

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA ELETRICA
DEPARTAMENTO DE TELEMATICA

PROBLEMA DE AUTOVALORES, OTIMIZAÇÃO
DE FUNÇÕES MATRICIAIS E ROBUSTEZ DE
SISTEMAS DINAMICOS : UMA ABORDAGEM
ALGORITMICA

Adilson Oliveira do Espírito Santo

Orientador:

Prof. Dr. Akebo Yamakami

Tese submetida à Faculdade de
Engenharia Elétrica da Univer-
sidade Estadual de Campinas
como parte dos requisitos
exigidos para obtenção do
título de DOUTOR em Engenha-
ria Elétrica.

- Fevereiro 1988 -

*Este exemplar
corresponde à redação final
da tese defendida pelo
Adilson O. Espírito Santo
e aprovada pela Comissão Julgadora
em fevereiro de 1988.*

Yamakami

A Lucidalva,

ao Alan,

à Aline,

ao Alex.

AGRADECIMENTOS

Temos uma enorme dívida de gratidão com numerosas pessoas e instituições, que contribuíram de forma decisiva para que aprendêssemos a pesquisar. São tantos que é impossível citar todos, mas queremos expressar nossos agradecimentos para aqueles cujo auxílio foi decisivo para realização deste trabalho.

Ao professor Paulo Roberto Oliveira de Souza da UFPa pela oportunidade de dar continuidade aos meus estudos e aos professores José Mario Martinez, José Claudio Geromel e Secundino Soares Filho pelas discussões sempre proveitosas.

Ao professor Mario Serra, já falecido, pelo exemplo de dedicação à pesquisa.

De modo especial agradeço ao meu orientador, Akebo Yamakami, pela paciência e fundamentalmente pela sua amizade.

Aos colegas Adriano, Celso Carnieri, Celso Coelho, Hélio, Herminio, Lourival, Marinho, Renato e Tadeu que durante os anos de pós-graduação muito me ajudaram.

Aos meus pais e meus irmãos pelo incentivo e apoio nos momentos decisivos.

A Universidade Federal do Pará, UFPa, e à CAPES pelo auxílio financeiro durante minha permanência na UNICAMP.

Índice

Introdução Geral.....	1
Capítulo 1. Problema de Autovalores, Minimização e Estabilidade de Sistemas Dinâmicos: Resultados Preliminares.....	3
1.1. Introdução.....	3
1.2. Autovalores e Autovetores.....	4
1.3. Normas de Vetores e Matrizes.....	10
1.4. Localização de Autovalores.....	15
1.5. Transformação de Householder e Rotações Planas.....	17
1.6. Sequência de Sturm.....	22
1.7. Funções Escalares com Sinal Definido.....	26
1.8. Conjuntos e Funções Convexas.....	27
1.9. Método Gradientes Conjugados.....	31
1.10. Sistemas Dinâmicos : Estabilidade.....	36
Capítulo 2. Problema Completo e Especial de Autovalores de uma Matriz Real Simétrica.....	48
2.1. Introdução.....	48
2.2. Cálculo do Autosistema de uma Matriz Simétrica.....	50
2.2.1 Algoritmo.....	50
2.2.2 Redução a Forma Tridiagonal.....	51
2.2.2.a Relações que Definem o Método	52

2.2.3	Cálculo dos Autovalores das Matrizes T e A	58
2.2.3.a	Menores Principais da Matriz $(A - T)$	59
2.2.3.b	Uma iteração do Método da Bisseção.....	62
2.2.3.c	Intervalo Inicial.....	63
2.2.4	Cálculo dos Autovetores da Matriz A	64
2.2.5	Exemplos.....	67
2.3.	Um Problema Especial de Autovalores.....	70
2.3.1	Minimização (Maximização) da Função R	70
2.3.2	Algoritmos.....	73
2.3.2.a	Algoritmo $A1$	74
2.3.2.b	Algoritmo $A2$	77
2.3.3	Implementação dos Algoritmos $A1$ e $A2$	81
2.3.3.a	Escolha do ponto Inicial.....	81
2.3.3.b	Uso da Homogeneidade da Função R	82
2.3.3.c	Escolha de h no Algoritmo $A1$	82
2.3.3.d	Busca Unidimensional em $A2$	82
2.3.3.e	Reinicialização Prematura em $A2$	83
2.3.4	Exemplos.....	84

Capítulo 3. Problemas de Otimização com Restrições Matriciais

	Semidefinidas.....	86
3.1.	Introdução.....	86
3.2.	Restrição Matricial Semidefinida.....	87
3.3.	Método dos Hiperplanos de Cortes.....	89
3.3.1	Vetor que Caracteriza o Corte.....	90
3.4.	Problema do Teste Educacional.....	94

3.4.1	Expressão do Vetor $c(y)$ e suas Propriedades.....	96
3.4.2	Algoritmo.....	98
3.4.3	Exemplos.....	100
3.4.4	Convergência.....	104
3.5.	Solução Diagonal Positiva da Equação de Lyapunov.....	107
3.5.1	Expressão do Vetor $c(y)$ e suas Propriedades.....	108
3.5.2	Grau de Estabilidade.....	110
3.5.3	Algoritmo.....	113
3.5.4	Exemplos.....	114

Capítulo 4. Robustez de Sistemas Dinâmicos Lineares Contínuos

	no Tempo.....	117
4.1.	Introdução.....	117
4.2.	Condições Suficientes.....	118
4.3.	Análise Numérica de $D(B)$	124
4.3.1	Algoritmo.....	125
4.4.	Exemplos.....	127
	Conclusões Gerais.....	130
	Referências.....	133

RESUMO: Neste trabalho abordamos questões referentes ao problema de autovalores e autovetores de uma matriz simétrica, otimização de funções matriciais e de robustez de sistemas dinâmicos lineares contínuos no tempo.

O Problema de autovalores e autovetores é abordado segundo dois pontos de vista distintos: decomposição da matriz do sistema onde sugerimos uma nova implementação para o cálculo dos autovetores e otimização da função quociente de Rayleigh onde dois novos algoritmos baseados numa combinação dos métodos de Newton e gradientes conjugados são apresentados.

Para resolver uma classe de problemas de otimização de funções matriciais, é sugerido uma metodologia baseada no método dos hiperplanos de corte e aplicada a dois problemas disponíveis na literatura, o problema do teste educacional que aparece em estatística e a determinação da solução diagonal positiva da equação de Lyapunov.

Sobre a robustez de sistemas dinâmicos lineares contínuos no tempo são fornecidas condições suficientes para existência de uma matriz constante de ganhos de realimentações, de maneira que o sistema de malha fechada seja robusto quanto a inserção no modelo de perturbações não lineares dependentes do estado. Para determinação da matriz de ganho propomos um procedimento numérico.

ABSTRACT: In this work we analyse three problems. In the first, we present some algorithms to solve the eigenvector and eigenvalue problems of the symmetric matrix, in the second we analyse the optimization problems with matricial constraints and finally in the third some robustness properties of linear continuous time dynamic systems are studied.

The eigenvector and eigenvalue problems are analysed using two methods: the decomposition technique on the matrix of the system and the optimization of the Rayleigh quotient. In the second one we propose two algorithms based on the Newton method and conjugated gradient method.

To solve a class of the optimization problems with matricial constraints we propose a methodology based on the cutting plane technique. Two examples are treated.

Finally, for a given linear continuous time dynamic systems we determine sufficient conditions for the existence of a constant feedback matrix such that the closed-loop system is robust in the sense that the perturbed system is asymptotically stable.

Introdução Geral

A realização deste trabalho foi motivada pelo estudo da estabilidade de sistemas dinâmicos. Aqui abordamos algumas questões que vem surgindo recentemente na literatura, como consequência das pesquisas sobre a estabilidade de sistemas dinâmicos.

Dentre as questões que despertaram nosso interesse, temos aquela de otimizar uma função matricial sujeito a restrições matriciais do tipo semidefinidas.

Outra diz respeito a robustez do sistema, através da qual é possível estabelecer condições para um sistemas dinâmico, sujeito a perturbações não-lineares dependentes do estado, permanecer estável.

Imbuído da vontade de resolver tais questões, deparamos com a necessidade de abordar problemas referentes ao estudo de autosistemas de uma matriz, isto é, calcular seus autovalores e respectivos autovetores.

As três questões citadas são o tema principal deste trabalho. Deste modo, procuramos abordá-las segundo uma ordem que na nossa opinião é a mais adequada para apresentação, apesar de diferir bastante da maneira como elas foram trabalhadas ao longo do tempo.

Assim, dividimos o trabalho em quatro capítulos. No primeiro, definições, teoremas e as notações foram estabelecidos. Podemos considerar este capítulo, como uma coletânea de resultados que fundamentam aqueles que serão apresentados nos capítulos subsequentes.

No capítulo dois, iniciamos a apresentação daquilo que pode ser considerado a essência deste trabalho. Lá tratamos o problema de calcular o autosistema de uma matriz real e simétrica, sob o ponto de

vista algorítmico. Duas questões são abordadas. Primeiro o problema de calcular o autovalor completo da matriz. Depois um problema especial, o de calcular os autovalores e autovetores associados localizados nas extremidades do espectro da matriz.

O terceiro capítulo concerne, ainda sob o ponto de vista algorítmico, à solução de uma classe de problemas de otimização matricial com restrições do tipo semidefinidas, cujas variáveis estão localizadas na diagonal principal da matriz incógnita. Uma metodologia baseada no método de planos de corte é proposta para resolver estes problemas.

Ainda com referência a este capítulo, dois problemas existentes na literatura são resolvidos utilizando esta metodologia. O primeiro diz respeito à confiabilidade de um exame constituído de vários subtestes, aos quais um grupo de estudantes é submetido. Este problema aparece na área de estatística e é denominado problema do Teste Educacional. O segundo problema é referente à investigação quanto à existência ou não de uma solução diagonal positiva da equação matricial de Lyapunov. Este problema aparece muito frequentemente em economia e sistemas de controle.

O último capítulo deste trabalho é dedicado ao estudo de robustez de sistemas dinâmicos lineares contínuos invariantes no tempo. Condições suficientes para existência de uma matriz de ganho de realimentação, tal que a matriz do sistema de malha fechada admita uma solução diagonal positiva da equação de Lyapunov são apresentadas. Um procedimento numérico, para calcular a matriz de ganho, é proposto.

Finalmente, algumas conclusões gerais e novas propostas para futuros desenvolvimentos encerram o trabalho.

Capítulo 1

Problema de Autovalores, Otimização e Estabilidade de Sistemas Dinâmicos: Resultados Preliminares

1.1. Introdução

O propósito principal deste capítulo, é apresentar os conceitos e definições utilizados ao longo do trabalho.

Uma coletânea de resultados, obedecendo um encadeamento lógico, é exposta com intuito de tornar a leitura do texto um processo menos árido. Exemplos explicativos são fornecidos, naqueles pontos onde achamos que era necessário esclarecer mais detalhadamente alguma ou outra idéia que influenciaram, decisivamente, nos resultados obtidos nos capítulos subsequentes.

Sobre o problema de autovalores e autovetores de uma matriz real, apresentamos resultados que conduzem à análise do autosistema da matriz, segundo dois pontos de vista, distintos. Primeiro com relação a decomposição da matriz e segundo, através de uma abordagem variacional.

Em se tratando de um trabalho em que predomina o aspecto algorítmico, os conceitos de normas vetoriais e matriciais não poderiam deixar de ser apresentados, inclusive com certa extensão. De imediato, estes conceitos foram aplicados para localização de autovalores. Com o mesmo objetivo, apresentamos um resultado devido a Gerschgorin.

Algumas transformações elementares, como transformação de Householder, rotações no plano e a definição de sequência de Sturm, serão

utilizadas na elaboração de um algoritmo para resolver o problema completo ou parcial de autovalores de uma matriz simétrica. Um resumo destes conceitos será fornecido nas seções 5 e 6.

Funções escalares com sinal definido, conjuntos e funções convexas e resultados correlacionados, aparecem nas seções 7 e 8.

O método dos gradientes conjugados para funções quadráticas, é resumido na seção 9. Ele serve de referência para os dois últimos algoritmos apresentados no próximo capítulo.

Completando o capítulo, na última seção, os conceitos de ponto de equilíbrio, estabilidade, estabilidade assintótica referentes ao estudo de sistemas dinâmicos, serão fornecidos. Os teoremas sobre estabilidade de sistemas dinâmicos, discretos e contínuos no tempo, devidos a Lyapunov, serão também incluídos. Para finalizar temos a definição de matriz estável e diagonalmente estável.

1.2. Autovalores e Autovetores

Um problema que aparece frequentemente nas aplicações da Álgebra Linear, é aquele de encontrar valores de um parâmetro escalar λ , para os quais existem vetores x , não-nulos, do \mathbb{R}^n que satisfazem a equação:

$$Ax = \lambda x \quad (1.2.1)$$

onde A é uma matriz de ordem n .

Def.1.2.1 - Um parâmetro escalar λ em (1.2.1), é um autovalor (valor característico) da matriz $A(n \times n)$ e o correspondente vetor não-nulo x é um autovetor (vetor característico) de A .

Def.1.2.2 - O conjunto $W = \{(\lambda_i, x_i) : \lambda_i \text{ é um escalar e } x_i \in \mathbb{R}^n, i=1,2,\dots,n\}$ é chamado autossistema da matriz $A(n \times n)$.

Para um dado valor de λ a equação (1.2.1) é equivalente a equação linear homogênea

$$(A - \lambda I)x = 0 \quad (1.2.2)$$

onde I é a matriz Identidade de ordem n . Sendo assim, uma condição necessária e suficiente para que exista uma solução não-trivial de (1.2.2) é que

$$\det(A - \lambda I) = 0. \quad (1.2.3)$$

Desenvolvendo este determinante, obtemos um polinômio de grau n em λ .

Def.1.2.3 - A equação (1.2.3) é denominada equação característica da matriz A e o polinômio de grau n , $P(\lambda) = \det(A - \lambda I)$, polinômio característico de A .

Pelo teorema fundamental da álgebra, sabemos que uma equação polinomial de grau n , $P(\lambda)=0$, tem n raízes, nem todas necessariamente distintas. Se uma raiz é contada um número de vezes igual à sua multiplicidade, então existem n raízes que podem ser tanto reais como complexas.

Neste trabalho a matriz do sistema (1.2.1) é sempre real e simétrica. Assim, podemos afirmar que as raízes do polinômio característico $P(\lambda) = 0$, são sempre números reais e que existe um conjunto

completo de n autovetores x_i , $i=1,2,\dots,n$ que forma uma base do \mathbb{R}^n .

Teo.1.1 : Se a matriz real A na equação (1.2.1) é simétrica, os autovalores de A são reais.

Prova: Suponhamos que λ é um autovalor da matriz A . Então existe pelo menos um autovetor x tal que

$$Ax = \lambda x \quad (1.2.4)$$

Tomando o conjugado complexo de (1.2.4), obtemos

$$A\bar{x} = \bar{\lambda} \bar{x} \quad (1.2.5)$$

(A é uma matriz real). Pré-multiplicando (1.2.4) por \bar{x}' , (1.2.5) por x' e subtraindo membro a membro temos:

$$\bar{x}'Ax - x'A\bar{x} = (\lambda - \bar{\lambda})x'\bar{x} \quad (1.2.6)$$

dado que $\bar{x}'x = x'\bar{x}$. Como A é simétrica, $\bar{x}'Ax = x'A\bar{x}$ e $x'\bar{x}$ é real e positivo. Concluimos de (1.2.6) que $\lambda = \bar{\lambda}$, sendo assim um número real.

