema anterior. Alternativamente, podemos utilizar o mecanismo de extrapolação para gerar bons pontos iniciais, ou seja, gerar

pontos iniciais que estejam bem próximos à trajetória central. A esse mecanismo denominamos *extrapolação inversa*. Com a utilização desse mecanismo, desenvolvido abaixo, esperamos acelerar o processo de resolução do subproblema irrestrito através da redução do número de iterações na busca da solução ótima do subproblema.

3.2.1 Extrapolação Inversa

A fórmula de extrapolação também pode ser usada para estimar o próximo mínimo da função irrestrita $F(x, \lambda, \tau_{k+1})$ após já terem sido computados k mínimos. Por exemplo, o $(k + 1)$ -ésimo mínimo, baseado na informação fornecida pelos k mínimos é estimado por:

$$x_{k+1,0} = x(\tau_0 \cdot \rho^k) = a_0 + \dots + a_{j-1}(\tau_0 \cdot \rho^k)^{j-1} + \dots + a_{k-1}(\tau_0 \cdot \rho^k)^{k-1}. \quad (3.35)$$

Apesar dos coeficientes não serem explicitamente calculados, é possível utilizar as relações (3.33) e (3.34) de modo a calcular inversos em direção à estimativa do $(k + 1)$ -ésimo mínimo $x_{k+1,0}$. Isso é feito pela substituição $i = k + 1$ na equação (3.33) e isolando o termo $x_{k+1,j-1}$, o que gera a relação recursiva

$$\hat{x}^{k+1} = x_{k+1,j-1} = (1 - \rho^j)x_{k+1,j} + \rho^j x_{k,j-1}. \quad (3.36)$$

Vamos assumir a hipótese que $a_0 = x_{k,k-1} = x_{k+1,k-1}$, pois $x_{k,k-1}$ é a melhor estimativa para o valor de a_0 na iteração k e é natural supor, que esse valor será mantido na iteração seguinte. Assim, com os valores obtidos pela expressão (3.33), podemos utilizá-los para avaliar a expressão (3.36) em $j = k - 1, k - 2, \dots, 1$. O último valor computado será tomado como a estimativa do ponto de mínimo do próximo subproblema penalizado e dado por $\hat{x}_{k+1,0}$.

Essa estimativa será utilizada como ponto inicial para a minimização do $(k + 1)$ -ésimo subproblema irrestrito. Como vários mínimos são obtidos ao longo do processo iterativo, espera-se que as estimativas sejam cada vez melhores. Nossa expectativa é de que essas estimativas reduzam o esforço computacional para minimizar a função irrestrita $F(x, \lambda, \tau_{k+1})$, de forma que serão dados menos passos até encontrar o ótimo do subproblema, pois o ponto inicial adotado está mais próximo do ótimo do que o ponto convencionalmente utilizado.

Segundo Fiacco [3], deve-se limitar o número de estimativas possíveis, já que para estimativas de alta ordem, podemos nos deparar com situações indesejáveis tais como erro de arredondamento. Por esse motivo, neste trabalho fixamos o número de estimativas em, no máximo 7, o que permite obter polinômios de até grau 6.

Um exemplo do funcionamento do mecanismo de extrapolação inversa está representada pelo gráfico da figura (3.1). O elemento na coluna mais à direita e na última linha, representa a melhor estimativa da solução ótima. Por este motivo, é copiado integralmente para a linha abaixo, onde será iniciado o processo inverso como mostrado por (3.1).

O termo $\hat{x}_{4,0}$ representa a estimativa de ordem 3 do ponto ótimo do $(k + 1)$ -ésimo subproblema. Podemos ainda realizar estimativas de ordem 2 com processo análogo ao exposto pela figura (3.1), copiando-se o termo $x_{3,2}$ para a linha seguinte e combinando-o aos termos subseqüentes aplicando a relação (3.36). Este mecanismo viabiliza a obtenção de um conjunto de $k - 1$ estimativas do ponto inicial da próxima iteração.

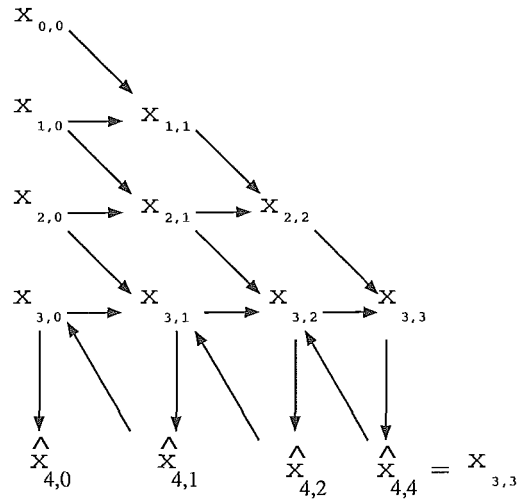


Figura 3.1: Estimativa do Ótimo do próximo subproblema penalizado

De posse destas $k - 1$ estimativas, escolhemos o ponto que leva ao menor valor da função objetivo modificada $F(x, \lambda, \tau)$, sem levarmos em consideração a questão da inviabilidade, que será tratada pelo método da Penalização Hiperbólica.

Capítulo 4

Resultados Computacionais

Neste capítulo temos o objetivo de apresentar os experimentos computacionais obtidos na resolução de um conjunto de problemas-teste apresentados por Schittkowsi [7], bem como, realizar comparações dos resultados obtidos nos métodos de extrapolação e extrapolação inversa com os resultados apresentados na literatura.

Estudaremos também, nesse capítulo, as variações da trajetória central descrita pelos minimizadores $x(\tau)$, associadas às mudanças dos valores do parâmetro ρ . Nesse trabalho utilizamos os valores $\rho = 1/2, 1/4$ ou $1/10$, que são os valores apresentados na literatura. Fizemos ainda um estudo para verificar qual desses valores do parâmetro ρ se adequa melhor ao conjunto de problemas resolvidos.

Como o mecanismo de extrapolação gera um conjunto de polinômios, realizamos escolhas dos polinômios que modelam a trajetória central de cada componente do vetor $x \in R^n$. Observamos que a escolha do grau do polinômio é um mecanismo importante e que está relacionado à estimativa da solução ótima fornecida pela extrapolação. Sendo assim, analisamos quais os graus de polinômios mais adequados.

Segundo Ariela [13], os polinômios quadráticos são os mais freqüentes. Analisamos também uma possível relação que possa haver entre o grau do polinômio escolhido com o grau da função objetivo e das restrições ativas.

Os problemas-teste foram resolvidos de duas maneiras distintas. Numa primeira análise, os problemas foram resolvidos sem a inclusão das rotinas de extrapolação. Em seguida, os problemas foram solucionados com a inclusão das rotinas de extrapolação. Desse modo, objetivamos medir os ganhos obtidos quando o mecanismo de extrapolação é aplicado.

Como descrito pela teoria, os problemas penalizados tornam-se crescentemente mal condicionados e por esse motivo nos utilizamos da expressão apresentada por Schittkowski [7] para realizar uma medida do condicionamento intrínscico (MC) dos problemas . A expressão é dada por:

$$MC = \frac{\mu_{max}}{\mu_{min}}, \quad (4.1)$$

onde μ_{max} e μ_{min} , são respectivamente, o maior e o menor multiplicador.

Nas seções seguintes apresentaremos os resultados computacionais obtidos com a implementação do método de extrapolação em Penalização Hiperbólica através de tabelas comparativas. O método foi implementado utilizando-se a rotina de BFGS e as rotinas de penalização hiperbólica desenvolvida por Xavier [21] em Visual Fortran Compaq 6.0.