Teo.1.2 : Sejam λ_i , $i=1,2,\dots,m$, autovalores distintos da matriz simétrica A . Então o conjunto de autovetores correspondentes x_i , $i=1,2,\dots,m$ é linearmente independente.

Prova : Vamos supor que o conjunto de autovetores é linearmente dependente. Então existem combinações lineares não-nulas desses vetores que são iguais ao vetor zero. Entre essas combinações lineares, vamos selecionar aquela que tem o número mínimo de coeficientes não-nulos. Sem perda de generalidade, pode ser assumido que esses coeficientes correspondem aos primeiros k autovetores, e que o primeiro coeficiente é igual a 1. Isto é, a relação é da forma

$$x_1 + \sum_{i=2}^k a_i x_i = 0. \quad (1.2.7)$$

para algum conjunto de a_i 's, $i=1,2,\dots,k$, $a_i \neq 0$.

Multiplicando (1.2.7) por A obtemos

$$Ax_1 + \sum_{i=2}^k a_i Ax_i = 0. \quad (1.2.8)$$

Usando o fato que os x_i 's são autovetores, (1.2.8) é equivalente a

$$\lambda_1 x_1 + \sum_{i=2}^k a_i \lambda_i x_i = 0 \quad (1.2.9)$$

Multiplicando (1.2.7) por λ_1 e subtraindo de (1.2.9) temos

$$\sum_{i=2}^k a_i (\lambda_i - \lambda_1) x_i = 0$$

Portanto, é uma combinação linear de somente $(k-1)$ termos, contradizendo a hipótese de k ser o número mínimo de coeficientes não-nulos. Donde concluímos, que o conjunto de autovetores x_i , $i=1,2,\dots,m$ é linearmente independente.

Pelo teo.1.2 podemos concluir que se os n autovalores determinados através do polinômio característico de uma matriz simétrica, são todos distintos, o conjunto de n autovetores associados formam uma base para o \mathbb{R}^n . No caso de existirem autovalores repetidos, ainda assim é possível determinar um conjunto de n autovetores associados linearmente independente [33].

Teo.1.3 : Se a matriz real A na equação (1.2.1) é simétrica, os autovetores correspondentes a autovalores distintos são ortogonais.

Prova: Sejam (λ_i, x_i) e (λ_j, x_j) dois elementos distintos do autosistema da matriz A . Por hipótese

$$Ax_i = \lambda_i x_i \quad \text{e} \quad Ax_j = \lambda_j x_j$$

então,

$$x_j' Ax_i = \lambda_i x_j' x_i \quad \text{e} \quad x_i' Ax_j = \lambda_j x_i' x_j$$

Subtraindo membro a membro estas igualdades e usando o fato de $x_j' Ax_i = x_i' Ax_j$, obtemos

$$(\lambda_i - \lambda_j) x_j' x_i = 0$$

Como $\lambda_i \neq \lambda_j$, temos que $x_j' x_i = 0$. Logo x_i e x_j são ortogonais.

Em geral, o comprimento de um autovetor não é de interesse. Por conveniência, sempre assumiremos que os autovetores de uma matriz simétrica são de comprimento unitário, isto é, eles são normalizados. Um conjunto de dois ou mais autovetores x_i normalizados e correspon-

dentes a diferentes autovalores satisfazem a equação $x_i'x_j = \delta_{ij}$. Qualquer conjunto de vetores satisfazendo esta equação é chamado um conjunto ortonormal.

Os resultados obtidos até agora dizem respeito somente ao número de autovalores e autovetores de uma matriz, no caso, real e simétrica, ficando a questão : como determinar cada um deles explicitamente?

Quando a ordem da matriz assume valores maiores ou igual a 5, em geral, precisamos usar um método numérico para determinar um, alguns ou todos os elementos de seu autosistema. Vários enfoques podem ser dados para resolver este problema. Um deles é através da decomposição da matriz do sistema, A , em matrizes mais simples, para efeito de cálculos, e que tenham os mesmos autovalores da matriz original (matrizes similares).

Outro enfoque que pode ser dado ao problema de autovalores é através do uso de técnicas de otimização. Para ver como isto é possível, vamos tomar a equação (1.2.1) e multiplicar ambos os membros pelo vetor não-nulo x'

$$x'Ax = \lambda x'x \quad x \neq 0 \quad (1.2.10)$$

temos então

$$\lambda = (x'Ax)/(x'x) \quad x \neq 0 \quad (1.2.11)$$

Def.1.2.4 - A função de valor escalar $R : \mathbb{R}^n \rightarrow \mathbb{R}$, denominada quociente de Rayleigh, é definida por

$$R(x) = (x'Ax)/(x'x) \quad x \neq 0. \quad (1.2.12)$$

Explorando algumas propriedades, peculiares, do quociente de Rayleigh, é possível resolver alguns problemas especiais de autovalores e autovetores [36].

Ainda outra maneira com que podemos abordar o problema de autovalores, será através do cálculo do polinômio característico da matriz A , sendo esta abordagem desaconselhada na maioria dos casos, a não ser em situações muito propícias.

1.3. Normas de Vetores e Matrizes

Para fazer uso de processos numéricos iterativos na solução de problemas práticos, precisamos de uma forma ou de outra, examinar a convergência de seqüências de vetores e matrizes geradas por esses processos. Daí, concluímos sobre a necessidade de um conceito geral de distância. Este conceito se torna mais evidente através da definição de norma de vetores e matrizes.

Def.1.3.1 - Uma seqüência de vetores do R^n , $(x^{(i)}, i=1,2,\dots)$, converge para o vetor x , se cada uma das seqüências formadas pelos componentes dos vetores $x^{(i)}$ converge para respectiva componente do vetor x .

Um tipo especial de seqüência de vetores convergentes, que vamos explorar muito, é aquele de seqüência de vetores convergentes direcionalmente.

Def.1.3.2 - Uma seqüência de vetores do R^n , $(x^{(i)}, i=1,2,\dots)$ é dita ser direcionalmente convergente se a seqüência de vetores normalizados $(x^{(i)}/c_i, i=1,2,\dots)$, com c_i 's escalares não-nulos, converge como na

definição 1.3.1.

Def.1.3.3 - A norma $N(x) = \|x\|$ de um vetor x , é uma função escalar real que satisfaz as seguintes condições :

$$(1) - \|x\| \geq 0 \quad \text{com } \|x\| = 0 \text{ somente para } x = 0$$

$$(2) - \|c \cdot x\| = |c| \cdot \|x\| \quad \text{para todo escalar } c$$

$$(3) - \|x + y\| \leq \|x\| + \|y\|$$

Através da def.1.3.3, qualquer função escalar que satisfaz as condições (1-3) pode ser usada como uma norma vetorial.

Exemplo 1: Algumas normas vetoriais usadas neste trabalho são :

$$\|x\|_1 = \sum_{i=1}^n |x_i| \quad (1.3.1)$$

$$\|x\|_2 = \left(\sum_{i=1}^n |x_i|^2 \right)^{1/2} \quad (1.3.2)$$

$$\|x\|_\infty = \max_i |x_i| \quad (1.3.3)$$

As normas 1, 2 e ∞ são também conhecidas respectivamente como norma octaedral, esférica ou euclideana e cúbica.

Usando a definição de norma vetorial, dizemos que uma sequência de vetores do \mathbb{R}^n , $(x^{(i)}, i=1, 2, \dots)$, converge para $x \in \mathbb{R}^n$ se e somente se

$$\lim_{i \rightarrow \infty} \|x^{(i)} - x\| = 0 \quad (1.3.4)$$

para qualquer norma arbitrária.

Def.1.3.4 - Uma sequência de matrizes $(A^{(i)}, i=1,2,\dots)$ converge para matriz A , se cada uma das seqüências formadas pelos elementos das matrizes $A^{(i)}$ converge para o respectivo elemento da matriz A .

Def.1.3.5 - A norma $N(A) = \|A\|$ de uma matriz A é uma função escalar real que satisfaz as seguintes condições :

(1) - $\|A\| \geq 0$, com $\|A\| = 0$ somente para $A = 0$

(2) - $\|c.A\| = |c|. \|A\|$ para todo escalar c

(3) - $\|A + B\| \leq \|A\| + \|B\|$

(4) - $\|A.B\| \leq \|A\|. \|B\|$

Exemplo 2: Algumas normas matriciais possíveis são :

$$\|A\|_1 = \max_j \sum_{i=1}^n |a_{ij}| \quad (1.3.5)$$

$$\|A\|_2 = \left(\sum_{i,j=1}^n |a_{ij}|^2 \right)^{1/2} \quad (1.3.6)$$

$$\|A\|_\infty = n \cdot \max_{i,j} |a_{ij}| \quad (1.3.7)$$

De modo análogo ao caso de vetores, dizemos que uma seqüência de matrizes $(A^{(i)}, i=1,2,\dots)$ converge para matriz A se e somente se

$$\lim_{i \rightarrow \infty} \|A^{(i)} - A\| = 0 \quad (1.3.8)$$

para qualquer norma arbitrária.

Def.1.3.6 - Uma norma matricial, $\|A\|$, é compatível com uma norma vetorial $\|x\|$ se para todo vetor x , se verifica :

$$\|Ax\| \leq \|A\| \cdot \|x\| \quad (1.3.9)$$

Def.1.3.7 - A função escalar real definida por

$$N(A) = \max_{x \neq 0} \|Ax\| / \|x\| = \max_{\|x\|=1} \|Ax\| \quad (1.3.10)$$

é uma norma matricial subordinada a norma vetorial $\|x\|$.

Exemplo 3: A norma matricial subordinada a norma vetorial esférica ou Euclídeana é $\|A\| = \sqrt{\lambda_{\max}(A'A)}$, onde λ_{\max} é o maior autovalor de $A'A$. Esta norma é chamada norma espectral de A .

Para mostrar isso, através de (1.3.2) temos que $\|x\|^2 = (x, x)$ e por (1.3.10),

$$\begin{aligned} N(A)^2 &= \max_{\|x\|=1} \|Ax\|^2 \\ &= \max_{\|x\|=1} (Ax, Ax) \\ &= \max_{\|x\|=1} (x, A'Ax) \end{aligned}$$

a matriz $A'A$ é simétrica e positiva semidefinida (ver seq.1.7), desde que $(Ax, Ax) \geq 0$ para todo vetor x . A matriz $A'A$ tem n autovalores reais, não negativos $\lambda_{\max} = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n = \lambda_{\min}$ e um correspondente sistema ortonormal de autovetores x_1, x_2, \dots, x_n . Um vetor qualquer x , normado pode ser representado por uma combinação linear dos autovetores

$$x = c_1 x_1 + c_2 x_2 + \dots + c_n x_n.$$

A ortogonalidade dos autovetores produz

$$\|x\|^2 = (x, x) = c_1^2 + c_2^2 + \dots + c_n^2 = 1.$$

Podemos assim, estimar o valor de $(x, A'Ax)$

$$\begin{aligned} (x, A'Ax) &= \left(\sum_{i=1}^n c_i x_i, A'A \sum_{i=1}^n c_i x_i \right) \\ &= \left(\sum_{i=1}^n c_i x_i, \sum_{i=1}^n c_i \lambda_i x_i \right) \\ &= \sum_{i=1}^n c_i^2 \lambda_i \\ &< \lambda_1 \sum_{i=1}^n c_i^2 \\ &< \lambda_1 \end{aligned}$$

especificamente, $(x_1, A'Ax_1) = \lambda_1$ tal que $N(A)^2 = \max_{\|x\|=1} (x, A'Ax) = \lambda_1$.

1.4. Localização de Autovalores

Em alguns métodos numéricos, para resolver o problema de autovalores, há necessidade da determinação de um intervalo que contenha todos os autovalores da matriz do problema. Com este objetivo, duas formas distintas, mas muito simples, são apresentadas. A primeira baseada na definição de norma matricial e a segunda usando os conhecidos círculos de Gerschgorin.

Teo.1.4 : Para toda norma matricial, $\|A\|$ é um limite superior para o valor absoluto dos autovalores da matriz $A(n \times n)$.

Prova: Seja A uma matriz arbitrária de ordem n e (λ, x) um elemento de seu autosistema. Por definição $Ax = \lambda x$. Além disso, seja $\|A\|$ uma norma matricial qualquer e $\|x\|$ a correspondente norma vetorial compatível. Então

$$\|Ax\| = \|\lambda x\| = |\lambda| \cdot \|x\| \quad (1.4.1)$$

De (1.3.9)

$$\|Ax\| \leq \|A\| \cdot \|x\| \quad (1.4.2)$$

Portanto, temos que

$$|\lambda| \cdot \|x\| \leq \|A\| \cdot \|x\| \quad (1.4.3)$$

Sendo x um autovetor, é não-nulo e portanto $\|x\| > 0$. Assim,

$$|\lambda| \leq \|A\| \quad (1.4.4)$$

para todo autovalor da matriz A.

Antes de provar o teorema de Gerschgorin, vamos apresentar um resultado devido a Hadamard [36].

Teo.1.5 : (Hadamard) Seja uma matriz A(nxn) com

$$|a_{ii}| \geq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \quad i=1,2,\dots,n \quad (1.4.5)$$

então, o $\det(A)$ é diferente de zero.

Prova: Por contradição, vamos supor que $\det(A)=0$. Assim, o sistema linear

$$Ax = 0 \quad (1.4.6)$$

admite um conjunto infinito de soluções não-nulas. Seja x_i a componente de maior valor absoluto de uma qualquer dessas soluções, isto é,

$$|x_i| > |x_j| \quad i \neq j \quad j=1,2,\dots,n$$

Da i-ésima equação de (1.4.6), temos que

$$|a_{ii}| |x_i| \leq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| |x_j|$$

portanto,

$$|a_{ii}| < \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| |x_j| / |x_i| < \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|$$

que contradiz (1.4.5). Assim $\det(A)$ é diferente de zero.

Teo.1.6 : (gerschgorin) O conjunto de autovalores de uma matriz A(nxn) está contido na união dos n círculos de centro a_{ii} e raios

respectivamente

$$r_i = \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \quad i=1,2,\dots,n.$$

Prova: Pelo teorema de Hadamard se

$$|a_{ii} - \lambda| > r_i$$

o $\det(A - \lambda I) = 0$. Como os autovalores da matriz A são as raízes de $\det(A - \lambda I) = 0$, cada autovalor λ_i de A está localizado no interior ou no contorno do círculo de raio r_i . Assim, o conjunto de todos os autovalores λ_i 's da matriz A está contido na união de todos os círculos de raios r_i 's.

1.5. Transformações de Householder e Rotações no Plano

Def.1.5.1 - Uma matriz (transformação) real $A(n \times n)$ tal que $A'A = AA' = I$ é chamada uma matriz ortogonal.

Nesta definição se as colunas (linhas) da matriz A forem tomadas como vetores do R^n , a condição $A'A = I$ ($AA' = I$) diz que as colunas (linhas) de A formam um conjunto ortonormal.

Teo.1.7 : Se $A(n \times n)$ é uma matriz ortogonal, então para todos os vetores x, y do R^n , temos:

(1) - $\|Ax\| = \|x\|$.

(2) - $\|A\| = 1$.

Prova: Vamos tomar o produto

$$(Ax, Ay) = (x, A'Ay)$$

$$= (x, y)$$

$$= (x, y)$$

fazendo $x = y$, temos $\|Ax\|_2 = \|x\|_2$ o que prova a parte (1). Usando a def.1.3.7 temos

$$\|A\|_2 = \max_{\|x\|_2=1} (\|Ax\|_2 / \|x\|_2) = 1.$$

provando assim a parte (2).

Geométricamente, o teo.1.7 afirma que uma transformação (matriz) ortogonal preserva tanto ângulo como comprimento Euclidiano.

Def.1.5.2 - Uma transformação elementar é em geral representada por uma matriz da forma

$$I + \alpha wv' \tag{1.5.1}$$

com α um escalar, w e v vetores do \mathbb{R}^n e wv' um produto diádico.

Para o caso de $w = v$ vetores unitários, ($\|w\|_2 = 1$) e $\alpha = -2$ temos uma transformação (matriz) de Householder ou refletor elementar.

$$H = I - 2ww', \quad \|w\|_2 = 1. \tag{1.5.2}$$

Teo.1.8 : A matriz H é simétrica e ortogonal.

Prova: É imediato provar que H é simétrica. Para mostrar que H é ortogonal basta calcular o produto

$$\begin{aligned}
H'H &= (I-2ww')(I-2ww') \\
&= I-4ww'+4ww'ww' \\
&= I.
\end{aligned}$$

Teo.1.9 : Para quaisquer dois vetores do \mathbb{R}^n , não-nulos de mesmo comprimento Euclidiano, existe uma transformação de Householder que transforma um vetor no outro.

Prova: Seja x e y vetores do \mathbb{R}^n de mesmo comprimento. Tomando H como em (1.5.2) temos

$$\begin{aligned}
Hx &= (I-2ww')x \\
&= x - 2w(w'x) \\
&= y \qquad \qquad \qquad (1.5.3)
\end{aligned}$$

Segundo (1.5.3) o vetor w tem a mesma direção que o vetor $(x-y)$. Note também que o vetor $Hx=y$, é simplesmente a diferença entre o vetor original x e um múltiplo do vetor w . Consequentemente, o vetor y é igual ao vetor x em todos as componentes onde o vetor de Householder w tem componentes nulos. Ainda de (1.5.3), se $w'x=0$ o vetor x será inalterado pela transformação de Householder.

Algumas propriedades da transformação H são facilmente observadas através da figura (1.1). Veja que geometricamente a transformação linear representada por H é uma reflexão em relação ao hiperplano ortogonal ao vetor w .

Geométricamente, a transformação linear R corresponde a uma rotação bidimensional de ângulo θ no plano gerado pelas direções de índices p e q . Podemos notar, também, que a transformação R será equivalente a uma transformação de Householder H , quando o vetor w que caracteriza a matriz H tem elementos não-nulos somente nas posições (p,p) , (p,q) , (q,p) e (q,q) .

A matriz R , pode ser caracterizada pelo terno (p,q,θ) com $1 \leq p < q \leq n$ e θ o ângulo de rotação. Seus elementos são definidos por

$$\begin{cases} r_{ii} = 1 & i \neq p, q \\ r_{pp} = \cos\theta \\ r_{pq} = \sin\theta \\ r_{qp} = -\sin\theta \\ r_{qq} = \cos\theta \\ r_{ij} = 0 & \text{outros} \end{cases} \quad (1.5.4)$$

Note que devido a estrutura da matriz R , quando pré-multiplicamos um dado vetor x do R^n pela matriz R , obtemos um outro vetor que difere do original somente nas p -ésima e q -ésima coordenadas permanecendo as outras inalteradas. Assim, se x é um dado vetor do R^n usualmente escolhemos o ângulo θ , tal que o elemento, por exemplo, na q -ésima posição seja nulo. Isto pode ser obtido escolhendo

$$s = \pm x_q / \eta \quad e \quad c = \pm x_p / \eta \quad (1.5.5)$$

onde $\eta = (x_p^2 + x_q^2)^{1/2}$. Em geral, para efeito de uso prático, usamos somente o sinal positivo.

1.6. Sequência de Sturm

Def. 1.6.1 - Uma sequência de funções reais $(f_0(x), f_1(x), \dots, f_n(x))$ na variável real x , é uma sequência de Sturm se:

S1 - As funções $f_i(x)$, $i=0,1,2,\dots,n$ são contínuas.

S2 - Para $x \in [a,b]$, $f_0(x)$ não troca de sinal.

S3 - Dois termos sucessivos da sequência não se anulam no mesmo ponto.

S4 - Se $f_i(\bar{x}) = 0$, $\bar{x} \in [a,b]$, então,

$$f_{i-1}(\bar{x}) \cdot f_{i+1}(\bar{x}) < 0 \quad i=1,2,\dots,n-1 \quad (1.6.1)$$

S5 - Para $f_n(\bar{x}) = 0$, $\bar{x} \in [a,b]$ e $h > 0$ suficientemente pequeno

$$\text{Sign}(f_n(\bar{x}-h)/f_{n-1}(\bar{x}-h)) = -1 \quad \text{e} \quad \text{Sign}(f_n(\bar{x}+h)/f_{n-1}(\bar{x}+h)) = +1 \quad (1.6.2)$$

A condição S5, significa que em todo zero de $f_n(x)$, o sinal de $f_{n-1}(\bar{x})$ corresponde ao sinal da derivada de $f_n(x)$, em $x = \bar{x}$ se ela existe.

Exemplo 4: A sequência de funções $(f_0(x), f_1(x), f_2(x), f_3(x))$ definidas por

$$f_0(x) = 1 \quad f_1(x) = x - 1 \quad f_2(x) = x^2 - 3x + 1$$

$$f_3(x) = x^3 - 3x^2 + 1$$

é uma sequência de Sturm.

As condições (S3 - S5), podem ser verificadas através da figura (1.2).

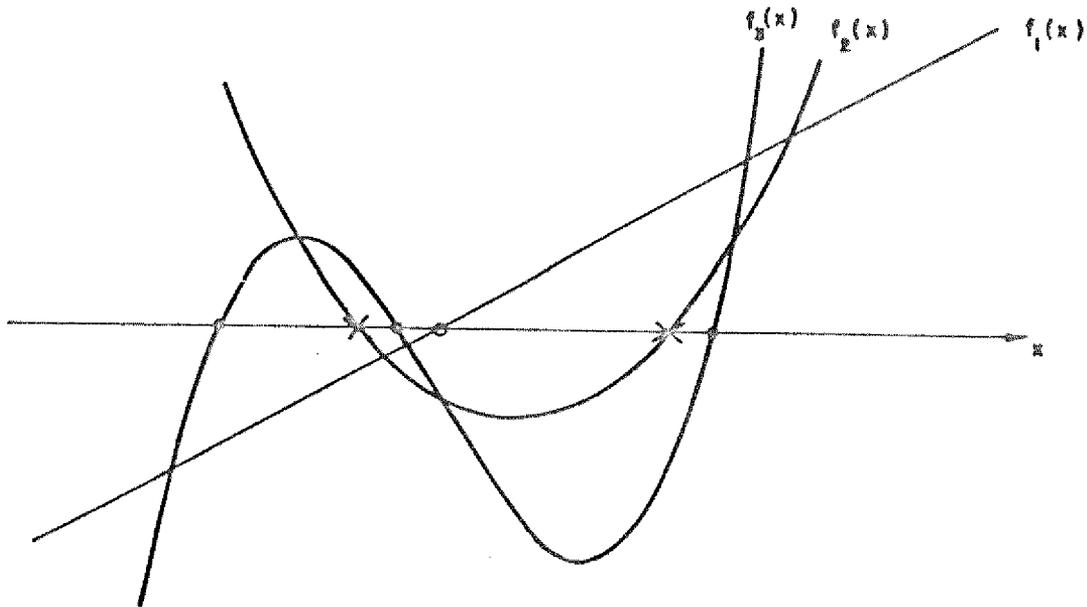


fig. (1.2)

Teo.1.10 : Seja $(f_0(x), f_1(x), \dots, f_n(x))$ uma sequência de Sturm, e $M(x)$ o número de trocas de sinais nos termos da sequência. Então, para um dado valor de x no intervalo $[a, b]$ o número de zeros, m , de $f_n(x)$ neste intervalo é igual a:

$$m = M(a) - M(b) \quad (1.6.3)$$

Prova: Considere o número de trocas de sinais na sequência $(f_0(x), f_1(x), \dots, f_{n-1}(x), f_n(x))$ como uma função de x , quando x cresce monotonicamente de a para b . Pela continuidade das funções na sequência de Sturm, o valor $H(x)$ se modifica somente se uma ou mais funções passa por um zero. Desde que por definição $f_0(x)$ não troca de sinal, dois casos devem ser considerados :

(a) Zeros oriundos somente das funções "internas" da sequência, isto é $(f_i(x), i=1,2,\dots,n-1)$.

Pelas condições S1,S3 e S4, somente as combinações de sinais mostradas na tabela (1.1) serão possíveis para $h > 0$ suficientemente pequeno.

x	$f_{l+1}(x)$	$f_l(x)$	$f_{l-1}(x)$	x	$f(x)$	$f(x)$	$f(x)$
$\bar{x}-h$	+	\pm	-	$\bar{x}-h$	-	\pm	+
\bar{x}	+	0	-	\bar{x}	-	0	+
$\bar{x}+h$	+	\mp	-	$\bar{x}+h$	-	\mp	+

tab.(1.1)

Em qualquer situação, o número de trocas de sinais permanece inalterado. Isto também será verdade se $f_i(x)$ tem um zero múltiplo. Assim, todo zero de uma função "interna" da sequência deixa o número de trocas de sinais invariante.

(b) Zeros de $f_n(x)$.

Pelas condições S1,S3 e S5, somente as combinações de sinais esboçadas na tabela (1.2) serão possíveis para $h > 0$ suficientemente pequeno.

x	$f_n(x)$	$f_{n-1}(x)$	x	$f_n(x)$	$f_{n-1}(x)$
$\bar{x}-h$	+	-	$\bar{x}-h$	-	+
\bar{x}	0	-	\bar{x}	0	+
$\bar{x}+h$	-	-	$\bar{x}+h$	+	+

tab.(1.2)

Em ambos os casos, uma troca de sinal é subtraída com o crescimento de x . Assim, todo zero de $f_n(x)$ reduz o número de trocas de sinais de um.

Resumindo temos que m zeros de $f_n(x)$ no intervalo $[a,b]$ diminui o número de mudanças de sinais $M(x)$ na sequência de Sturm de m . Portanto, m é a diferença entre $M(a)$ e $M(b)$, o que conclui a prova.

Exemplo 5: Usando a mesma sequência do exemplo 4, podemos determinar o intervalo que contém algumas ou todas as raízes da função $f_3(x)$, tab.(1.3).

x	$f(x)$	$f(x)$	$f(x)$	$f(x)$	$M(x)$
-1.	+	-	+	-	3
0.	+	-	+	+	2
.5	+	-	-	+	2
1.	+	0	-	-	1
3	+	+	+	+	0

tab.(1.3)

Para mostrar isto, sejam A_1 e A_2 dois elementos quaisquer de S e t um real, $0 \leq t \leq 1$. Por (1.7.3) tA_1 e $(1-t)A_2$ são elementos de S e por conseguinte $A = tA_1 + (1-t)A_2$, ainda por (1.7.3) pertence a S .

Def.1.8.2 - Um conjunto S é um cone, se para todo x pertencente a S e $t \geq 0$ tx pertence a S . Se o conjunto S também é convexo, dizemos que S é um cone convexo.

Exemplo 7: O conjunto S das matrizes reais simétricas de ordem n negativas semidefinidas é um cone convexo.

A demonstração é trivial. Será mais interessante dar uma visão geométrica deste conjunto.

Para $n=2$, vamos tomar a matriz

$$A = \begin{bmatrix} x_1 & x_3 \\ x_3 & x_2 \end{bmatrix},$$

pelo critério de Sylvester temos:

$$x_1 \leq 0 \quad , \quad x_2 \leq 0$$

$$x_1 x_2 - x_3^2 \geq 0.$$

Esboçando em R^3 o conjunto definido por essas relações, obtemos a fig.1.3. Observe que as matrizes negativas definidas estão localizadas no interior do cone.

De acordo com a tab.(1.3), na última coluna temos os valores de $M(x)$, donde concluímos que para cada intervalo $[-1,0]$, $[.5,1]$ e $[1,3]$ existe respectivamente, um zero de $f_3(x)$. O conjunto de todas os zeros de $f_3(x)$ está contido no intervalo $[-1,3]$.

1.7. Funções Escalares com Sinal Definido

Def.1.7.1 - Uma função escalar, $V : \mathbb{R}^n \rightarrow \mathbb{R}$, definida numa região $S \subseteq \mathbb{R}^n$, contendo a origem, é positiva definida se $V(x) > 0$ para todo x não-nulo em S e $V(0) = 0$.

Def.1.7.2 - Uma função escalar V como da def.1.7.1. é negativa definida se $-V(x) > 0$ para todo x não-nulo e $V(0) = 0$.

Def.1.7.3 - Uma função escalar, $V : \mathbb{R}^n \rightarrow \mathbb{R}$, definida numa região $S \subseteq \mathbb{R}^n$, contendo a origem, é positiva semidefinida se $V(x) \geq 0$ para todo $x \in S$.

Def.1.7.4 - Uma função escalar V como da def.1.7.3 é negativa semidefinida se $-V(x) \geq 0$ para todo $x \in S$.

Se uma função escalar não obedece nenhuma das definições anteriores ela é dita ser indefinida.

Funções escalares com sinal definido, de grande interesse, são as chamadas formas quadráticas. Estas funções escalares são definidas explicitamente por

$$f(x) = x'Ax \quad (1.7.1)$$

com $x \in \mathbb{R}^n$ e A uma matriz real simétrica de ordem n .

No caso particular de formas quadráticas, é muito comum o uso da seguinte notação: $A > 0$, para indicar que a função $f(x)$ é positiva definida. No caso de $f(x)$ ser positiva semidefinida denotamos simplesmente por $A \geq 0$. Esta notação será largamente usada nos capítulos subsequentes.

Em virtude de sua definição, o sinal de uma forma quadrática pode ser determinado através de alguns critérios estabelecidos sobre a sua matriz. Por exemplo :

1. (Critério de Silvester) Se todos os menores principais líderes da matriz A são positivos, a forma quadrática definida por A é positiva definida.

2. (Critério de autovalores) Se todos os autovalores da matriz A que define a forma quadrática são positivos, a forma quadrática definida por A é positiva definida.

1.8. Conjuntos e funções convexas

Def.1.8.1 - Um conjunto S é convexo, se para dois elementos quaisquer, x_1, x_2 de S e um escalar t com $0 \leq t \leq 1$, o elemento $x = tx_1 + (1-t)x_2$ pertence a S .

Exemplo 6: O conjunto S das matrizes reais simétricas de ordem n , positivas semidefinidas ($A \geq 0$) é um conjunto convexo.