As tabelas comparativas foram construídas a partir da seguinte notação:

NPCE = número de passos obtidos com a utilização da rotina de extrapolação;

NPSE = número de passos obtidos sem a utilização da rotina de extrapolação;

$f(x(\tau))$ = valor da função objetivo no ponto $x(\tau)$;

GRAU = grau do polinômio escolhido.

4.1 Problema Teste 1

Embora a teoria tenha sido desenvolvida para problemas de programação convexa, o método de extrapolação foi usado com sucesso nesse primeiro problema teste, HS 116 [7] que não é convexo.

Nesse experimento, realizamos uma comparação entre o número de passos dados a cada resolução de um subproblema com a inclusão do método de extrapolação e sem a inclusão do mesmo. O nosso objetivo é avaliar o ganho proporcionado pela extrapolação inversa na estimativa de ótimos da próxima iteração.

Estudamos também, a resolução do problema com diferentes valores para o parâmetro ρ com o objetivo de observar os valores estimados pela extrapolação, os graus dos polinômios obtidos e a convergência em direção ao ótimo em cada caso.

Realizamos ainda, uma comparação com os resultados obtidos para os diferentes valores de ρ num ponto de *corte*. Entende-se por *corte*, um valor pré-fixado de τ . Com esse procedimento, para $\rho = 1/10$, inicializamos τ com valor 100. Para $\rho = 1/4$, inicializamos $\tau = 1.024$. Dessa forma, no ponto de *corte* $\tau = 10^{-3}$ o número de extrapolações realizadas é o mesmo em ambos os casos, o que nos permite analisar, em igualdade de condições, o desempenho dessas reduções

de τ . O resultado obtido está descrito na tabela (4.4).

Os parâmetros iniciais τ_0 e λ_0 para os dois casos $\rho = 1/4$ e $\rho = 1/10$ referentes ao problema HS 116 estão na tabela (4.1).

ρ	τ_0	λ_0
$\frac{1}{4}$	1.024	150
$\frac{1}{10}$	1.000	150

Tabela 4.1: Dados Iniciais de HS 116

O ponto inicial foi tomado exatamente igual ao apresentado por Schittkowski [7]:

$$x_0 = (0.5, 0.8, 0.9, 0.1, 0.14, 0.5, 489, 80, 650, 450, 150, 150, 150). \quad (4.2)$$

Como se trata de um problema com 13 variáveis, iremos suprimir os valores obtidos para o vetor x durante a minimização. Assim, por concisão, apresentaremos somente a solução ótima.

Vejamos a definição do problema e a aplicação do mecanismo de extrapolação no problema não convexo HS 116. Os resultados obtidos com a Penalização Hiperbólica com extrapolação estão descritos na tabela 4.2, onde a primeira coluna representa os valores do parâmetro de penalização interior e cada linha com **extrap** representa a extrapolação realizada.

$$\textit{Minimizar} \quad f(x) = x_{11} + x_{12} + x_{13}$$

sujeito a

$$x_3 - x_2 \geq 0 \quad (4.3)$$

$$x_2 - x_1 \geq 0 \quad (4.4)$$

$$1 - 0.002x_7 + 0.002x_8 \geq 0 \quad (4.5)$$

$$50 \leq f(x) \leq 250 \quad (4.6)$$

$$x_{13} - 1.262626x_{10} + 1.231059x_3x_{10} \geq 10 \quad (4.7)$$

$$x_5 - 0.3475x_2 - 0.975x_2x_5 + 0.00975x_2^2 \geq 0 \quad (4.8)$$

$$x_6 - 0.03475x_3 - 0.975x_3x_6 + 0.00975x_3^2 \geq 0 \quad (4.9)$$

$$x_5x_7 - x_1x_8 - x_4x_7 + x_4x_8 \geq 0 \quad (4.10)$$

$$1 - 0.002(x_2x_9 + x_5x_8 - x_1x_8 - x_6x_9) - x_5 - x_6 \geq 0 \quad (4.11)$$

$$x_2x_9 - x_3x_{10} - x_6x_9 - 500x_2 + 500x_6 + x_2x_{10} \geq 0 \quad (4.12)$$

$$x_2 - 0.9 - 0.002(x_2x_{10} - x_3x_{10}) \geq 0 \quad (4.13)$$

$$x_4 - 0.03475x_1 - 0.975x_1x_4 + 0.00975x_1^2 \geq 0 \quad (4.14)$$

$$x_{11} - 1.62626x_8 + 1.231059x_1x_8 \geq 0 \quad (4.15)$$

$$x_{12} - 1.262626x_9 + 1.231059x_2x_9 \geq 0 \quad (4.16)$$

A solução ótima obtida após a minimização foi o ponto viável

$$x^* = (0.6921320530, 0.9403979200, 0.9403979203, 0.05960207997, \\ 0.1000000015, 0.2902993364, 534.1122570634, 34.1122570661, \\ 500.0, 0.1000000001, 14.0054900890, 52.4703384864, \\ 0.0104940677).$$

Discussão dos resultados

As cinco primeiras colunas das tabelas (4.2) e (4.3) já foram explicadas anteriormente. A sexta coluna, mostra o erro absoluto entre o valor da função

τ	$f(x(\tau))$	NPCE	NPSE	Grau	$ f(x^k) - f(x^*) $
1.0	67.5147048222	271	271	-	1.0
1.0e-1	66.5883864124	66	66	-	$1.0e - 1$
Extrap1	66.4854621450	-	-	1	$8.0e - 4$
1.0e-2	66.4965219197	88	77	-	$1.0e - 2$
Extrap2	66.4863147540	-	-	1	$8.0e - 6$
1.0e-3	66.4873424284	61	63	-	$1.0e - 3$
Extrap3	66.4863224850	-	-	1	$2.0e - 7$
1.0e-4	66.4864246167	70	64	-	$1.0e - 4$
Extrap4	66.4863226380	-	-	1	$4.0e - 9$
1.0e-5	66.4863328394	38	66	-	$1.0e - 5$
Extrap5	66.4863226420	-	-	1	0
1.0e-6	66.4863236618	59	101	-	$1.0e - 6$
Extrap6	66.4863226420	-	-	1	0
1.0e-7	66.4863227441	23	53	-	$1.0e - 7$
Extrap7	66.4863226420	-	-	1	0
1.0e-8	66.4863226523	59	88	-	$1.0e - 8$
Extrap8	66.4863226420	-	-	3	0
1.0e-9	66.4863226431	17	49	-	$1.0e - 9$
Extrap9	66.4863226420	-	-	5	0
1.0e-10	66.4863226422	65	66	-	$2.0e - 10$
Extrap10	66.4863226420	-	-	4	0
1.0e-11	66.4863226421	91	89	-	$1.0e - 10$
Extrap11	66.4863226420	-	-	4	0

Tabela 4.2: Resultados do problema HS 116 com $\rho = \frac{1}{10}$

objetivo associado à especificada linha e o valor ótimo dessa função objetivo.

Devemos observar primeiramente que as linhas associadas ao uso da extrapolação mostram cabal e claramente a eficiência desse mecanismo. Para ilustrarmos o ganho proporcionado pela extrapolação apresentamos nas tabelas (4.5) e (4.6) o resultado obtido para $\Delta f(x^k) = |f(x^k) - f(x^*)|$ antes da utilização da extrapolação (coluna 2) e o resultado obtido após o uso do mecanismo (coluna 3) na resolução do problema HS 116 com parâmetro $\rho = \frac{1}{10}$ e $\rho = \frac{1}{4}$, respectivamente. A última coluna representa o ganho proporcionado pelo mecanismo de extrapolação.

τ	$f(x(\tau))$	NPCE	NPSE	Grau	$ f(x^k) - f(x^*) $
1.024	67.5395596537	174	174	-	1.05
2.56e-1	66.7479967925	35	35	-	$2.62e - 1$
extrap1	66.4841425050	-	-	1	$2.18e - 3$
6.40e-2	66.5516250411	39	56	-	$6.53e - 2$
extrap2	66.4861677910	-	-	1	$1.55e - 4$
1.60e-2	66.5026424342	43	31	-	$1.63e - 2$
extrap3	66.4863148990	-	-	1	$7.74e - 6$
4.00e-3	66.4904020425	47	52	-	$4.08e - 3$
extrap4	66.4863219120	-	-	1	$7.30e - 7$
1.00e-3	66.4873424270	27	40	-	$1.02e - 3$
extrap5	66.4863225550	-	-	1	$8.70e - 8$
2.50e-4	66.4865775866	29	52	-	$2.55e - 4$
extrap6	66.4863226400	-	-	1	$2.00e - 9$
6.25e-5	66.4863863769	20	46	-	$6.37e - 5$
extrap7	66.4863226400	-	-	6	$2.000e - 9$
1.56e-5	66.4863385759	33	68	-	$1.59e - 5$
extrap8	66.4863226420	-	-	1	0
3.91e-6	66.4863266255	48	45	-	$3.98e - 6$
extrap9	66.4863226420	-	-	6	0
9.77e-7	66.4863236379	34	50	-	$9.96e - 7$
extrap10	66.4863226420	-	-	1	0
2.44e-7	66.4863228911	32	30	-	$2.49e - 7$
extrap11	66.4863226420	-	-	1	0
6.10e-8	66.4863227043	40	39	-	$6.23e - 8$
extrap12	66.4863226420	-	-	6	0
1.53e-8	66.4863226577	37	40	-	$1.57e - 8$
extrap13	66.4863226420	-	-	1	0
3.81e-9	66.4863226460	56	35	-	$4.00e - 9$
extrap14	66.4863226420	-	-	1	0
9.54e-10	66.4863226430	38	60	-	$1.05e - 9$
extrap15	66.4863226420	-	-	6	0
2.38e-10	66.4863226423	43	89	-	$3.42e - 10$
extrap16	66.4863226420	-	-	1	0
5.96e-11	66.4863226421	139	60	-	$1.48e - 10$
extrap17	66.4863226420	-	-	6	0

Tabela 4.3: Resultados do problema HS 116 com $\rho = \frac{1}{4}$

ρ	$f(x)$	PASSOS	CONDICIONAMENTO
$\frac{1}{4}$	66.4873424270	27	$5.0e + 15$
$\frac{1}{10}$	66.4873611934	157	$5.0e + 15$

Tabela 4.4: Corte de HS 116 em $\tau = 1.0e - 3$

Extrapolação	Antes	Depois	$\log_{10} \frac{\text{Antes}}{\text{Depois}}$
1	$1.0e - 1$	$8.0e - 4$	2.1
2	$1.0e - 2$	$8.0e - 6$	3.1
3	$1.0e - 3$	$2.0e - 7$	3.7
4	$1.0e - 4$	$4.0e - 9$	4.4

Tabela 4.5: Análise de Ganho da Extrapolação com $\rho = \frac{1}{10}$

Os resultados isolados das tabelas (4.5) e (4.6), linha a linha, mostram *in limine* a pujança do mecanismo de extrapolação, que aumenta consideravelmente a velocidade de convergência em direção à solução ótima.

Com o método de Penalização Hiperbólica aliado ao uso da extrapolação obtemos o valor ótimo $f(x^*) = 66.486323$ em $\tau = 1.0e - 5$. Na literatura o valor apresentado é 97.588409.

Podemos observar na tabela (4.2) que, para $\tau = 1.0e - 5$, temos um valor extrapolado que gera o ótimo da função objetivo. Resolvendo o problema sem o uso da extrapolação obteríamos este valor somente em $\tau = 1.0e - 11$, quando o problema já apresenta avançado estágio de degenerescência da matriz Hessiana, e conseqüentemente há maior dificuldade de se resolver o problema irrestrito. Dessa forma, graças ao uso da extrapolação, conseguimos realizar uma estimativa para a solução ótima com a mesma precisão daquela obtida na forma ortodoxa, todavia com τ assumindo um valor 10^6 vezes maior.

Extrapolação	Antes	Depois	$\log_{10} \frac{\text{Antes}}{\text{Depois}}$
1	$2.62e - 1$	$2.18e - 3$	2.1
2	$6.53e - 2$	$1.55e - 4$	2.6
3	$1.63e - 2$	$7.74e - 6$	3.3
4	$4.80e - 3$	$7.30e - 7$	3.7
5	$1.02e - 3$	$8.70e - 8$	4.1
6	$2.55e - 4$	$2.00e - 9$	5.1
7	$6.37e - 5$	$2.00e - 9$	4.5

Tabela 4.6: Análise de Ganho da Extrapolação com $\rho = \frac{1}{4}$

A utilização do mecanismo de extrapolação inversa, que estima o ótimo do sub-problema seguinte e o utiliza como ponto inicial para o processo de minimização, apresentou nesse problema, conforme o previsto na teoria, resultados eficientes em relação à redução do número de passos dados em cada subproblema para alcançar o ótimo. Por exemplo, vejamos na tabela (4.2) que para $\tau = 1.0e - 7$ o número de passos obtidos pelo método de extrapolação é bem menor que o obtido no método sem a aplicação da extrapolação, o que mostra a eficiência do método de extrapolação inversa. Nos casos em que foi obtido um número de passos maior ocorreu inviabilidade do ponto extrapolado, como em $\tau = 1.0e - 2$.

Na tabela (4.4) observamos que, no ponto de corte proposto, o valor de função objetivo fornecido por $\rho = \frac{1}{4}$ é melhor que o valor apresentado por $\rho = \frac{1}{10}$. Entretanto, verificamos, experimentalmente, que esse não pode ser um único e decisivo fator para determinar o valor que mais se adequa ao parâmetro ρ no problema, pois nas iterações seguintes verificamos que o ótimo foi atingido primeiramente quando usamos $\rho = \frac{1}{10}$. Isso mostra que devemos levar em consideração outros critérios para avaliar qual o melhor valor de ρ que se adequa ao problema.

Verificamos ainda que a resolução do problema com diferentes valores do parâmetro ρ , provoca alteração na escolha dos polinômios. Observamos que para $\rho = \frac{1}{4}$, obtivemos 12 ocorrências do polinômio de grau 1 e 5 ocorrências do polinômio de grau 6. Já com $\rho = \frac{1}{10}$ a escolha do polinômio de grau 1 ocorreu 9 vezes, enquanto, os polinômios de grau 2, 3 e 6 obtiveram somente uma escolha. Para este último caso, vejamos como ficaria o polinômio após k mínimos da forma $x(\tau)$:

$$x(\tau_k) = a_0 + a_1\tau_k + a_2\tau_k^2 + a_3\tau_k^3 + \dots + a_{k-1}\tau_k^{k-1} \quad (4.17)$$

onde $\tau_k = \rho^{k-1}\tau_0$.

Como nos termos de maior ordem tornamos a parcela que multiplica os coeficientes a_j cada vez menor, é intuitivo que os termos de mais baixa ordem, como os termos a_0 e $a_1\tau_k$, sejam mais estáveis, pois os termos restantes são mais permeáveis aos efeitos deletérios do arredondamento.