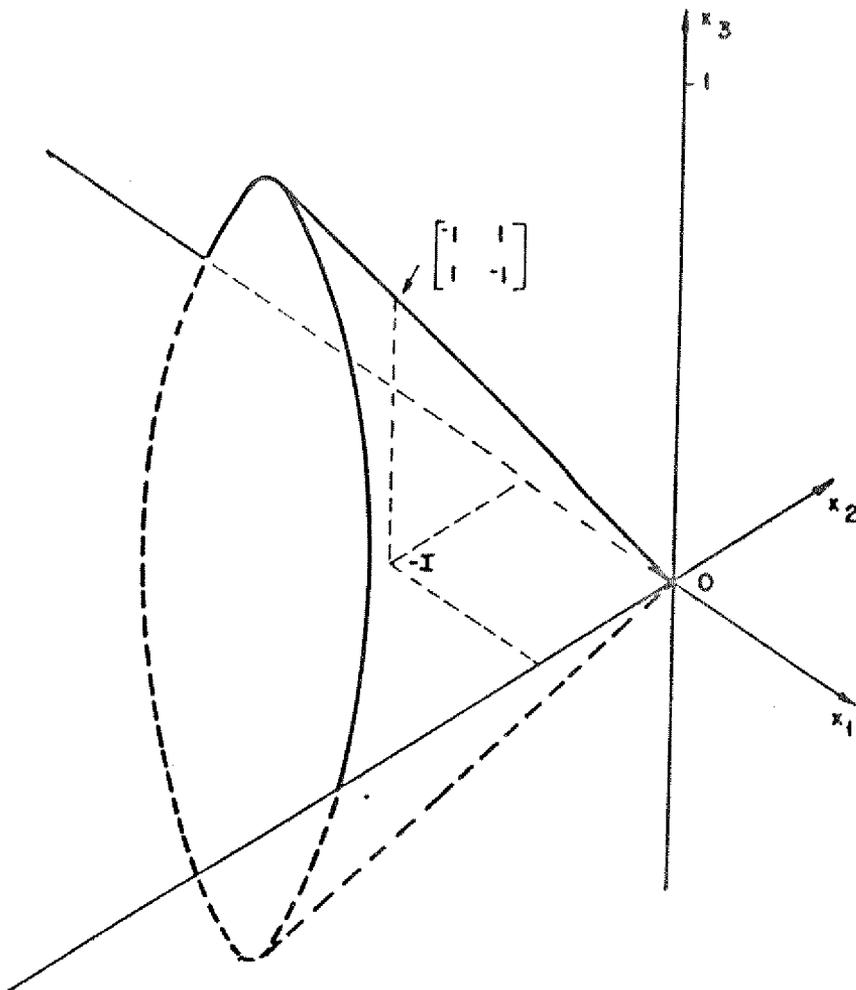


fig.(1.3)

Def.1.8.3 - Um conjunto $V \subseteq \mathbb{R}^n$ é uma variedade linear se, para dois elementos quaisquer x_1, x_2 de V e um escalar real t , o elemento $x = tx_1 + (1-t)x_2$ pertence a V .

Def.1.8.4 - Um hiperplano em \mathbb{R}^n é uma variedade linear de dimensão $(n-1)$.

Exemplo B: Seja a um vetor do \mathbb{R}^n e b um número real. O conjunto $H = \{x \in \mathbb{R}^n : a'x = b\}$ é um hiperplano no \mathbb{R}^n .

Através da linearidade da equação $a'x=b$ segue que H é uma variedade linear. Vamos tomar um ponto (vetor) qualquer x_0 em H . Transladando H de $x-x_0$, obtemos o conjunto H que é um subespaço linear do \mathbb{R}^n . Este subespaço é gerado por todos os vetores do \mathbb{R}^n que satisfazem $a'x = 0$, portanto, todos os vetores ortogonais ao vetor a . Daí, está claro que H tem dimensão $(n-1)$.

Def.1.8.5 - Um semi-espaço fechado positivo (negativo) associado a um hiperplano $H = \{x \in \mathbb{R}^n : a'x = b\}$ é um conjunto convexo dado por :

$$H^+ = \{x \in \mathbb{R}^n : a'x \geq b\} \quad (1.8.1)$$

$$H^- = \{x \in \mathbb{R}^n : a'x \leq b\} \quad (1.8.2)$$

No caso de desigualdade estrita ($>$, $<$) temos semi-espaços abertos positivos (negativos).

Def.1.8.6 - Um conjunto S , que pode ser escrito como a interseção de um número finito de semi-espaços fechados é denominado um polítopo convexo.

Def.1.8.7 - Uma função f definida sobre um conjunto convexo S é convexa, se para quaisquer elementos x_1, x_2 pertencentes a S e um escalar $t, 0 \leq t \leq 1$ a desigualdade

$$f(tx_1 + (1-t)x_2) \leq t f(x_1) + (1-t) f(x_2) \quad (1.8.3)$$

é sempre verdadeira.

Exemplo 9: A função norma $N : R^{n \times n} \rightarrow R$ definida por

$$N(A) = \|A\|$$

é uma função convexa.

Para mostrar isto, vamos tomar dois elementos A_1 e A_2 do $R^{n \times n}$ e $0 \leq t \leq 1$. Seja $A = tA_1 + (1-t)A_2$. Da definição de N temos

$$\begin{aligned} N(A) &= \|tA_1 + (1-t)A_2\| \\ &\leq t \|A_1\| + (1-t) \|A_2\| \\ &= t N(A_1) + (1-t) N(A_2). \end{aligned}$$

Donde se conclui que N é uma função convexa.

1.9. Método Gradientes Conjugados

Métodos gradientes conjugados foram inicialmente propostos, por M.R. Hestenes e E. Stiefel [25], para resolver sistemas de equações lineares, em seguida R. Fletcher e C.M. Reeves [16] apresentaram um algoritmo completo para minimização de funções não-quadráticas, sem restrições. Neste algoritmo, a cada iteração é necessário resolver um sistema linear pelo método de gradientes conjugados. O nome gradientes conjugados, surgiu pelo fato das direções de buscas dos métodos acima

serem direções conjugadas, geradas a partir dos gradientes da função envolvida no processo de minimização.

Def.1.9.1 - Dado uma matriz $A(n \times n)$ real e simétrica, dois vetores, p e q do \mathbb{R}^n , são ditas direções conjugadas (A -ortogonais) com relação à matriz A se

$$p' A q = 0. \quad (1.9.1)$$

Com esta definição, temos que a noção de direções conjugadas é uma generalização de ortogonalidade, em cujo caso a matriz A é igual a matriz identidade I .

Exemplo.10: O conjunto dos n autovetores de uma matriz $A(n \times n)$ real e simétrica é um conjunto de vetores conjugados com relação a A .

Vamos tomar dois autovetores quaisquer x_1 e x_2 da matriz A . Seja λ_1 o autovalor de A associado a x_1 . De (1.2.1) temos

$$A x_1 = \lambda_1 x_1$$

Pré-multiplicando esta igualdade por x_2' temos

$$x_2' A x_1 = \lambda_1 x_2' x_1$$

Como x_1 e x_2 são vetores ortogonais

$$x_2' A x_1 = 0.$$

Assim, x_1 e x_2 são vetores conjugados.

Teo.1.11 : Se $A(n \times n)$ é uma matriz real, simétrica e positiva definida e o conjunto de vetores não-nulos (p_1, p_2, \dots, p_m) é A -ortogonal, então os vetores p_i 's, $i=1,2,\dots,m$, são linearmente independentes.

Prova: Vamos supor que existem m escalares $a_i, i=1,2,\dots,m$ tais que

$$\sum_{i=1}^m a_i p_i = 0. \quad (1.9.2)$$

Pré-multiplicando (1.9.2) pelo vetor $p_i' A$, temos

$$a_i p_i' A p_i = 0. \quad (1.9.3)$$

Desde que A é positiva definida, $p_i' A p_i > 0$. De (1.9.3) temos que $a_i = 0$ $i=1,2,\dots,m$, o que conclui a prova.

Métodos de direções conjugadas, invariavelmente, são projetados e analisados para minimizar(maximizar) funções quadráticas do tipo

$$f(x) = (1/2) x' A x - b' x + c, \quad x \in R^n \quad (1.9.4)$$

onde $A(n \times n)$ é uma matriz real simétrica e positiva definida, b um vetor fixo do R^n e c um escalar.

O gradiente da função quadrática (1.9.4) num ponto x , é o vetor $df(x)/dx = Ax - b$. Um ponto crítico de $f(x)$ é um ponto tal que $df(x)/dx = 0$. Assim, x será um ponto crítico da função f se e somente se x é solução do sistema de equações

$$Ax = b.$$

(1.9.5)

Vamos denotar também o vetor $(b - Ax)$ como sendo o vetor resíduo, r , do sistema (1.9.5) em x . Note que $r = -df(x)/dx$.

Um importante resultado que relaciona direções conjugadas com minimização de funções quadráticas sem restrições é devido a M.J.D. Powell [40].

Teo.1.12 (Powell) : Seja (p_1, p_2, \dots, p_m) um conjunto de vetores não-nulos, mutuamente conjugados com relação a matriz $A(n \times n)$ real, simétrica e positiva definida. O ponto do R^n que minimiza a função quadrática (1.9.4) sobre a variedade linear gerada pelo ponto x_1 e os vetores p_1, p_2, \dots, p_m , é obtido fazendo buscas ao longo de cada uma das direções conjugadas somente uma vez.

Prova: Todo ponto x pertencente a variedade linear, pode ser escrito como

$$x = x_1 + \sum_{i=1}^m a_i p_i \quad (1.9.6)$$

com $a_i, i=1,2,\dots,m$, escalares. O ponto mínimo requerido será então

$$x^* = x_1 + \sum_{i=1}^m a_i^* p_i \quad (1.9.7)$$

onde os $a_i^*, i=1,2,\dots,m$, são escolhidos para minimizar a função

$$\begin{aligned} f(x) &= f\left(x_1 + \sum_{i=1}^m a_i p_i\right) \\ &= f(x_1) + \sum_{i=1}^m [a_i p_i' (Ax_1 - b) + (1/2) a_i^2 p_i' A p_i] \quad (1.9.8) \end{aligned}$$

Pelo teo.1.11 os vetores p_i 's geram um subespaço de dimensão m . Desde que, não existe termos cruzados em a_i , isto é, $a_i \cdot a_j$ para $i \neq j$. Os a_i 's, $i=1,2,\dots,m$, ótimos serão encontrados resolvendo m problemas de minimização do tipo

$$\min_{a_i} \{ f(x_1) + a_i p_i' (Ax_1 - b) + (1/2) a_i^2 p_i' A p_i \} \quad (1.9.9)$$

Mas (1.9.9) é equivalente a resolver

$$\min_{a_i} f(x_1 + a_i p_i) \quad i=1,2,\dots,m \quad (1.9.10)$$

Conseqüentemente, buscando o ponto mínimo nas m direções conjugadas somente uma vez, chegaremos a solução ótima a_i^* , $i=1,2,\dots,m$ e portanto, ao ponto mínimo de $f(x)$ restrito a variedade linear. Com isto, a prova está concluída.

Fazendo $m = n$ no teo.1.12, podemos então encontrar o mínimo da função quadrática (1.9.4) sobre R^n , simplesmente pesquisando ao longo de n direções, não-nulas, mutuamente conjugadas com relação a matriz A .

Quando as direções mutuamente conjugadas, p_i , $i=1,2,\dots,n$, são os vetores negativos dos gradientes ($r = -df(x)/dx$) nos pontos obtidos após cada pesquisa, temos o chamado método dos gradientes conjugados [22], que pode ser enunciado, da seguinte maneira:

Passo inicial. Selecione um ponto x_1 e calcule

$$p_1 = r_1 = b - Ax_1 \quad (1.9.11)$$

Passos iterativos. Tendo obtido x_k, r_k e p_k calcule x_{k+1}, r_{k+1} e p_{k+1} pelas fórmulas

$$a_k = (c_k / d_k), \quad d_k = p_k' A p_k, \quad c_k = p_k' r_k \quad (1.9.12)$$

$$x_{k+1} = x_k + a_k p_k, \quad r_{k+1} = r_k - a_k A p_k \quad (1.9.13)$$

$$b_k = - (p_k' A r_{k+1}) / d_k \quad (1.9.14)$$

$$p_{k+1} = r_{k+1} + b_k p_k \quad (1.9.15)$$

Término. No m -ésimo passo, $m < n$, se $\|r_m\| = 0$, pare. $x^* = x_m$, é o ponto de mínimo da função f .

1.10. Sistemas Dinâmicos: Estabilidade

Sistemas dinâmicos serão focalizados neste trabalho segundo seus aspectos de estabilidade. Para um estudo das propriedades, assim como para efeito de análise segundo o ponto de vista de estabilidade, torna-se necessário algumas noções básicas relativas ao próprio sistema.

Todos os conceitos e definições aqui apresentados serão restritos a classes de sistemas dinâmicos descritos por sistemas de equações a diferenças (sistemas discretos) ou sistemas de equações diferenciais (sistemas contínuos).

Def.1.10.1 - Um sistema dinâmico invariante no tempo será descrito através de uma das equações :

$$x(t+1) = f(x(t)) \quad ; \quad x(0) = x_0 \quad t=0,1,\dots \quad (1.10.1)$$

para o caso de sistemas discretos e

$$\dot{x}(t) = f(x(t)) \quad ; \quad x(0) = x_0 \quad t \geq 0 \quad (1.10.2)$$

para sistemas contínuos.

O vetor $x(.) \in R^n$ é o vetor de estado do sistema e $f(.):R^n \rightarrow R^n$ uma função vetorial contínua que satisfaz as condições de Lipschitz [26], isto é, para quaisquer dois estados x e y pertencentes a região do R^n ,

$$S(x) = (x \in R^n : \|x - x_0\| < \epsilon \text{ e } |t - t_0| < \delta)$$

a condição

$$\|f(x) - f(y)\| \leq \rho \|x-y\| \quad (1.10.3)$$

é satisfeita. A constante ρ depende somente de ϵ e δ .

Um caso particular de interesse é quando $f(.)$ é uma função linear com (1.10.1) e (1.10.2) tomando as seguintes formas

$$x(t+1) = Ax(t) \quad ; \quad x(0) = x_0 \quad t=0,1,\dots \quad (1.10.4)$$

e

$$\dot{x}(t) = Ax(t) \quad ; \quad x(0) = x_0 \quad t \geq 0 \quad (1.10.5)$$

onde A é uma matriz constante de ordem n .

Entre os conceitos que conduzem à análise de estabilidade de sistemas dinâmicos, está o de ponto de equilíbrio.

Def.1.10.2 - Um vetor de estado \bar{x} é um ponto de equilíbrio de um sistema dinâmico, se uma vez que o vetor de estado do sistema sendo igual a \bar{x} , ele permanece igual a \bar{x} indefinidamente.

Para o caso dos sistemas dinâmicos (1.10.1) e (1.10.2) temos:

$$\bar{x} = f(\bar{x}) \quad (1.10.6)$$

e

$$0 = f(\bar{x}) \quad (1.10.7)$$

Note que quando $f(\cdot)$ é linear como em (1.10.4) e (1.10.5), a determinação de \bar{x} será através da solução de um sistema de equações lineares.

Propriedades de estabilidade caracterizam o comportamento do sistema quando ele é inicializado próximo a um dado ponto de equilíbrio. Se o sistema for inicializado num estado exatamente igual ao ponto de equilíbrio, então por definição ele nunca se moverá. Quando inicializado próximo, o estado pode permanecer próximo a ele talvez até tendendo na direção do ponto de equilíbrio com o decorrer do tempo.

Para tornar a definição de estabilidade precisa, necessitamos saber o que significa um estado estar próximo de um ponto de equilíbrio.

Def.1.10.3 - Uma vizinhança $S(\bar{x}, R)$ no espaço de estados é o conjunto de todos os pontos no espaço que satisfazem

$$\|x - \bar{x}\|_2 \leq R \quad (1.10.8)$$

Segundo (1.10.8), $S(\bar{x}, R)$ é uma esfera de centro em \bar{x} e raio igual a R .

Def.1.10.4 - Um ponto de equilíbrio x de (1.10.1), (1.10.2) é estável se existe $R_0 > 0$ para o qual se verifica: para todo $R < R_0$, existe um r , $0 < r < R$, tal que se $x_0 \in S(\bar{x}, r)$ então $x(t)$ está dentro de $S(\bar{x}, R)$ para todo $t > 0$.