Como citamos acima, os polinômios de grau 1 prevaleceram com 70% das escolhas, o que mostra que, nesse problema, a repetida utilização de uma extrapolação linear seria suficiente para atingirmos boas aproximações da solução ótima. Observando este fato, Ariela e Sofer [13] desenvolveram trabalhos nessa linha, utilizando somente extrapolações lineares para realizar estimativas da solução ótima.

A partir dos resultados empíricos, intuimos ainda que a maior escolha dos polinômios lineares possa estar relacionada com as características do problema, como a função objetivo e as restrições ativas. Nesse problema em particular, temos a função objetivo linear com restrições ativas quadráticas que apresentam coeficientes mais expressivos nos termos lineares. Isto nos indica uma possível relação entre o

grau do polinômio escolhido e a função objetivo e as restrições ativas, porém não foi desenvolvida nenhuma teoria que sustente essa observação.

4.2 Problema Teste 2

Nessa seção apresentaremos os resultados obtidos para HS 118 [7], que trata-se de um problema com função objetivo quadrática e restrições lineares.

Assim como no problema anterior, esse problema foi implementado utilizando-se diferentes valores para o parâmetro ρ . Os valores de τ_0 foram escolhidos de maneira que pudéssemos realizar um *corte* em $\tau = 1.0e - 3$. A partir daí, para cada resolução do problema com um valor diferente de ρ garantimos que o número de extrapolações realizadas até o ponto de corte foi o mesmo, o que permite nos comparar a qualidade da solução fornecida por cada valor de ρ em igualdade de condições.

Os parâmetros iniciais para HS 118 estão na tabela (4.7) abaixo.

ρ	τ_0	λ_0
$\frac{1}{2}$	$3.2e - 2$	10
$\frac{1}{4}$	1.024	10
$\frac{1}{10}$	100	10

Tabela 4.7: Dados Iniciais de HS 118

Vejamos a definição do problema e os resultados obtidos que estão representados nas tabelas (4.8), (4.9) e (4.10).

Minimizar

$$\sum_{k=0}^4 (2.3x_{3k+1} + 0.0001x_{3k+1}^2 + 1.7x_{3k+2} + 0.0001x_{3k+2}^2 + 2.2x_{3k+3} + 0.00015x_{3k+3}^2) \quad (4.18)$$

sujeito a

$$0 \leq x_{3j+1} - x_{3j-2} + 7 \leq 13 \quad (4.19)$$

$$0 \leq x_{3j+2} - x_{3j-1} + 7 \leq 14 \quad (4.20)$$

$$0 \leq x_{3j+3} - x_{3j-3} + 7 \leq 13 \quad (4.21)$$

$$j = 1, \dots, 4$$

$$0 \leq x_1 + x_2 + x_3 - 60 \quad (4.22)$$

$$0 \leq x_4 + x_5 + x_6 - 50 \quad (4.23)$$

$$0 \leq x_7 + x_8 + x_9 - 70 \quad (4.24)$$

$$0 \leq x_{10} + x_{11} + x_{12} - 85 \quad (4.25)$$

$$0 \leq x_{13} + x_{14} + x_{15} - 100 \quad (4.26)$$

$$8 \leq x_1 \leq 21 \quad (4.27)$$

$$43 \leq x_1 \leq 57 \quad (4.28)$$

$$3 \leq x_1 \leq 16 \quad (4.29)$$

$$0 \leq x_{3k+1} \leq 90 \quad (4.30)$$

$$0 \leq x_{3k+2} \leq 120 \quad (4.31)$$

$$0 \leq x_{3k+2} \leq 120 \quad (4.32)$$

$$0 \leq x_{3k+2} \leq 60 \quad (4.33)$$

$$k = 1, \dots, 4$$

τ	$f(x(\tau))$	NPCE	NPSE	$ f(x^k) - f(x^*) $	Grau
1.0e+02	880.9686884378	30	30	$2.2e + 2$	-
1.0e+01	694.2814329129	29	29	$2.9e + 1$	-
extrap1	673.5552828400	-	-	8.7	1
1.0e+00	667.8752092251	41	43	3.1	-
extrap2	664.8567289700	-	-	$3.6e - 2$	2
1.0e-01	665.1251972513	63	41	$3.0e - 1$	-
extrap3	664.8187364900	-	-	$1.7e - 3$	3
1.0e-02	664.8509205242	19	44	$3.0e - 2$	-
extrap4	664.8204488800	-	-	$1.1e - 6$	1
1.0e-03	664.8234965791	11	37	$3.0e - 3$	-
extrap5	664.8204495100	-	-	$4.9e - 7$	1
1.0e-04	664.8207547455	16	42	$3.0e - 4$	-
extrap6	664.8204501000	-	-	$1.0e - 7$	1
1.0e-05	664.8204804770	22	48	$3.0e - 5$	-
extrap7	664.8204500000	-	-	0	6
1.0e-06	664.8204530470	33	47	$3.0e - 6$	-
extrap8	664.8204500000	-	-	0	4
1.0e-07	664.8204503047	14	39	$3.0e - 7$	-
extrap9	664.8204500000	-	-	0	1
1.0e-08	664.8204500305	28	44	$3.0e - 8$	-
extrap10	664.8204500000	-	-	0	2
1.0e-09	664.8204500031	9	55	$3.1e - 9$	-
extrap11	664.8204500000	-	-	0	2
1.0e-10	664.8204500003	24	51	$3.1e - 10$	-
extrap12	664.8204500000	-	-	0	2
1.0e-11	664.8204500000	58	33	$3.0e - 11$	-
extrap13	664.8204500000	-	-	0	3

Tabela 4.8: Resultados de HS 118 com $\rho = \frac{1}{10}$

τ	$f(x(\tau))$	NPCE	NPSE	$ f(x^k) - f(x^*) $	Grau
1.024	667.9484887459	93	93	3.1	-
2.56e-01	665.6010142848	31	31	$7.8e - 1$	-
extrap1	664.8194610200	-	-	$9.9e - 4$	1
6.40e-02	665.0154659651	16	34	$2.0e - 1$	-
extrap2	664.8203419600	-	-	$1.1e - 4$	1
1.60e-02	664.8691996794	13	37	$4.9e - 2$	-
extrap3	664.8204478800	-	-	$2.1e - 6$	1
4.00e-03	664.8326390613	9	33	$1.2e - 2$	-
extrap4	664.8204524200	-	-	$2.4e - 6$	1
1.00e-03	664.8234971175	14	25	$3.0e - 3$	-
extrap5	664.8204495800	-	-	$4.2e - 7$	5
2.50e-04	664.8212116142	4	34	$7.6e - 4$	-
extrap6	664.8204497800	-	-	$2.2e - 7$	2
6.25e-05	664.8206404219	16	27	$1.9e - 4$	-
extrap7	664.8204500200	-	-	$2.0e - 8$	1
1.56e-05	664.8204976156	5	35	$4.8e - 5$	-
extrap8	664.8204500100	-	-	$1.0e - 8$	6
3.91e-06	664.8204619022	29	29	$1.2e - 5$	-
extrap9	664.8204500000	-	-	0	6
9.77e-07	664.8204529754	23	42	$3.0e - 6$	-
extrap10	664.8204500000	-	-	0	1
2.44e-07	664.8204507439	10	30	$7.4e - 7$	-
extrap11	664.8204500000	-	-	0	1
6.10e-08	664.8204501860	14	31	$1.9e - 7$	-
extrap12	664.8204500000	-	-	0	4
1.53e-08	664.8204500465	6	38	$4.7e - 8$	-
extrap13	664.8204500000	-	-	0	2

Tabela 4.9: Resultados de HS 118 com $\rho = \frac{1}{4}$

τ	$f(x(\tau))$	NPCE	NPSE	$ f(x^k) - f(x^*) $	Grau
3.20e-02	664.9179518562	243	243	$9.8e - 2$	-
1.60e-02	664.8692003876	12	12	$4.9e - 2$	-
extrap1	664.8204507300	-	-	$7.3e - 7$	1
8.00e-03	664.8448253337	2	21	$2.4e - 2$	-
extrap2	664.8204507300	-	-	$7.3e - 7$	1
4.00e-03	664.8326383819	3	23	$1.2e - 2$	-
extrap3	664.8204515400	-	-	$1.5e - 6$	1
2.00e-03	664.8265433228	5	19	$6.1e - 3$	-
extrap4	664.8204461900	-	-	$3.8e - 6$	4
1.00e-03	664.8234969728	7	24	$3.0e - 3$	-
extrap5	664.8204506300	-	-	$6.3e - 7$	1
5.00e-04	664.8219736197	6	22	$1.5e - 3$	-
extrap6	664.8204497000	-	-	$3.0e - 7$	6
2.50e-04	664.8212118236	4	19	$7.6e - 4$	-
extrap7	664.8204499200	-	-	$8.0e - 8$	4
1.25e-04	664.8208309300	5	23	$3.8e - 4$	-
extrap8	664.8204500400	-	-	$4.0e - 8$	1
6.25e-05	664.8206404159	17	23	$1.9e - 4$	-
extrap9	664.8204498000	-	-	$2.0e - 7$	6
3.13e-05	664.8205452211	15	18	$9.5e - 5$	-
extrap10	664.8204500300	-	-	$3.0e - 8$	1
1.56e-05	664.8204976085	11	22	$4.8e - 5$	-
extrap11	664.8204499600	-	-	$4.0e - 8$	6
7.81e-06	664.8204738040	5	26	$2.4e - 5$	-
extrap12	664.8204500000	-	-	0	1
3.91e-06	664.8204619018	3	20	$1.2e - 5$	-
extrap13	664.8204500000	-	-	0	6

Tabela 4.10: Resultados de HS 118 com $\rho = \frac{1}{2}$

Extrapolação	Antes	Depois	$\log_{10} \frac{\textit{Antes}}{\textit{Depois}}$
1	$2.9e + 1$	8.7	0.52
2	3.1	$3.6e - 2$	1.94
3	$3.0e - 1$	$1.7e - 3$	2.25
4	$3.0e - 2$	$1.1e - 6$	4.4
5	$3.0e - 3$	$4.9e - 7$	3.8
6	$3.0e - 4$	$1.0e - 7$	3.48

Tabela 4.11: Análise de Ganho da Extrapolação com $\rho = \frac{1}{10}$

Extrapolação	Antes	Depois	$\log_{10} \frac{\textit{Antes}}{\textit{Depois}}$
1	$7.8e - 1$	$9.9e - 4$	2.9
2	$2.0e - 1$	$1.1e - 4$	3.3
3	$4.9e - 2$	$2.1e - 6$	4.4
4	$1.2e - 2$	$2.4e - 6$	3.7
5	$3.0e - 3$	$4.2e - 7$	3.8
6	$7.6e - 4$	$2.2e - 7$	3.5
7	$1.9e - 4$	$2.0e - 8$	4.0
8	$4.8e - 5$	$1.0e - 8$	3.7

Tabela 4.12: Análise de Ganho da Extrapolação com $\rho = \frac{1}{4}$

Discussão dos Resultados

Em todos os casos apresentados, a medida do mau condicionamento, calculado através da expressão (4.1), foi da ordem de $1.0e + 15$ avaliada no ponto $\hat{\tau}$, o que mostra que o problema está extremamente mal condicionado. Entende-se por $\hat{\tau}$ o valor de τ para o qual se obteve a melhor estimativa da solução ótima. Por exemplo, na tabela (4.8) o valor de $\hat{\tau}$ é dado por $\hat{\tau} = 1.0e - 5$.

O ganho apresentado pela extrapolção pode ser observado na tabela (4.8) na linha $\tau = 1.0e - 5$, onde foi alcançado o valor ótimo para o problema. Se o problema fosse resolvido sem a aplicação do mecanismo de extrapolção, o ótimo só seria

Extrapolação	Antes	Depois	$\log_{10} \frac{\textit{Antes}}{\textit{Depois}}$
1	$4.9e - 2$	$7.3e - 7$	4.8
2	$2.4e - 2$	$7.3e - 7$	4.5
3	$1.2e - 2$	$1.5e - 6$	3.9
4	$6.1e - 3$	$3.8e - 6$	4.2
5	$3.0e - 3$	$6.3e - 7$	3.7
6	$1.5e - 3$	$3.0e - 7$	3.7
7	$7.6e - 4$	$8.0e - 8$	4.0
8	$3.8e - 4$	$4.0e - 8$	4.0

Tabela 4.13: Análise de Ganho da Extrapolação com $\rho = \frac{1}{2}$

alcançado para $\tau = 1.0e - 10$. Podemos ainda observar que os valores das linhas de $\Delta f(x^k) = |f(x^k) - f(x^*)|$ juntamente com os resultados das tabelas (4.11), (4.12) e (4.13) mostram a eficiência e a eficácia do mecanismo de extrapolação, que ganha de 2 a 4 casas decimais de precisão em relação aos resultados obtidos sem a aplicação da extrapolação.

Um fator importante explicitado nos resultados obtidos é a redução significativa do número de passos ao longo do processo de minimização. Isso se deve à utilização do mecanismo de extrapolação inversa. Observamos na tabela (4.10) que apresenta a solução de HS 118 obtida para $\rho = \frac{1}{2}$, que o número de passos é, na maioria dos casos, $\frac{1}{3}$ menor quando a rotina de extrapolação é aplicada. O número de passos quando a extrapolação é utilizada, também é sempre menor em relação ao procedimento sem extrapolação. Em relação a $\rho = \frac{1}{10}$ e $\rho = \frac{1}{4}$ essa tendência continua sendo válida com algumas poucas exceções.

Verificamos que o polinômio linear apresentou maior frequência em todos os casos e com menor intensidade na tabela (4.8). Isto significa que o método utilizou com

mais freqüência os dois últimos pontos gerados para modelar a trajetória.

4.3 Problema Teste 3

Em Wright [17] o problema abaixo foi usado para mostrar que o passo de Newton pode ser inviável caso a extrapolação não seja usada. Esse problema também foi apresentado por Ariela [13] como aplicação de extrapolação linear. Usaremos aqui o mesmo exemplo para medir a eficiência do mecanismo de extrapolação com polinômios de até grau 6.

Definição do problema :

$$\begin{array}{ll}
 \min & -\frac{1}{8}x_1 + 2x_2 - x_3 \\
 \text{sujeito a :} & -\frac{1}{2}x_1^2 - x_2^2 - x_3^2 + \frac{41}{8} \geq 0 \\
 & \phantom{-\frac{1}{2}x_1^2} x_2^3 + 1 \geq 0 \\
 & x_1^2 + x_2^2 + x_3 - \frac{1}{2} \geq 0
 \end{array}$$

Os parâmetros iniciais para os quais o problema foi resolvido estão descritos na tabela (4.14).