Def.1.10.5 - Um ponto de equilíbrio \bar{x} de (1.10.1), (1.10.2) é assintoticamente estável se ele é estável e em adição existe um $R_0 > 0$ tal que se o estado for inicializado em $S(\bar{x}, R)$, ele tende a x quando t cresce.

Def.1.10.6 - Um ponto de equilíbrio x é instável se ele não é estável. Isto é, \bar{x} é instável se para algum $R > 0$ e qualquer $0 < r < R$ existe um ponto em $S(\bar{x}, r)$ tal que se o estado do sistema for inicializado neste ponto, eventualmente x move-se para fora de $S(\bar{x}, R)$.

Com o propósito de analisar a estabilidade de sistemas dinâmicos, o engenheiro russo A. M. Lyapunov publicou dois métodos. O chamado primeiro método ou método indireto de Lyapunov consiste em verificar a estabilidade através de uma versão linearizada do sistema. Esta linearização é baseada numa expansão em série de Taylor da função $f(\cdot)$ em torno de um ponto de equilíbrio do sistema [32].

O segundo método ou método direto de Lyapunov [26], consiste em analisar a estabilidade do sistema a partir de sua equação sem precisar resolvê-la.

Nos últimos anos, este segundo método vem recebendo uma atenção muito especial dos pesquisadores da área de sistemas dinâmicos, pelo fato de ser uma abordagem muito geral, não se restringindo a uma classe particular de sistemas.

Def.1.10.7 - Uma função escalar, $V: S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função de Lyapunov para um sistema dinâmico e um ponto de equilíbrio do sistema pertencente a S , se satisfaz as seguintes condições:

- a - V é uma função contínua
- b - V atinge seu único mínimo em $\bar{x} \in S$
- c - V é sempre decrescente ao longo de qualquer trajetória, do sistema, contida em S .

A primeira condição estabelece que o gráfico da função V , no espaço de estados não possui descontinuidades. A segunda, que V é minimizada em \bar{x} significa que seu gráfico tem seu ponto mais baixo em \bar{x} . A terceira condição é aquela que realmente relaciona a função V com o sistema. Quando o estado evolui no espaço de estados de acordo com as leis do movimento, associamos com cada instante de tempo t o correspondente valor da função $V(x(t))$. O comportamento no tempo da função V é um resumo do comportamento de $x(t)$. Assim, a terceira condição para uma função de Lyapunov é que o valor de V associado com o vetor de estado x , nunca cresça com o tempo quando o sistema evolui.

As figuras (1.4.a) e (1.4.b) sintetizam as noções geométricas de uma função de Lyapunov para um sistema de duas variáveis de estado. A primeira mostra o gráfico da função V , no espaço de estado. A segunda é um traçado das curvas de níveis de V , lugar dos pontos onde V é constante, no espaço de estado.

Note que estes comentários são válidos tanto para sistemas discretos como para sistemas contínuos.

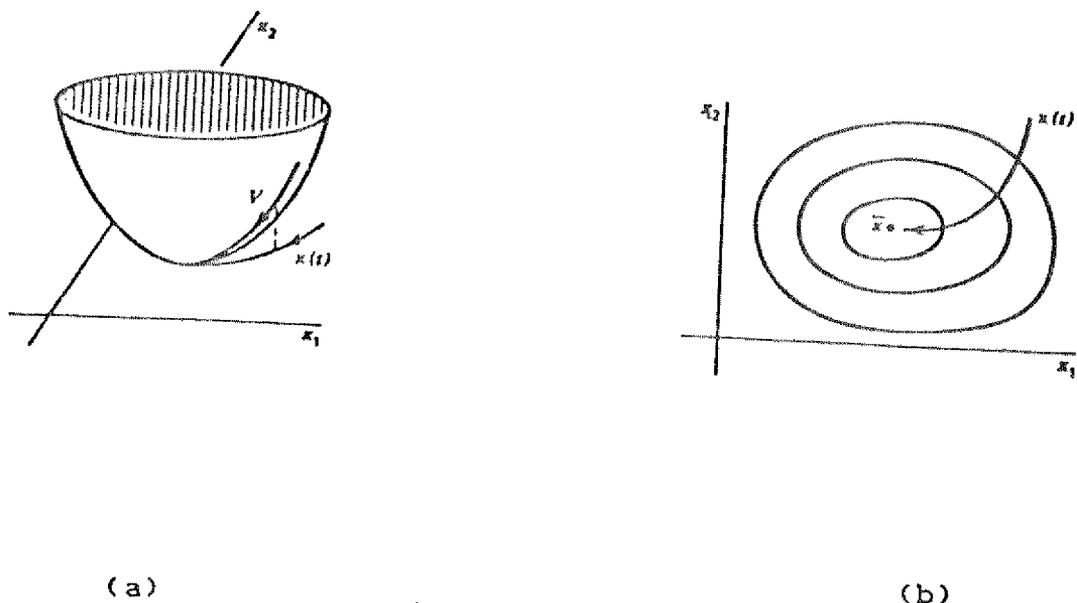


fig. (1.4)

Vamos agora focalizar, somente sistemas discretos no tempo do tipo (1.10.1) junto com um dado ponto de equilíbrio \bar{x} .

A condição (c), de que uma função de Lyapunov nunca cresça ao longo de uma trajetória pode ser traduzida através de uma relação matemática específica.

Se em qualquer instante t , o estado do sistema (1.10.1) é igual a x então, no próximo instante $(t+1)$ o estado será $f(x)$. Os valores da função de Lyapunov nestes pontos são respectivamente $V(x)$ e $V(f(x))$. Portanto, a variação no valor de V será

$$\Delta V(x) = V(f(x)) - V(x), \quad (1.10.9)$$

Se V é uma função de Lyapunov sobre uma região S do espaço de estados, esta variação é não-positiva para todo estado x em S . Em outras palavras, a condição de que a função de Lyapunov não cresça ao longo de uma trajetória se traduz através de:

$$\Delta V(x) = V(f(x)) - V(x) < 0 \quad (1.10.10)$$

para todo $x \in S$.

Def.1.10.8 - Uma função V , definida numa região S , do espaço de estados de $x(t+1)=f(x(t))$ e contendo o ponto de equilíbrio \bar{x} , é uma função de Lyapunov do sistema discreto no tempo, se ela satisfaz as seguintes condições :

- (1) - V é uma função contínua
- (2) - V atinge seu único mínimo em $\bar{x} \in S$
- (3) - A variação no valor da função $\Delta V(x) = V(f(x)) - V(x)$ satisfaz $\Delta V(x) < 0$ para todo $x \in S$.

Teo.1.13 : (Discreto no tempo) Se existe uma função de Lyapunov V numa região esférica $S(\bar{x}, R_0)$ com centro em \bar{x} , então o ponto de equilíbrio é estável. Se, além disso, a função ΔV é estritamente negativa ($\Delta V(x) < 0$) em todo ponto (exceto \bar{x}), então, a estabilidade é assintótica.

Prova: A prova será baseada em relações geométricas ilustradas na fig.1.5. Suponha que $V(x)$ é definida na região esférica $S(\bar{x}, R_0)$. Seja R arbitrário com $0 < R < R_0$. Seja $R_1 < R$, selecionado, tal que se $x \in S(x, R_1)$ então $f(x) \in S(x, R_0)$. Um tal R_1 existe porque f é contínua. Com esta

escolha, se o vetor de estado pertence a $S(\bar{x}, R_1)$, ele não passa da esfera maior $S(x, R_0)$ em um passo.

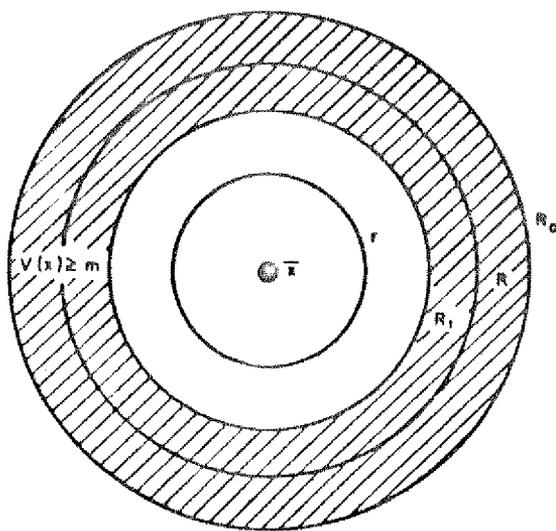


fig. (1.5)

Agora seja m o valor mínimo de $V(x)$ sobre a região definida por $R_1 \leq \|x - \bar{x}\| \leq R_0$. Este valor mínimo existe porque V é por hipótese contínua. Também temos $m > V(\bar{x})$ já que V tem um único mínimo em \bar{x} .

Sendo V contínua, é possível selecionar um $0 < r < R_1$ tal que para $x \in S(\bar{x}, r)$ temos $V(x) < m$, porque próximo a \bar{x} a função V deve tomar valores próximo a $V(\bar{x})$.

Agora suponha que $x(0)$ é tomado como um ponto arbitrário em $S(\bar{x}, r)$, então $V(x(0)) < m$. Desde que, $\Delta V(x) < 0$ o valor $V(x)$ não pode crescer com o tempo. Portanto, a trajetória nunca pode sair de $S(\bar{x}, R_1)$ e conseqüentemente ela nunca pode sair de $S(\bar{x}, R)$. Assim, para este arbitrário $R > 0$ encontramos $r > 0$, correspondente às condições de definição de estabilidade.

Se em adição $\Delta V(x) < 0$ para todo ponto exceto \bar{x} , então, V deve realmente decrescer continuamente (ou para todo t , ou enquanto o estado

alcança \bar{x} , se isto ocorrer num tempo finito). Assim, $V(x)$ converge para algum valor limite \bar{m} . A única questão que resta é saber se é possível para $\bar{m} > V(x)$. Isto não é possível pois, desde que $V(x(t))$ converge para \bar{m} , deve ser verdade que $\Delta V(x(t))$ converge para zero. Mas $\Delta V(x)$ é estritamente negativa exceto para \bar{x} . Assim, desde que $\Delta V(x)$ é contínua, $x(t)$ deve convergir para \bar{x} e $V(x(t))$ deve convergir para $V(\bar{x})$. Isto caracteriza estabilidade assintótica, o que prova o teorema.

Vamos considerar agora, sistemas contínuos no tempo (1.10.2) junto com um dado ponto de equilíbrio \bar{x} , assumindo ainda que f é contínua. No caso contínuo no tempo, o requisito de que o valor da função de Lyapunov nunca cresça ao longo de uma trajetória, é expresso em termos da derivada no tempo. Suponha que $x(t)$ é uma trajetória. Então $V(x(t))$ representa o correspondente valor de V ao longo da trajetória. Para que V não cresça, requeremos que $\dot{V}(x(t)) < 0$ para todo t . Usando a regra da cadeia para diferenciação temos

$$\begin{aligned} \dot{V}(x(t)) = & (\partial V / \partial x_1) \cdot \dot{x}_1(t) + (\partial V / \partial x_2) \cdot \dot{x}_2(t) + \\ & \dots + (\partial V / \partial x_n) \cdot \dot{x}_n(t). \end{aligned} \quad (1.10.11)$$

Usando a equação do sistema original, (1.10.11) torna-se

$$\begin{aligned} \dot{V}(x(t)) = & (\partial V / \partial x_1) \cdot f_1(x(t)) + (\partial V / \partial x_2) \cdot f_2(x(t)) + \\ & \dots + (\partial V / \partial x_n) \cdot f_n(x(t)). \end{aligned} \quad (1.10.12)$$

que pode ser escrito como

$$\dot{V}(x(t)) = \nabla V(x(t))'f(x(t)) \quad (1.10.13)$$

sendo $\nabla V(x(t))$ o vetor gradiente da função V .

Portanto, a condição de que V não cresça ao longo de qualquer trajetória do sistema, se traduz através de que $\dot{V}(x) = \nabla V(x)'f(x) < 0$ para todo $x \in S$.

Def.1.10.9 - Uma função V definida numa região S do espaço de estados de $\dot{x}(t)=f(x(t))$ e contendo o ponto de equilíbrio \bar{x} é uma função de Lyapunov do sistema contínuo no tempo, se ela satisfaz as seguintes condições:

- (1) V é uma função contínua e tem primeiras derivadas parciais contínuas.
- (2) V atinge seu único mínimo em $\bar{x} \in S$.
- (3) A função $\dot{V}(x) = \nabla V(x)'f(x)$ satisfaz $\dot{V}(x) < 0$ para todo $x \in S$.

Teo.1.14 : (Contínuo no tempo) Se existe uma função de Lyapunov $V(x)$ numa região esférica $S(\bar{x}, R_0)$ com centro em \bar{x} , então, o ponto de equilíbrio \bar{x} é estável. Se, além disso, a função $\dot{V}(x)$ é estritamente negativa em todo ponto (exceto \bar{x}), então a estabilidade é assintótica.

Prova: Similar a do teo.1.13.

Exemplo.11: Considere o sistema autônomo linear contínuo no tempo

$$\dot{x}(t) = Ax(t) \quad (1.10.14)$$

Vamos associar a este sistema e ao ponto de equilíbrio $\bar{x} = 0$ a forma quadrática

$$V(x) = x'Px \quad (1.10.15)$$

onde P é simétrica positiva definida. Esta função V é contínua e tem primeiras derivadas parciais contínuas. Além disto, desde que P é positiva definida, a origem é o único ponto mínimo de V . Assim, V em termos gerais é uma candidata em potencial para uma função de Lyapunov. Resta, portanto, determinar como $V(x)$ é influenciada pela dinâmica do sistema.

De (1.10.15)

$$\begin{aligned} \dot{V}(x) &= \dot{x}'Px + x'P\dot{x} \\ &= x'A'Px + x'PAx \\ &= x'(A'P + PA)x. \end{aligned}$$

Definindo

$$A'P + PA = -Q \quad (1.10.16)$$

temos

$$\dot{V}(x) = -x'Qx.$$

Sendo V uma função quadrática, a função V será uma função de Lyapunov se a matriz Q for positiva semidefinida. No caso da matriz Q ser positiva definida, o sistema (1.10.14) será assintoticamente estável.

Exemplo.12 : Para o caso de sistemas discretos

$$x(t+1) = Ax(t)$$

a equação matricial de Lyapunov apresenta a seguinte forma:

$$A'PA - P = -Q \quad (1.10.17)$$

com Q e P matrizes reais, simétricas e positiva definidas.

Def.1.10.10 - Uma matriz real $A(n \times n)$, é estável se o sistema (1.10.4) ((1.10.5)) é assintoticamente estável.

Através do teo.(1.8) ((1.9)) e da def. 1.10.10, temos que: uma matriz real A é estável se e somente se a eq.(1.10.17) ((1.10.16)) admite uma solução P positiva definida para qualquer matriz Q positiva definida.

Def.1.10.11 - Uma matriz $A(n \times n)$ real é diagonal estável se DA é uma matriz estável para qualquer matriz $D = (d_i, i=1,2,\dots,n)$ com $d_i > 0$.

Teo.1.15 : Se existe uma matriz diagonal P , tal que $A'P + PA$ é negativa definida, então A é diagonal estável.

Prova : Se $A'P + PA = -Q$ com $Q > 0$, então

$$(AD)'PD + DP(AD) = -DQD \quad (1.10.18)$$

para qualquer matriz diagonal positiva D . Desde que, PD e DQD são ambas positivas definidas, AD é uma matriz estável.(Ex.11).

O mesmo enunciado pode ser dado para o caso da eq.(1.10.17).

Capítulo 2

Problema Completo e Especial de Autovalores de uma Matriz Real e Simétrica

2.1. Introdução

O problema de calcular o autosistema de uma matriz, para efeito didático, divide-se em duas categorias. A primeira engloba os problemas de achar todos os elementos do conjunto $((\lambda_i, x_i), i=1,2,\dots,n)$ onde λ_i são os autovalores da matriz e x_i os autovetores associados como na def.1.2.2, sendo este o chamado problema completo de autovalores e autovetores. Na segunda categoria, estão os chamados problemas especiais de autovalores onde um ou alguns, mas não todos os elementos (λ_i, x_i) do autosistema da matriz devem ser calculados.

Indubitavelmente, qualquer problema da segunda categoria pode ser resolvido pelos métodos propostos para resolver aqueles da primeira. Ocorre que os métodos existentes para resolver o problema completo, em geral, exigem um esforço computacional muito grande. Daí, o empenho no sentido de elaborar métodos numéricos específicos para resolver problemas especiais.

Neste capítulo, vamos abordar problemas das duas categorias, propondo algoritmos numéricos para calcular o autosistema de uma matriz simétrica, assim como, resolver um tipo particular de problemas especiais, que consiste em calcular o menor ou maior autovalor e seu respectivo autovetor.

O primeiro algoritmo a ser apresentado, determina o autosistema de uma matriz real simétrica $A(n \times n)$ de uma forma muito peculiar. No cálculo individualizado de cada elemento (λ_i, x_i) do autosistema, está a razão principal pela qual este algoritmo torna-se um ferramenta muito útil na determinação e análise da estrutura de autosistemas, tão necessária na análise de convergência dos processos de otimização.

A abordagem efetuada por este algoritmo consiste em decompor a matriz $A(n \times n)$ num produto de outras matrizes. Esta decomposição é feita através de transformações ortogonais (Householder), sendo que, a cada nova iteração, obtemos uma matriz similar à matriz anterior. Este processo de decomposição continua até se obter uma matriz na forma tridiagonal. A partir deste ponto, passamos a trabalhar com a matriz tridiagonal, calculando seus autovalores e respectivos autovetores. Note que os autovalores da matriz tridiagonal são os mesmos da matriz original, não ocorrendo o mesmo para os autovetores.

A determinação dos autovalores é feita através de uma sequência de polinômios, cujo termo de maior grau é o polinômio característico da matriz tridiagonal. Mostramos que esta sequência polinomial, é uma sequência de Sturm. Em seguida, obtendo um intervalo que contenha todas as raízes do polinômio característico (autovalores), é possível através de um processo de encontrar raízes de um polinômio, obter os autovalores. O processo usado é conhecido como método da bissecção intervalar repetida.

Para obter todos os autovetores da matriz tridiagonal, propomos resolver n sistemas lineares tridiagonais, usando rotações planas.

Em seguida, conhecendo os vetores que definem as transformações ortogonais do processo de tridiagonalização e os autovetores das matrizes tridiagonais, completamos a última etapa, deste algoritmo,

deduzindo uma relação de recorrência, para calcular os autovetores da matriz original A.

Os outros algoritmos propostos abordam especificamente um problema especial. Esta abordagem é feita através do uso de técnicas de otimização que exploram as propriedades da função quociente de Rayleigh.

A utilização de direções conjugadas (gradientes conjugados), com objetivo de minimizar (maximizar) a função quociente de Rayleigh permite, que esses algoritmos sejam aplicados a problemas cuja dimensão da matriz seja de valor considerável: os chamados problemas de grande porte. Como em geral, as matrizes dos problemas de grande porte apresentam estruturas esparsas, será perfeitamente viável, através destes algoritmos explorar esta esparsidade.

2.2. Cálculo do Autosistema de uma Matriz Simétrica.

2.2.1 Algoritmo

Seja uma dada matriz $A(n \times n)$ simétrica. Vamos considerar o problema de calcular todos os pares (λ_i, x_i) , $i=1,2,\dots,n$ com x_i não-nulo que satisfazem a equação

$$Ax_i = \lambda_i x_i \quad (2.2.1)$$

Para resolver este problema, propomos um algoritmo que consiste basicamente de três etapas:

E1 - Usando transformações elementares ortogonais de Householder, obter uma matriz simétrica tridiagonal T, a partir da matriz original A.

E2 - Calcular os autovalores da matriz tridiagonal T - que por similaridade são os mesmos da matriz A - usando o método da bissecção intervalar repetida.

E3 - Determinar os autovetores da matriz A , associados aos seus respectivos autovalores, a partir dos autovetores da matriz tridiagonal T .

Este algoritmo difere de um inicialmente proposto por Givens-Householder e posteriormente completado por Wilkinson [53,54]. Aqui, o intervalo inicial contendo todos os autovalores é determinado usando o teo.1.6 de Gerschgorin. Os autovetores em E3, propomos que sejam calculados a partir da solução de n sistemas lineares simétricos tridiagonais usando rotações planas.

2.2.2 Redução a Forma Tridiagonal

Entre os métodos disponíveis para obter uma matriz simétrica tridiagonal T a partir de uma matriz simétrica $A(n \times n)$, o método proposto por A.S.Householder [50] mostra ser o mais eficiente de todos, apesar da sua aparente complexidade.

Com o mesmo propósito J.W.Givens [23], inspirado na idéia inicial de J.C.Jacobi, sugeriu o uso de rotações planas para anular elemento por elemento da matriz A , fora das posições tridiagonais. Mesmo sendo uma idéia muito simples, no computo geral o número de operações realizadas pelo método de Givens, é maior que o realizado pelas transformações de Householder [48], que consiste em anular simultaneamente todos os elementos de uma linha (colunas) fora das posições tridiagonais em apenas um passo.

A partir de (2.2.3), obtemos a matriz na forma tridiagonal quando $k=(n-2)$.

Para caracterizar uma iteração, vamos obter A_1 que é típica. Suponha que queremos anular os elementos a_{1k} , $k=3,4,\dots,n$, e por simetria a_{k1} . Escolhemos o vetor w com a primeira componente nula, ou seja $w'=[0, w_2, w_3, \dots, w_n]$. Isto porque desejamos que a matriz A_1 tenha elementos nulos fora das posições tridiagonais de sua primeira linha (coluna). Assim, a matriz que representa a transformação de Householder será:

$$H = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 - 2w_2^2 & -2w_2 w_3 & \dots & -2w_2 w_n \\ 0 & -2w_2 w_3 & 1 - 2w_3^2 & \dots & -2w_3 w_n \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & -2w_2 w_n & -2w_3 w_n & \dots & 1 - 2w_n^2 \end{bmatrix} \quad (2.2.4)$$

Para determinar as componentes não-nulas do vetor w , vamos efetuar o produto $H_1' A H_1 = H_1 A H_1$ e representar o elemento genérico deste produto por \tilde{a}_{ij} . De imediato verificamos que a primeira linha do produto $H_1 A$ permanece inalterada e os elementos da primeira linha da matriz $A_1 = H_1 A H_1$ são:

$$\begin{aligned} \tilde{a}_{11} &= a_{11} \\ \tilde{a}_{12} &= (1 - 2w_2^2) a_{12} - 2w_2 w_3 a_{13} - \dots - 2w_2 w_n a_{1n} \\ \tilde{a}_{13} &= -2w_2 w_3 a_{12} + (1 - 2w_3^2) a_{13} - \dots - 2w_3 w_n a_{1n} \end{aligned} \quad (2.2.5)$$

$$\tilde{g}_{1n} = -2w_2 w_n a_{12} - 2w_3 w_n a_{13} - \dots + (1-2w_n^2) a_{1n}$$

Como o vetor w tem a primeira componente nula, pelo teo.1.7. temos que

$$\sum_{j=2}^n \tilde{g}_{1j}^2 = \sum_{j=2}^n a_{1j}^2 = s^2 \quad (2.2.6)$$

Definindo

$$f \doteq w_2 a_{12} + w_3 a_{13} + \dots + w_n a_{1n} \quad (2.2.7)$$

Temos de (2.2.5) que

$$\tilde{g}_{1j} = a_{1j} - 2w_j f, \quad j=2,3,\dots,n. \quad (2.2.8)$$

De (2.2.6) e $\tilde{g}_{1j} = 0$ para $j=3,4,\dots,n$, temos

$$\tilde{g}_{12}^2 = s^2 \quad (2.2.9)$$

Portanto, de (2.2.8) segue que

$$a_{12} - 2w_2 f = \pm s. \quad (2.2.10)$$

Como o objetivo é $\tilde{g}_{1j} = 0$, $j=3,4,\dots,n$, ainda por (2.2.8)

$$a_{1j} - 2w_j f = 0 \quad j=3,4,\dots,n. \quad (2.2.11)$$

Multiplicando (2.2.10) por w_2 , cada uma das equações de (2.2.11) pelo correspondente w_j , adicionando as equações resultantes e levando em conta que $\|w\|=1$, temos que

$$f = \bar{f} w_2 s \quad (2.2.12)$$

De (2.2.10) e (2.2.12) podemos calcular o valor de w_2 e com isto (2.2.10) permite calcular o valor de f . Usando as equações (2.2.11) temos w_j , $j=3,4,\dots,n$.

Resumindo

$$w_2 = (.5[1. \bar{f}(a_{12}/s)])^{1/2} ;$$

$$f = \bar{f} w_2 s ;$$

$$w_j = .5(a_{1j}/f) \quad j=3,4,\dots,n \quad (2.2.13)$$

Note que existem duas determinações para o vetor w , dependendo do sinal escolhido para s . Adotaremos para s , sempre o mesmo sinal de a_{12} . Note também que através das relações (2.2.13) podemos determinar o vetor w e conseqüentemente a matriz H_1 .

Resta agora completar a transformação de similaridade da matriz A , usando H_1 . A matriz produto $A_1 = H_1' A H_1$ pode ser determinada sem encontrar a matriz H_1 explicitamente.

De (2.2.4) o elemento genérico da matriz $H(h_{ij})$ pode ser escrito como

$$h_{ij} = \delta_{ij} - 2w_i w_j \quad i, j=1,2,\dots,n \quad (2.2.14)$$

onde δ_{ij} é o delta de Kronecker.

Os elementos da matriz $H_1' A = H_1 A$ são:

$$\begin{aligned} \bar{a}_{ij} &= \sum_{k=1}^n h_{ik} a_{kj} \\ &= \sum_{k=1}^n (\delta_{ik} - 2w_i w_k) a_{kj} \\ &= a_{ij} - 2w_i \sum_{k=1}^n w_k a_{kj}. \end{aligned} \quad (2.2.15)$$

Vamos definir

$$\begin{aligned} p_j &= \sum_{k=1}^n w_k a_{kj} \\ &= \sum_{k=2}^n w_k a_{kj} \quad j=1, 2, \dots, n \end{aligned} \quad (2.2.16)$$

Note que de (2.2.7) e (2.2.16) temos $f = p_1$, visto que $w_1 = 0$.

Assim, (2.2.15) pode ser escrito como:

$$\bar{a}_{ij} = a_{ij} - 2w_i p_j, \quad i, j=1, 2, \dots, n \quad (2.2.17)$$

Os elementos \bar{a}_{ij} da matriz $A_1 = H_1' A H_1$ são:

$$\begin{aligned} \bar{a}_{ij} &= \sum_{k=1}^n a_{ik} h_{kj} \\ &= \sum_{k=1}^n (a_{ik} - 2w_i p_j) (\delta_{kj} - 2w_k w_j) \\ &= a_{ij} - 2w_j \sum_{k=1}^n a_{ik} w_k - 2w_i p_j + 4w_i w_j \sum_{k=1}^n p_k w_k. \end{aligned}$$

Por simetria e da eq. (2.2.16)

$$\sum_{k=1}^n w_k a_{ki} = \sum_{k=1}^n a_{lk} w_k = p_i.$$

Definindo o escalar g independente de i, j como:

$$\begin{aligned} g &= \sum_{k=1}^n p_k w_k \\ &= \sum_{k=2}^n p_k w_k \end{aligned} \quad (2.2.18)$$

podemos escrever

$$\begin{aligned} \tilde{a}_{ij} &= a_{ij} - 2w_j p_i - 2w_i p_j + 4w_i w_j g \\ &= a_{ij} - 2w_i (p_j - gw_j) - 2w_j (p_i - gw_i) \\ &= a_{ij} - w_i q_j - w_j q_i \quad i, j=1, 2, \dots, n \end{aligned} \quad (2.2.19)$$

onde

$$q_j = 2(p_j - gw_j), \quad j=1, 2, \dots, n$$

Resumindo, temos que os elementos da matriz $A_1(\tilde{a}_{ij})$ podem ser calculados por

$$p_j = \sum_{k=2}^n w_k a_{kj};$$

$$g = \sum_{k=2}^n p_k w_k;$$

$$q_j = 2(p_j - gw_j);$$

$$a_{ij} = a_{ij} - w_i q_j - w_j q_i \quad i, j = 1, 2, \dots, n. \quad (2.2.20)$$

Com as relações (2.2.13) e (2.2.20) obtemos a matriz A na forma desejada:

$$\begin{bmatrix} x & * & 0 & 0 & \dots & 0 \\ x & * & * & * & \dots & * \\ 0 & * & * & * & \dots & * \\ 0 & * & * & * & \dots & * \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & * & * & * & \dots & * \end{bmatrix}.$$

Após completar a primeira iteração, realizamos o mesmo procedimento com a submatriz de ordem $(n-1)$ formada pelas linhas e colunas $2, 3, \dots, n$ da matriz A_1 . O mesmo processo é realizado com as submatrizes de ordem $(n-2), (n-3), \dots, 2$. Na última iteração atingiremos uma matriz na forma tridiagonal.

Note que, para a k -ésima iteração o vetor de Householder w , tem suas primeiras k componentes iguais a zero.

2.2.3 Cálculo dos Autovalores das Matrizes T e A .

Após a redução da matriz A à forma tridiagonal, o próximo passo é calcular os autovalores da matriz tridiagonal T . Como T e A são matrizes similares, elas possuem os mesmos autovalores. Assim, nenhum esforço adicional será necessário para determinar os autovalores da matriz A .

2.2.3.a - Menores Principais da Matriz $(\lambda I - T)$

Seja T uma matriz simétrica tridiagonal. Vamos denotar os elementos da diagonal principal de T por a_i , $i=1,2,\dots,n$ e os subdiagonais por b_i , $i=1,2,\dots,n-1$.

$$T = \begin{bmatrix} a_1 & b_1 & & & & \\ b_1 & a_2 & b_2 & & & \\ & & \cdot & \cdot & \cdot & \\ & & & b_{n-2} & a_{n-1} & b_{n-1} \\ & & & & b_{n-1} & a_n \end{bmatrix} \quad (2.2.21)$$

Vamos supor $b_i \neq 0$, $i=1,2,\dots,n-1$. Os menores principais de ordem 1 até n da matriz $(\lambda I - T)$ são:

$$\begin{aligned} p_1(\lambda) &= (\lambda - a_1) \\ p_2(\lambda) &= (\lambda - a_2)(\lambda - a_1) - b_1^2 \\ &= (\lambda - a_2) p_1(\lambda) - b_1^2 \\ p_3(\lambda) &= (\lambda - a_3) p_2(\lambda) - b_2^2 p_1(\lambda) \\ &\vdots \\ p_n(\lambda) &= (\lambda - a_n) p_{n-1}(\lambda) - b_{n-1}^2 p_{n-2}(\lambda) \end{aligned} \quad (2.2.22)$$

Definindo $p_0(\lambda) = 1$, temos a sequência $(p_0(\lambda), p_1(\lambda), \dots, p_n(\lambda))$ com termo geral

$$p_i(\lambda) = (\lambda - a_i) p_{i-1}(\lambda) - b_{i-1}^2 p_{i-2}(\lambda) \quad i=1,2,\dots,n \quad (2.2.23)$$

Note que $p_n(\lambda)$ é o polinômio característico da matriz T .