ρ	τ_0	λ_0
$\frac{1}{10}$	100	10

Tabela 4.14: Dados Iniciais do problema teste 3

Para efeito de comparação, utilizamos o mesmo ponto inicial apresentado em Ariela [13]:

$$x^0 = \begin{pmatrix} 0.50108 \\ -0.99933 \\ 1.99992 \end{pmatrix}$$

τ	$\ \hat{x} - x^*\ _2$	$f(\hat{x})$	$ f(\hat{x}) - f(x^*) $	NPCE	NPSE	Grau
1	3.11e-02	-3.9929796963	6.95e-02	26	26	-
1.00e-01	3.11e-03	-4.0555447074	6.96e-03	17	17	-
extrap1	1.57e-04	-4.0624963753	3.62e-06	-	-	1
1.00e-02	3.11e-04	-4.0618044595	6.96e-04	8	15	-
extrap2	5.66e-08	-4.0625000241	2.41e-08	-	-	2
1.00e-03	3.11e-05	-4.0624304672	6.95e-05	2	15	-
extrap3	5.66e-08	-4.0625000241	2.41e-08	-	-	2
1.00e-04	3.11e-06	-4.0624930443	6.96e-06	4	19	-
extrap4	4.16e-08	-4.0624999973	2.66e-09	-	-	1
1.00e-05	3.12e-07	-4.0624993045	6.96e-07	5	15	-
extrap5	4.12e-08	-4.0625000000	1.37e-11	-	-	5
1.00e-06	5.02e-08	-4.0624999305	6.95e-08	5	23	-
extrap6	4.12e-08	-4.0625000000	1.37e-11	-	-	1
1.00e-07	4.12e-08	-4.0624999930	6.96e-09	7	27	-
extrap7	4.12e-08	-4.0625000000	1.25e-11	-	-	1
1.00e-08	4.12e-08	-4.0624999993	6.94e-10	9	16	-
extrap8	4.12e-08	-4.0625000000	1.25e-11	-	-	4
1.00e-09	4.12e-08	-4.0624999999	6.97e-11	12	23	-
extrap9	4.12e-08	-4.0625000000	1.25e-11	-	-	1
1.00e-10	4.12e-08	-4.0625000000	7.29e-12	10	22	-
extrap10	4.12e-08	-4.0625000000	1.25e-11	-	-	1

Tabela 4.15: Resultados do problema teste 3 com $\rho = \frac{1}{10}$

O problema tem solução ótima conhecida em:

$$x^* = \begin{pmatrix} 0.5 \\ -1.0 \\ 2.0 \end{pmatrix}$$

com $f(x^*) = -4.0625$. Os resultados computacionais obtidos com a utilização da extrapolação aliada à Penalização Hiperbólica estão apresentados na tabela (4.15).

Discussão dos Resultados

Podemos observar na tabela (4.15) que na segunda extrapolação realizada, extrapolação quadrática, obtemos uma precisão de 7 dígitos do valor ótimo da função objetivo.

Para $\tau = 1.0e-5$, a extrapolação com um polinômio de grau 5 gerou o resultado ótimo e a partir desse valor de τ todas as extrapolações geraram o ótimo. Caso a rotina de extrapolação não fosse executada só obteríamos mesmo resultado para $\tau = 1.0e-10$, o que significa que para τ 100 000 vezes maior conseguimos realizar uma boa estimativa da solução ótima.

O mau condicionamento do problema medido pela expressão (4.1) no ponto $\tau = 1.0e-5$ é da ordem de $1.0e+13$, enquanto para $\tau = 1.0e-10$ o condicionamento é da ordem de $1.0e+16$.

Em carácter comparativo o problema foi executado até $\tau = 1.0e-24$ com o método de penalização hiperbólica sem interrupções provocadas pelo mau condicionamento do problema. Já no artigo de Ariela [13], o método apresenta falhas já em $\tau = 1.0e-8$. Isso mostra um certo retardo no mau condicionamento da matriz Hessiana da função irrestrita proposta pelo método de Penalização Hiperbólica por Xavier [23].

Os resultados em (4.15) nos indicam também uma redução no número de passos, quando utilizado o mecanismo de extrapolação. Isto deve-se à implementação da rotina de extrapolação inversa que estima pontos iniciais próximos à trajetória central. Presumimos que esta questão ganhe maior relevância em problemas de grande porte, onde o tempo de processamento reduzido é particularmente desejado.

Resumo de Resultados

Nessa parte do trabalho, nosso objetivo é trazer um resumo geral dos resultados obtidos para os diversos problemas testados. Os resultados estão descritos na tabela

(4.16). Vejamos que em todos os casos, as estimativas do ponto ótimo apresentadas pela extrapolação são pontos viáveis, pois apresenta violação $r(x(\hat{\tau}_{CE}))$ nula.

	$\hat{\tau}_{CE}$	$f(x(\hat{\tau}_{CE}))$	$r(x(\hat{\tau}_{CE}))$	$\hat{\tau}_{SE}$	$f(x(\hat{\tau}_{SE}))$	$f_{HS}(x^*)$	$r_{HS}(x^*)$
HS19	1.0e-6	-6961.81390	0	1.0e-11	-6961.81388	-6961.81381	0
HS21	1.0e-3	-99.9600000	0	1.0e-8	-99.9600000	-99.9600000	0
HS30	1.0e-5	1.000000000	0	1.0e-10	1.000000000	1.000000000	0
HS83	1.0e-3	-30665.5387	0	1.0e-6	-30665.5387	-30665.5387	0
HS95	1.0e-7	0.015619250	0	1.0e-10	0.015619525	0.01561951	0.2e-9
HS96	1.0e-6	0.014614093	0	1.0e-11	0.015619525	0.01561951	0.2e-9
HS97	1.0e-7	3.135808700	0	1.0e-11	3.135809100	3.135809100	0
HS98	1.0e-5	3.135809100	0	1.0e-9	3.1358091000	3.13580910	0
HS108	1.0e-6	-0.86388662	0	1.0e-39	-0.863136571	-0.8660254	0.4e-9
HS113	1.0e-5	24.30620890	0	1.0e-8	24.30620910	24.3062091	0.1e-8
HS116	1.0e-5	66.48632264	0	1.0e-11	66.48632264	97.5884090	0
HS117	1.0e-5	32.34867899	0	1.0e-8	32.34867898	32.3486790	0
HS118	1.0e-5	664.8204500	0	1.0e-11	664.8204500	664.8204500	0
Ariela	1.0e-5	-4.06250000	0	1.0e-10	-4.06250000	-4.06250000	0
Lootsma	1.0e-7	1.414213562	0	1.0e-11	1.414213562	1.414213562	0

Tabela 4.16: Resultados Gerais

Na tabela (4.16), $r(x(\hat{\tau}_{CE}))$ representa a violação fornecida pelo ponto ótimo estimado pela extrapolação em $\hat{\tau}$. A coluna f_{HS} representa a solução ótima e $r_{HS}(x^*)$ representa a violação da solução ótima apresentada por Schittkowski [7]. Os valores $\hat{\tau}_{SE}$ representam os pontos onde obtemos a solução ótima sem o uso do mecanismo de extrapolação. De maneira análoga, a coluna $\hat{\tau}_{CE}$ representa os valores de τ onde foram obtidos os pontos ótimos por extrapolação.

A tabela (4.16) mostra um conjunto de resultados dos problemas o qual aplicamos a rotina de extrapolação. Podemos notar, por exemplo, que para HS108 conseguimos através do mecanismo de extrapolação realizar a estimativa para o ótimo em $\tau = 1.0e - 6$ sem violação. Isso significa, que com τ assumindo um

valor 10^{33} vezes maior, conseguimos realizar um boa estimativa da solução ótima.

Para os problemas *HS95*, *HS96*, *HS116* as soluções obtidas por extrapolação apresentam valores de função objetivo menores que os apresentados na literatura, enquanto nos outros problemas a solução é igual ou ligeiramente melhor.