A sequência de polinômios $(p_i(\lambda), i=0,1,\dots,n)$ gerada pelos menores principais, conduz à elaboração de um método para calcular os autovalores da matriz T.

Antes de apresentar o método, precisamos demonstrar alguns resultados sobre os elementos desta sequência polinomial.

P1. - $p_i(\lambda), i=0,1, \dots, n$ são funções contínuas.

Prova : Pela definição de menores principais, $p_i(\lambda)$ são reais e como polinômios são funções contínuas.

P2. - $p_i(\lambda)$, não troca de sinal.

P3. - Se $b_i \neq 0, i=1,2,\dots,n-1$, as raízes dos $p_i(\lambda)$ são reais e únicas. Para $1 \leq i \leq n-1$, as raízes de $p_i(\lambda)$ separam as raízes de $p_{i+1}(\lambda)$.

Prova : A primeira parte decorre do fato dos $p_i(\lambda)$'s serem polinômios característicos de submatrizes simétricas de T. A demonstração da segunda parte - usando o princípio da indução finita no índice i - se encontra em [48].

P4 - Se $p_i(\bar{\lambda}) = 0$ então $p_{i-1}(\bar{\lambda}) \cdot p_{i+1}(\bar{\lambda}) < 0, i=0,1,\dots,n-1$.

Prova : Pelo princípio da indução finita:

1) Para $i=1$, se $p_1(\bar{\lambda}) = 0$ então $p_0(\bar{\lambda}) \cdot p_2(\bar{\lambda}) = p_2(\bar{\lambda}) = -b_1^2 < 0$.

2) Vamos supor verdadeiro para $i=k-1$, isto é,

se $p_{k-1}(\bar{\lambda}) = 0$ então $p_{k-2}(\bar{\lambda}) \cdot p_k(\bar{\lambda}) < 0$.

3) Para $i=k$, se $p_k(\bar{\lambda}) = 0$ segue de P3 que $p_{k-1}(\bar{\lambda}) \neq 0$, então

$$\begin{aligned} p_{k-1}(\bar{\lambda}) \cdot p_{k+1}(\bar{\lambda}) &= p_{k-1}(\bar{\lambda}) [(\bar{\lambda} - a_{k+1}) p_k(\bar{\lambda}) - b_k^2 p_{k-1}(\bar{\lambda})] \\ &= - (b_k^2 p_{k-1}(\bar{\lambda}))^2 < 0. \end{aligned}$$

P5. - $p_i(\lambda) > 0$, quando $\lambda \rightarrow +\infty$ $i=1, 2, \dots, n$

Prova : consequência direta de (2.2.22)

Com estas cinco propriedades, verificamos que a sequência $(p_0(\lambda), p_1(\lambda), \dots, p_n(\lambda))$ é uma sequência de Sturm (def.1.6.1.).

Corolário 1. - O número m de raízes do polinômio característico da matriz T , $p_n(\lambda)$, que são maiores que um dado valor $\lambda = \bar{\lambda}$, é igual ao número de trocas de sinais $A(\bar{\lambda})$ da sequência polinomial $(p_i(\lambda), i=0, 1, \dots, n)$.

Prova : Suponhamos que, as raízes do polinômio característico -autovalores da matriz T - estejam ordenados da seguinte maneira:

$$\lambda_1 > \lambda_2 > \dots > \lambda_n \quad (2.2.24)$$

Para qualquer valor b maior que λ_1 , $p_i(b) > 0$, $i=0, 1, \dots, n-1$. Assim, $A(b) = 0$ e portanto $m = A(\bar{\lambda}) - A(b) = A(\bar{\lambda})$ e a prova está concluída.

Note que este resultado permite uma grande flexibilidade quanto a determinação dos autovalores da matriz T . Podemos determinar um, alguns ou todos os autovalores de T dentro de um intervalo pré-fixado.

2.2.3.b - Uma iteração do Método da Bissecção

Vamos agora caracterizar uma iteração do método da bissecção. Suponhamos que as raízes do polinômio característico $p_n(\lambda)$ estejam ordenadas como em (2.2.24). A desigualdade estrita é garantida pela hipótese $b_i \neq 0$, $i=1,2,\dots,n-1$. Como o método localiza uma raiz (autovalor) independente das outras, vamos supor que estamos interessado na k -ésima raiz λ_k do polinômio $p_n(\lambda)$. Além disso, supomos também que λ_k pertence ao intervalo $[a,b]$ contendo todas as raízes. Pelo corolário 1, $A(a) = n$ e $A(b) = 0$. Vamos dividir o intervalo ao meio e calcular o número de trocas de sinais $A(\bar{\lambda})$ neste ponto médio, $\bar{\lambda} = (a + b)/2$.

Duas hipóteses podem ser feitas:

1) Se $A(\bar{\lambda}) = m \geq k$, então $\lambda_k \in [\bar{\lambda}, b]$ e $\bar{\lambda}$ representa um limite inferior para λ_k .

2) Se $A(\bar{\lambda}) = m < k$, então $\lambda_k \in [a, \bar{\lambda}]$ e $\bar{\lambda}$ representa um limite superior para λ_k .

Em seguida $\bar{\lambda}$ substitui a ou b de acordo com o caso 1 ou 2, e o processo é repetido.

De um modo geral, no j -ésimo passo temos:

Se $A(\lambda_j) \geq k$ com $\lambda_j = (a_j + b_j)/2$ fazemos $a_{j+1} = (a_j + b_j)/2$ e $b_{j+1} = b_j$

Se $A(\lambda_j) < k$ com $\lambda_j = (a_j + b_j)/2$ fazemos $a_{j+1} = a_j$ e $b_{j+1} = (a_j + b_j)/2$

e o processo continua até que o tamanho do intervalo $(b_j - a_j)$ satisfaça uma tolerância pré-estabelecida.

Após p repetições do processo, o intervalo $[a, b]$ se reduz a outro de tamanho $(b - a)/2^p$.

2.2.3.c - Intervalo Inicial

Sobre a determinação do intervalo inicial $[a, b]$ que supomos conhecido a priori, podemos usar um dos teoremas 1.4 ou 1.6.

Para o caso de usar o teo.1.4, qualquer norma (1.3.5)-(1.3.7) pode ser escolhida. Por exemplo, sugerimos $\|T\|_\infty$. Fazemos $b = \|T\|_\infty$ e $a = -b$. Como T é uma matriz tridiagonal

$$\|T\|_\infty = \max(|a_1| + |b_1|, |b_{i-1}| + |a_i| + |b_i|, |b_{n-1}| + |a_n|) \quad (2.2.25)$$

Com relação ao uso do teo. de Gershgorin - usado para implementar este algoritmo - para matrizes simétricas tridiagonais, os raios r_i serão determinados por

$$\begin{aligned} r_1 &= |b_1| \\ r_i &= |b_{i-1}| + |b_i| \quad i=2,3,\dots,n-1 \\ r_n &= |b_{n-1}| \end{aligned} \quad (2.2.26)$$

de tal forma que podemos usar os seguintes limites

$$a = \min_i (a_i - r_i) \quad e \quad b = \max_i (a_i + r_i).$$

Em geral, estes limites são melhores que os limites produzidos quando usamos as normas matriciais.

2.2.4 Cálculo dos Autovetores da Matriz A

Conhecendo os autovalores da matriz A, uma das alternativas para calcular seus autovetores é através da solução de n sistemas lineares homogêneos do tipo

$$(A - \lambda_i I)x = 0 \quad i=1,2,\dots,n. \quad (2.2.27)$$

Para utilizar essa estratégia, devemos notar que a eficiência do algoritmo global está diretamente vinculada ao fato de podermos usar toda a informação até agora obtida, isto é, o cálculo individualizado dos autovalores pelo método da bisseção intervalar repetida, assim como, as transformações de Householder utilizadas para obter a matriz tridiagonal T.

Propomos neste trabalho que, para obtermos todos os autovetores x_i , $i=1,2,\dots,n$ da matriz A, sejam resolvidos inicialmente n sistemas lineares tridiagonais da forma

$$(T - \lambda_i I)y_i = 0 \quad (2.2.28)$$

e em seguida, utilizando as matrizes H_i , $i=1,2,\dots,n-2$ da primeira etapa, calcular os autovetores x_i , $i=1,2,\dots,n$ da matriz A.

A solução dos sistemas homogêneos tridiagonais (2.2.28), será obtida através de sistemas auxiliares não-homogêneos.

Vamos agora descrever como resolver um sistema tridiagonal do tipo (2.2.28). Para isto suponha $i=k$. Temos então que obter o vetor não-nulo y_k (autovetor de T associado ao autovalor λ_k).

$$\begin{cases} g_{i,i+1} \leftarrow c \cdot g_{i,i+1} + b \cdot g_{i+1,i+1} \\ g_{i+1,i+1} \leftarrow -b \cdot g_{i,i+1} + c \cdot g_{i+1,i+1} \\ d_i \leftarrow c \cdot d_i + b \cdot d_{i+1} \\ d_{i+1} \leftarrow -b \cdot d_i + c \cdot d_{i+1} \end{cases} \quad i=1,2,\dots,n-1 \quad (2.2.31)$$

onde

$$c = g_{ii} / u \quad , \quad b = g_{i+1,i} / u \quad , \quad u = \sqrt{g_{ii}^2 + g_{i+1,i}^2} .$$

O número de passos necessários para obter o sistema na forma triangular superior será no máximo $(n-1)$, caso em que todos os $b_i \neq 0$. No caso de $b_i = 0$, $i=2,3,\dots,n$, não efetuamos a transformação.

Obtido os autovetores da matriz tridiagonal T , soluções dos sistemas (2.2.28), vamos descrever como proceder para encontrar os autovetores da matriz A .

De acordo com a descrição da etapa E1, encontramos $(n-2)$ vetores w que definem as transformações H_i , $i=1,2,\dots,n-2$. Através destes vetores, é possível determinar todos os autovetores da matriz A . O processo consiste no seguinte:

Através de (2.2.2) para $k=(n-2)$ temos

$$T = A_{n-2} = (H_1 H_2 \dots H_{n-2})' A (H_1 H_2 \dots H_{n-2}) \quad (2.2.32)$$

fazendo $H = H_1 H_2 \dots H_{n-2}$, (2.2.32) fica $T = H' A H$. Se y_k é um dos autovetores da matriz T associado ao autovalor λ_k , temos

$$T y_k = \lambda_k y_k$$

$$H T y_k = \lambda_k H y_k$$

$$HTH'Hy_k = \lambda_k Hy_k$$

$$AHy_k = \lambda_k Hy_k \quad (2.2.33)$$

portanto, $x_k = Hy_k$ é um autovetor da matriz A associado ao autovalor λ_k .

Seja $H_j = I - 2w_j w_j'$ e u_j um dado vetor não-nulo do R^n . Então

$$\begin{aligned} H_j u_j &= u_j - 2w_j w_j' u_j \\ &= u_j - 2(w_j' u_j) w_j. \end{aligned} \quad (2.2.34)$$

Definindo

$$\beta_j \doteq (w_j' u_j)$$

de (2.2.34) temos

$$H_j u_j = u_j - 2\beta_j w_j. \quad (2.2.35)$$

Substituindo o vetor genérico u_j pelo autovetor y_k da matriz T, através da relação de recorrência

$$u_{j-1} = u_j - 2\beta_j w_j, \quad j=(n-2), \dots, 2, 1 \quad (2.2.36)$$

obtemos o autovetor da matriz A, que será dado por $x_k = u_0$.

2.2.5 Exemplos

Algumas experiências numéricas com matrizes testes foram realizadas. Apresentaremos somente dois exemplos que comprovam o

desempenho do algoritmo proposto. Os cálculos foram realizados em precisão simples e os resultados finais foram arredondados para seis casas decimais.

Exemplo 1 (Cálculo do autosistema completo) A matriz A é dada por:

$$A = \begin{bmatrix} 2.0 & 1.0 & 3.0 & 4.0 \\ 1.0 & -3.0 & 1.0 & 5.0 \\ 3.0 & 1.0 & 6.0 & -2.0 \\ 4.0 & 5.0 & -2.0 & -1.0 \end{bmatrix}$$

Solução:

$$\text{Autovalores} \left\{ \begin{array}{l} \lambda_1 = 7.932905 \\ \lambda_2 = 5.668864 \\ \lambda_3 = -1.573190 \\ \lambda_4 = -8.028580 \end{array} \right.$$

$$\text{Autovetores} \left\{ \begin{array}{l} x'_1 = [.560145, .211633, .776708, .195382] \\ x'_2 = [-.378703, -.362419, .537943, -.660199] \\ x'_3 = [.688048, -.624123, -.259801, -.675573] \\ x'_4 = [.263462, .659041, -.199634, -.675573] \end{array} \right.$$

Exemplo 2 (Cálculo de dois elementos do autosistema) A matriz $A(a_{ij})$ de ordem 11 é definida por:

$$a_{ij} = \begin{cases} 6 & \text{para } i = j \\ 3 & \text{" } |i - j| = 1 \\ 1 & \text{" } |i - j| = 2 \\ 1 & \text{" } |i - j| = 3 \\ 0 & \text{" } |i - j| > 3 \end{cases}$$

Exceto $a_{11} = a_{11,11} = 5$
 $a_{21} = a_{12} = a_{10,11} = a_{11,10} = 2$

Solução:

Autovalores $\left\{ \begin{array}{l} \lambda_1 = \lambda_{\max} = 14.941821 \\ \lambda_{11} = \lambda_{\min} = .522281 \end{array} \right.$

Autovetores $\left\{ \begin{array}{l} x'_1 = x'_{\max} = (-1) \cdot [.105662, .204124, .288675, \\ \quad .353553, .394338, .408248, \\ \quad .394337, .353553, .288675, \\ \quad .204124, .105662] \\ x'_{11} = x'_{\min} = [.105663, -.204124, .2288673, \\ \quad -.353554, .394338, -.408248, \\ \quad .394337, -.353553, .288677, \\ \quad -.204125, .105663] \end{array} \right.$

2.3. Um Problema Especial de Autovalores

Entre os muitos problemas de calcular um elemento do autovalor de uma matriz simétrica, aquele de determinar o par $(\lambda_{\min}, x_{\min})$ ou eventualmente o par $(\lambda_{\max}, x_{\max})$ desempenha um papel proeminente nos algoritmos propostos nos capítulos subsequentes.

Para resolver este problema, utilizamos a função quociente de Rayleigh. Através de sua minimização (maximização) encontramos a solução desejada. Esta minimização é levada a efeito através de um ou outro algoritmo que serão aqui propostos.

2.3.1 Minimização (Maximização) da Função R.

Vamos mostrar inicialmente, que para calcular o autovalor mínimo (máximo) e seu autovetor associado de uma matriz simétrica $A(n \times n)$, é suficiente minimizar (maximizar) a função quociente de Rayleigh.

Teo. 2.1. : Os autovetores de uma matriz simétrica $A(n \times n)$ são dados pelos pontos críticos da função quociente de Rayleigh, R . Os correspondentes autovalores de A são os valores críticos de R .