A tabela (4.16) torna explícito ainda o fato de que em todos os problemas as estimativas de solução ótima geradas pela extrapolação estão em valores de τ onde o condicionamento ainda não é crítico.

4.4 Análise dos resultados

Antes de tudo, os resultados obtidos têm a natureza da preliminaridade, pois os experimentos computacionais realizados foram limitados.

Os resultados obtidos com a classe de problemas que testamos, evidenciam a eficiência do mecanismo de extrapolação nas estimativas das soluções ótimas, conforme o previsto na teoria desenvolvida nos capítulos anteriores. A redução no número de passos internos na minimização irrestrita é um resultado inequívoco e também evidencia a eficiência da implementação da rotina de extrapolação inversa acelerando o processo de convergência ao ótimo de um subproblema.

Verificamos, na apresentação dos resultados, que a variação no parâmetro ρ acarreta mudanças ao longo da minimização. Decorrente deste fato, destacamos duas questões importantes: o número de passos internos de BFGS, a cada resolução de um subproblema, é bem menor quando $\rho = \frac{1}{2}$ que nos outros casos. Essa redução lenta no parâmetro τ , faz com que a distância entre a seqüência de pontos

obtidos internamente por BFGS, seja pequena e portanto o número de passos fica reduzido. Esta é uma característica interessante e vem ainda agregada ao fato de que reduções mais lentas retardam a degenerescência da matriz Hessiana. Entretanto, esta pequena redução, pode provocar a falha na minimização. Para problemas não convexos e mal condicionados, caso o método obtenha ao longo do processo iterativo um ponto inviável, através de passos pequenos fica mais difícil escapar da região de inviabilidade. No exemplo de HS 116, um problema não convexo, a minimização não foi realizada com sucesso para $\rho = \frac{1}{2}$. Presumimos então que o fato acima citado tenha sido um dos grandes motivos para este insucesso.

Outro fato importante observado foi a escolha da extrapolação linear na maioria dos casos e dos problemas resolvidos. Vale ressaltar que em HS 118, com parâmetro $\rho = \frac{1}{10}$ obtivemos grande incidência do polinômio de grau 2. Para $\rho = \frac{1}{2}$, observamos grande incidência do polinômio de grau máximo 6, apesar de a prevalência ser, em ambos os casos, do polinômio linear.

A literatura traz, por exemplo Ariela [13], para os problemas que satisfazem as condições impostas **C1**, **C2**, **C3** e **C4** descritas no capítulo 3, o polinômio de grau 2 como o que mais se adequa na maioria dos casos. Observamos a ocorrência deste fato para problemas pequenos que satisfazem todas as condições impostas, porém para problemas onde as condições não são satisfeitas em sua plenitude, a extrapolação linear, apresentou-se como a mais eficiente. Esta pode ser também a grande razão pelo qual encontramos artigos publicados onde somente extrapolações lineares são utilizadas para obter boas estimativas do ótimo, como é o caso de Ariela e Sofer [13].

4.5 Análise de ρ

Diante da possibilidade apresentada em escolher os diferentes valores de ρ (1/2, 1/4, 1/10) , construímos um esquema para medir qual o parâmetro de redução que melhor se adequou ao problema. Considere a expressão:

$$t_i = \frac{\hat{\tau}_{CE}}{\hat{\tau}_{SE}} \quad (4.34)$$

onde i é o valor do parâmetro de redução 2, 4 ou 10, $\hat{\tau}_{CE}$ é o valor de τ para o qual obtivemos a estimativa da solução ótima no método com extrapolação e $\hat{\tau}_{SE}$ é o valor de τ para o qual obtivemos a solução ótima no método de Penalização Hiperbólica sem extrapolação.

Verificamos então:

$$\hat{t} = \max\{t_2, t_4, t_{10}\} \quad (4.35)$$

e estabelecemos o parâmetro de redução mais adequado. Com isso, priorizamos aquele parâmetro que fornece as melhores estimativas para valores altos de τ , o que é desejado em virtude do crescente mau condicionamento do problema à medida que τ se aproxima de zero.

Vejamos então a partir de (4.34), qual o melhor parâmetro de redução para este critério, no problema HS 118:

$$t_{10} = \frac{10^{-5}}{10^{-11}} = 1.0e + 6$$

$$t_4 = \frac{3.91e - 6}{1.49e - 11} = 2.62e + 5$$

$$t_2 = \frac{3.91e - 6}{1.49e - 11} = 2.62e + 5$$

De acordo com este critério, o parâmetro $\rho = \frac{1}{10}$ apresentou melhor resultado.

Utilizamos ainda outro mecanismo para analisar a melhor redução de τ . Vejamos a expressão:

$$v_i = \hat{\tau} * C * \{ \sum NPSE - \sum NPCE \} - \{ f(\hat{x}^*) - f(x(\hat{\tau})) \} \quad (4.36)$$

onde C é uma constante.

A idéia da expressão (4.36) é favorecer a escolha dos parâmetros que tiveram reduções mais significativas no número de passos. Portanto, escolhemos como melhor aquele que tem maior valor para v_i .

Pelo segundo critério (4.36) temos:

$$v_{10} = 830.00$$

$$v_4 = 578.13$$

$$v_2 = 1250.00$$

A partir dos resultados acima, podemos observar que o melhor parâmetro de redução, neste caso específico, foi $\rho = \frac{1}{2}$. Isto deve-se à incorporação do número de passos na expressão, pois ao utilizar este valor para $\rho = \frac{1}{2}$, reduzimos lentamente o parâmetro de penalização interior τ gerando pontos intermediários x bem próximos e obtemos assim um número de passos bem reduzido, aumentando o valor da expressão (4.36).

Vale ressaltar, que estes são apenas dois dos vários mecanismos que poderíamos utilizar para estabelecer a melhor redução de τ .

Vejam agora o problema HS 116, onde analisaremos os valores de t e v nas tabelas (4.2) e (4.3).

$$t_{10} = 1.0e + 06$$

$$t_4 = 4.2e + 06$$

$$\hat{t} = \max\{t_4, t_{10}\} = t_4.$$

E ainda,

$$v_4 = 11500.00$$

$$v_{10} = 130.00.$$

Observando os critérios acima, podemos concluir que para o problema HS 116, o parâmetro de redução $\rho = \frac{1}{4}$ apresentou melhores resultados na resolução do problema.

Repetindo essa análise para todos os problemas teste estudados, observamos que o valor ideal de ρ varia caso a caso. Assim, analisamos que a escolha ótima de ρ deve ser determinada empiricamente.

Conclusões e Observações

Para todos os problemas testados, observamos que o mecanismo de extrapolação apresentou um comportamento satisfatório, o que mostramos nas tabelas 4.2, 4.8, 4.9 e 4.10.

Os resultados mostraram que o mecanismo de extrapolação pode ser efetivamente usado para acelerar a convergência de métodos de otimização, amenizando os perigos do mau condicionamento.

Conforme o previsto na teoria, através do método de extrapolação obtivemos boas estimativas para o ótimo. No problema HS116, em particular, os resultados obtidos pela extrapolação foram bastante relevantes em relação aos resultados apresentados na literatura.

Apesar de muitos problemas apresentarem elevado número de condição na matriz Hessiana da função irrestrita penalizada, o que pode dificultar e até mesmo impedir o sucesso da minimização, verificamos que, na grande maioria das vezes, com τ da ordem de 10^{-6} , já temos uma boa estimativa da solução ótima utilizando o esquema de extrapolação de Richardson-Romberg. Fiacco [3] narra em seu trabalho, que os erros computacionais são crescentes e podem ser críticos para estimativas de mais

alta ordem, o que verificamos em nosso experimento.

Obtivemos bons resultados ainda, na utilização da extrapolação inversa associada à extrapolação. Esta tinha o objetivo de gerar bons pontos iniciais para a resolução de cada subproblema penalizado, ou seja, gerar pontos bem próximos à trajetória central e acelerar o processo de convergência em direção ao ótimo do subproblema penalizado. O sucesso desse processo foi comprovado empiricamente pela redução do número de passos na minimização interna de BFGS. Tomemos a tabela 4.10 como caso característico. Nas duas primeiras iterações, como não há extrapolação, há paridade no número de passos. Já nas iterações seguintes, podemos observar que o ganho é bastante relevante. Entretanto, para $\tau = 6.25e-5$ e nas iterações seguintes, o ganho no número de passos não foi tão relevante. Conjectura-se que o mesmo não tenha sido tão marcante, como o esperado, devido aos erros de processamento proporcionalmente maiores no cálculo das diferenças no esquema de Richardson-Romberg.

Na escolha dos polinômios durante a extrapolação, pudemos observar nos exemplos numéricos, que para problemas onde não há convexidade, a extrapolação linear obteve vantagem em relação aos outros polinômios de até grau 6, grau máximo em que a extrapolação mostrou eficácia. Essa estrutura de permitir grau máximo 6 dos polinômios foi criada após observarmos que o ganho de função objetivo para polinômios de ordem mais elevada não eram tão relevantes.

A partir dos casos testados, conjecturamos que a escolha dos polinômios possa estar relacionada ao grau da função objetivo e ao grau das restrições ativas. Em grande parte dos problemas observamos a ocorrência de tal fato.

Em trabalhos futuros, podemos estabelecer aprimoramentos do estudo da trajetória central. Podemos ainda utilizar o mecanismo de extrapolação para os multiplicadores de Lagrange, não só para realizar estimativas mais precisas, como pode ser usado em articulação com o método do Lagrangeano Hiperbólico proposto por Xavier [20] em 1992.

Outro aspecto importante, que pode vir a ser explorado no futuro é tentar estabelecer uma conexão teórica entre a característica do problema definida pela função objetivo com as restrições ativas e o grau do polinômio escolhido na extrapolação. A ampliação dos experimentos computacionais também deve ser feita através da realização de testes com problemas maiores e consultas a outras fontes.

A extensão dos resultados teóricos mostrados por Ariela [13], quanto à viabilidade do ponto extrapolado para problemas convexos, também é um ponto de interesse. Nesse artigo, Ariela mostra que usando-se o método de Newton e realizando-se sempre extrapolações lineares, os pontos extrapolados obtidos são sempre pontos viáveis. Conjecturamos que resultado análogo possa ser também válido em nosso caso, pois os resultados computacionais preliminares indicaram essa validade.

Observamos que nos problemas que satisfazem as condições de **C1** a **C4**, os pontos extrapolados obtidos são pontos viáveis, assim como os pontos extrapolados inversos. Não encontramos nenhuma base teórica que sustente a hipótese de que em problemas convexos os pontos gerados pela extrapolação sejam sempre viáveis. No trabalho de Ariela [13], o que há desenvolvido é a garantia de que ao utilizarmos o método de Newton na minimização irrestrita, as *extrapolações lineares* produzem sempre pontos viáveis. Conjecturamos que isso seja válido também para o nosso

caso, o que pode ser investigado em trabalhos futuros. Adicionalmente, o exemplo HS116, nos indica que a extrapolação deve ser usada acompanhada com teste de viabilidade.

Referências Bibliográficas

- [1] APOSTOL, T.M., *Calculus*, 2 ed., vol. 2, Wiley International Edition, 1969.
- [2] AVRIEL, M., *Nonlinear Programming Analysis and Methods*. Englewood Cliffs, Prentice-Hall, 1976.
- [3] FIACCO, A.V., McCORMICK, G.P., “Extensions of SUMT for Nonlinear Programming: Equality Constraints and Extrapolation”. *Management Science*, Vol. 12, nº 11, pp. 816-828, jul. 1966.
- [4] FLETCHER, R., *Practical Methods of Optimization*. 2 ed., Wiley , 1987.
- [5] GILL, P.E., MURRAY, W., Wright, M.H., *Practical Optimization*, 1 ed. , Academic Press, 1981.
- [6] HENRICI, P., *Applied and Computacional Complex Analysis*. vol. 2 ,New York, John Wiley and Sons, 1977.
- [7] HOCK, W., SCHITTCKOWSKI, K., *Test Examples for Nonlinear Programming Codes*. Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, 1981.
- [8] KARMAKAR, N.K., “A New Polynomial Time Algorithm for Linear Programming”. *Combinatorica*, nº 4, pp. 373-395, 1984.

- [9] LUENBERGER, D.G., *Introduction to Linear and Nonlinear Programming*. Stanford University, Addison Wesley, 1973.
- [10] LOOTSMA, F.A., "Hessian matrices of penalty functions for solving constrained optimization problems". *Philips Res. Repts 24*, Eindhoven, The Netherlands, pp.322-331, 1969.
- [11] LOOTSMA, F.A., "Extrapolation in Logarithmic Programming". *Philips Res. Repts. 23*, pp. 108-116, 1968.
- [12] MARTINEZ, J.M., SANTOS, S.A., "Métodos Computacionais de Otimização". 20^o Colóquio Brasileiro de Matemática, IMPA , pp. 24-28, jul, 1995.
- [13] NASH, S.G., SOFER, A. - "Why Extrapolation Helps Barrier Methods". Disponível em : <http://www.gmu.edu/departments/ore/sofer.html>. Acesso em: 30 jul. 2000.
- [14] POLAK, E., *Computational Methods in Optimization*. Mathematics in Science and Engineering, vol. 77, Academic Press, 1971.
- [15] SIDI, A., "A User-Friendly Extrapolation Method for Oscillatory Infinity Integrals". *Mathematics of Computation*, vol. 51, n^o 183, pp.249-266, jul. , 1988.
- [16] SIDI, A., "Some Properties of a Generalization of the Richardson Extrapolation Process". *Journal of the Institute of Mathematics and its Applications* , Vol. 24, pp. 327-346, mar. , 1979.

- [17] WRIGHT, M. H., “Why a pure primal Newton barrier step may be infeasible”. *SIAM Journal of Optimization*, vol.5, n^o 1, pp. 1-12, 1995.
- [18] XAVIER, A.E., “Hyperbolic Penalty: A New Method for Nonlinear Programming With Inequalities”. In: *International Transaction in Operational Research*, n^o. 8, pp. 1-13, 2001.
- [19] XAVIER, A.E., OLIVEIRA, A. A. F., *Optimum Covering of Plane Domains by Circles Via Hyperbolic Smoothing Method: Computational Results*. Dept. of Engineering and Computer Science, COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, 1999.
- [20] XAVIER, A.E., MACULAN, N., “Hyperbolic Lagrangean: A New Method of Multipliers”. *Technical Report*, COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, 1995.
- [21] XAVIER, A.E., *Penalização Hiperbólica*. Tese de D.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, 1992.
- [22] XAVIER, A.E., MACULAN, N., “Extrapolação em Penalização Hiperbólica”. In: *Anais do II Congresso Latino Americano de Investigacion Operativa e Ingenieria de Sistemas*, pp. 24-38, Buenos Aires, 1984.
- [23] XAVIER, A.E., *Penalização Hiperbólica - Um Novo Método para Resolução de Problemas de Otimização*. Tese de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, 1982.