Prova: De acordo com a def.1.2.4., a função R é definida pela razão

$$R(x) = N(x)/D(x) \quad (2.3.1)$$

onde $N(x) = x'Ax$ e $D(x) = x'x$ com $x \neq 0$. O vetor gradiente de R , no ponto x , $\nabla R(x)$, será dado por

$$\nabla R(x) = [D(x)(dN(x)/dx) - N(x)(dD(x)/dx)]/[D(x)] \quad (2.3.2)$$

Sendo $dN(x)/dx = 2Ax$ e $dD(x)/dx = 2x$, substituindo estas expressões em (2.3.2) e simplificando temos

$$\nabla R(x) = 2 \cdot [Ax - R(x)x] / (x'x), \quad x \neq 0 \quad (2.3.3)$$

Como, por definição, os pontos críticos da função quociente de Rayleigh são aqueles para os quais $\nabla R(x) = 0$ se $x \neq 0$ então $\nabla R(x) = 0$ se e somente se

$$Ax = R(x)x, \quad x \neq 0 \quad (2.3.4)$$

Comparando (2.3.4) com (1.2.1), temos que os pontos críticos da função R são os autovetores da matriz A e $\lambda = R(x)$, seus correspondentes autovalores.

A função R é uma função homogênea de grau zero, isto é, $R(ax) = R(x)$, $x \neq 0$ e a um escalar. Um caso particular de interesse, para os nossos propósitos é quando $a = 1/\|x\|$. Neste caso temos:

$$R(x/\|x\|) = R(x) \quad (2.3.5)$$

Desde que $x/\|x\|$ é um vetor unitário, segue que os valores da função R sobre qualquer variedade linear V no R^n serão determinados pelos vetores unitários em V .

Lema 1 : Seja V uma variedade linear no espaço R^n . Existem vetores unitários x_{\min} e x_{\max} em V tal que

$$R(x_{\min}) \leq R(x) \leq R(x_{\max}) \quad (2.3.6)$$

para todo $x \neq 0$ em V . Os valores $\lambda_{\min} = R(x_{\min})$ e $\lambda_{\max} = R(x_{\max})$ são os valores mínimos e máximos de R em V .

Prova: Vamos representar por W o conjunto de todos os vetores unitários em V . W é um conjunto compacto. Desde que a função R é contínua em W , segue que R atinge seu valor mínimo, λ_{\min} e máximo, λ_{\max} , respectivamente nos pontos x_{\min} e x_{\max} pertencentes a W . Temos desta maneira que

$$\lambda_{\min} = R(x_{\min}) \leq R(x) \leq R(x_{\max}) = \lambda_{\max}$$

para todo x não-nulo em W .

Lema 2 : Os gradientes

$$\nabla R(x_{\min}) = 2(Ax_{\min} - \lambda_{\min} x_{\min})$$

(2.3.7)

$$\nabla R(x_{\max}) = 2(Ax_{\max} - \lambda_{\max} x_{\max})$$

de R nos pontos x_{\min} e x_{\max} como no Lema 1, são ortogonais a variedade linear V no \mathbb{R}^n .

Prova : Sendo V uma variedade linear no \mathbb{R}^n , o vetor gradiente $\nabla R(x_{\min})$ é ortogonal a V [33], assim como para o vetor $\nabla R(x_{\max})$. Desde que $\|x_{\min}\| = \|x_{\max}\| = 1$, as expressões para $\nabla R(x_{\min})$ e $\nabla R(x_{\max})$ decorrem de (2.3.3)

Teo. 2.2 : Um ponto mínimo, x_{\min} , da função quociente de Rayleigh em \mathbb{R}^n (com $x \neq 0$) é um autovetor da matriz $A(n \times n)$. O valor $R(x_{\min})$ da função neste ponto é o autovalor mínimo de A .

Prova : Se no Lema 2, $V = \mathbb{R}^n$, temos que o gradiente $\nabla R(x_{\min})$ é ortogonal a todo vetor $x \in \mathbb{R}^n$ e, portanto, $\nabla R(x_{\min}) = 0$. Pelo teo.2.1, x_{\min} é um autovetor da matriz A . Pelo Lema 1, temos que $\lambda_{\min} = R(x_{\min})$

é o autovalor mínimo de A , o que prova o teorema.

Um teorema análogo pode ser enunciado para o par $(\lambda_{\max}, x_{\max})$.

2.3.2 Algoritmos

Após ter estabelecido os resultados que conduzem a formulação do problema especial de autovalores - P.E.A. - como um problema de otimização vamos agora focalizar a questão de obter a solução deste problema sob o ponto de vista numérico.

Sendo a solução do P.E.A., os pontos extremos da função quociente de Rayleigh, é muito natural que qualquer uma das técnicas usadas para otimizar funções não lineares seja aplicada para encontrar esta solução. Entretanto, certos compromissos de eficiência devem ser respeitados. Por exemplo, taxa de convergência, espaço em memória de computadores, precisão numérica para citar alguns.

Entre os métodos disponíveis, se a prioridade é a taxa de convergência, o método de Newton será prontamente o escolhido. Ocorre que, para usar este método, precisamos em cada iteração, calcular a matriz Hessiana - matriz de derivadas segundas - da função R e inverter esta matriz ou resolver um sistema de equações lineares para encontrar a direção de busca. Por outro lado, se a escolha do método é pelo aspecto de simplicidade, o método do gradiente ótimo - "steepest descent" - seria o preferido. Contudo, este método apesar de simples, apresenta sérios problemas de convergência [33].

Tendo em vista satisfazer o maior número de fatores que conduzem a um eficiente algoritmo de otimização, propomos, como mostrado a seguir, para resolver o P.E.A., dois algoritmos baseados na idéia de direções conjugadas. Além disso, a idéia de aplicar métodos de

direções conjugadas para minimizar (maximizar) funções não-quadráticas - caso da função quociente de Rayleigh - é fundamentada no princípio de que se um algoritmo funciona bem para otimização de funções quadráticas, ele pode ser modificado para se obter um eficiente algoritmo de otimização para funções não-quadráticas.

2.3.2.a Algoritmo A1

Vamos assumir que a matriz Hessiana da função R - denotada por G - seja positiva definida no domínio de busca do ponto de mínimo.

Assim, podemos enunciar o algoritmo A1.

Passo 1. - (Inicialização) Seleccione um ponto x e $h > 0$. Faça

$$z_1 = 0, \quad r_1 = -g_1 \quad \text{e} \quad p_1 = r_1$$

Passo 2. - (Ciclo Gradientes Conjugados) Para $k=1,2,\dots,n$, faça

$$1 - s_k = [g(x_1 + h_k p_k) - g(x_1)]/h_k, \quad h_k = h/\|p_k\| \quad (2.3.8)$$

$$2 - a_k = c_k/d_k, \quad c_k = p_k' r_k, \quad d_k = p_k' s_k \quad (2.3.9)$$

$$3 - z_{k+1} = z_k + a_k p_k, \quad r_{k+1} = -g_{k+1} \quad (2.3.10)$$

$$4 - p_{k+1} = r_{k+1} + b_k p_k, \quad b_k = \|r_{k+1}\|^2 / \|r_k\|^2 \quad (2.3.11)$$

Passo 3. - (Pare ou Reinicialize) Faça $x^* = x_1 + z_{n+1}$. Se

$\|g(x^*)\| \leq \epsilon$, $\epsilon > 0$, pare. Se não, faça $x_1 = x^*$ e repita o passo 2.

Comentários : 1) O algoritmo A1 pode ser considerado como uma variante do método de Newton. Vale lembrar que o método de Newton para minimização de funções é definido pela relação

$$x_{k+1} = x_k - G_k^{-1} g_k$$

onde x_{k+1} é o ponto crítico (mínimo) da aproximação quadrática

$$F_k(x) = f(x_k) + g'_k(x - x_k) + .5 (x - x_k)' G(x - x_k) \quad (2.3.12)$$

da função f próximo de x_k . Assim, em cada iteração, o método minimiza uma quadrática do tipo (2.3.12) e toma este ponto mínimo de F_k como a melhor estimativa para o ponto mínimo de f .

O método de Newton tem a vantagem de convergir, rapidamente, quando está próximo à solução, apresentando, em geral, convergência superlinear ou quadrática quando a função é de classe C^2 . Por outro lado, a cada iteração, exige um esforço computacional que o torna inviável para determinados problemas, como é o caso do P.E.A..

Para contornar esta desvantagem do método de Newton, sem perder sua excelente taxa de convergência, o algoritmo A1, no lugar de calcular segundas derivadas da função R , usa aproximações como dadas pela relação (2.3.8). Note que esta relação aproxima o limite

$$\lim_{h \rightarrow 0} [g(x_1 + hp_k) - g(x_1)]/h = G_1 p_k.$$

2) No passo 2, resolvemos um problema quadrático aproximação da função R , em torno do ponto x_k .

3) O escalar a_k é o tamanho do passo, na direção p_k . z_{k+1} é uma aproximação da direção de Newton. r_k é o resíduo, direção de descida. p_{k+1} é a nova direção conjugada, combinação linear do gradiente no ponto x_{k+1} e a direção anterior p_k . A fórmula usada para b_k , é equivalente àquela para o caso quadrático $b_k = - (g_k' r_{k+1}) / d_k$ como pode facilmente ser verificado.

4) No passo 3, temos a reinicialização (restart) do ciclo de gradientes conjugados, fazendo $p_{k+1} = -g_{k+1}$. Note que isto é uma parte importante do algoritmo, visto que assegura que um ciclo de gradientes conjugados aproxima um passo do método de Newton. Além do mais, na pior das hipóteses, o algoritmo A1 pode ser pensado como passos espaçados do método "Steepest Descent".

Com relação a convergência do algoritmo, podemos dizer que a convergência global é garantida pelo fato de A1, em situação mais desfavorável, ser considerado um método Steepest Descent.

Com objetivo de garantir que a sequência funcional (R_k) , seja sempre decrescente e também acelerar o processo de convergência do algoritmo, no fim de cada ciclo gradientes conjugados, um parâmetro α_k é selecionado como sendo o maior α pertencente ao intervalo, $0 < \alpha \leq 1$, para o qual a desigualdade

$$R(x_1 + \alpha z_{n+1}) - R(x_1) \leq \eta \alpha g_1' z_n, \quad 0 < \eta \leq 1/2$$

se verifica.

Após a obtenção de α_k , o novo ponto x^* é dado por

$$x^* = x_1 + \alpha_k z_{n+1}$$

x' , no caso, será a solução se o critério de parada for satisfeito ou será o novo ponto inicial para mais uma iteração do algoritmo.

2.3.2.b - Algoritmo A2

Algumas modificações podem ser feitas no algoritmo A1, permitindo a elaboração de outros efetivos algoritmos de direções conjugadas.

No lugar de aproximar $G_1 p_k$, vamos fazer

$$x_k = x_1 + z_k \quad (3.2.13)$$

e substituir a fórmula (3.2.8) por

$$s_k = [g(x_k + h_k p_k) - g(x_k)]/h_k \quad (3.2.14)$$

Temos assim, uma aproximação de $G_k p_k$. O novo ciclo de gradientes conjugados é modificado para:

$$s_k = [g(x_k + h_k p_k) - g(x_k)]/h_k, \quad h_k = h/\|p_k\| \quad (3.2.15)$$

$$a_k = c_k/d_k, \quad c_k = p_k' r_k, \quad d_k = p_k' s_k \quad (3.2.16)$$

$$x_{k+1} = x_k + a_k p_k, \quad r_{k+1} = -g_{k+1} \quad (3.2.17)$$

Quando k/n é um inteiro, fazemos $p_{k+1} = r_{k+1}$. Se não

$$p_{k+1} = r_{k+1} + b_k p_k, \quad b_k = -(s_k' r_{k+1})/d_k. \quad (3.2.18)$$

Entretanto, a menos que a função objetivo seja quadrática, o ponto x_{k+1} como em (3.2.17) não necessariamente minimiza a função na direção $x = x_k + tp_k$. Novamente modificamos o algoritmo, de modo que selecionamos $t^* = a_k$, com t minimizando a função de uma variável

$$f_k(t) = R(x_k + tp_k). \quad (3.2.19)$$

Assim, a obtenção do tamanho do passo a_k será feita resolvendo um problema de minimização de uma variável, conhecido como busca unidimensional.

Se $t^* = a_k$ minimiza a função f , através de (3.2.19), a derivada de f no ponto $t^* = a_k$ será

$$\begin{aligned} df_k/dt &= g(x_k + a_k p_k)' p_k \\ &= g_{k+1}' p_k \end{aligned} \quad (3.2.20)$$

que é igual a zero, dado que a_k minimiza f_k . Assim sendo, $r_{k+1} = -g_{k+1}$ é ortogonal a direção p_k e

$$\begin{aligned} p_{k+1}' r_{k+1} &= (r_{k+1} + b_k p_k)' r_{k+1} \\ &= \|r_{k+1}\|^2, \end{aligned} \quad (3.2.21)$$

De (3.2.16) e (3.2.21) concluímos que $c_k = \|r_k\|^2$ para todo k .

Redefinindo o vetor s_k por

$$s_k = [g(x_k + a_k p_k) - g(x_k)]/a_k \quad (3.2.22)$$

temos de (3.2.17) que

$$s_k = (r_k - r_{k+1})/a_k \quad (2.2.23)$$

e

$$d_k = p'_k s_k$$

$$d_k = [p'_k (r_k - r_{k+1})]/a_k$$

$$d_k = (p'_k r_k)/a_k$$

$$d_k = c_k/a_k$$

Como no algoritmo A1, obtemos $a_k = c_k/d_k$.

Uma nova expressão para b_k pode ser obtida pois:

$$s'_k r_{k+1} = [(r_k - r_{k+1})/a_k] r_{k+1}$$

e

$$b_k = - (s'_k r_{k+1})/d_k$$

$$= [(r_{k+1} - r_k)/c_k] r_{k+1}$$

Então

$$b_k = [||r_{k+1}||^2 - (r'_k r_{k+1})]/||r_k||^2 \quad (2.2.24)$$

De posse das relações (3.2.17) e (3.2.24) podemos enunciar o algoritmo A2.

Passo 1. - (Inicialização) Selecione o ponto x_1 . Faça

$$r_1 = -g_1, \quad p_1 = r_1$$

Passo 2. - (Ciclo Gradientes Conjugados) Para $k = 1, 2, \dots$

1. - Encontre $t^* = a_k$ que minimize $f_k(t) = R(x_k + tp_k)$

$$2. - x_{k+1} = x_k + a_k p_k, \quad r_{k+1} = -g_{k+1} \quad (2.2.25)$$

3. - Se k/n é um inteiro, faça $p_{k+1} = r_{k+1}$. Caso contrário, faça

$$4. - p_{k+1} = r_{k+1} + b_k p_k, \quad b_k = (\|r_{k+1}\|^2 - r_k' r_{k+1}) / \|r_k\|^2 \quad (2.2.26)$$

Passo 3. - (Pare ou Reinicialize) Se no k -ésimo passo $\|g_k\| \leq \epsilon$, $\epsilon > 0$, pare. Se não, com $x_1 = x_{k+1}$ volte ao passo 2.

Comentários : 1) No passo 2.1, em cada ciclo de gradientes conjugados temos que resolver um problema de minimização unidimensional. Para uma função não-linear qualquer, esta minimização é feita através de rotinas especializadas, por exemplo, interpolação cúbica, seção áurea, busca de Fibonacci etc.. No caso da função R , é possível resolver este problema de forma analítica, como veremos mais adiante. 2) Assim como para o algoritmo A1, o processo de reinicialização é parte importante de A2, pelos mesmos motivos já citados. 3) No passo 3, temos a possibilidade do algoritmo atingir o ponto ótimo desejado, antes de completar o ciclo de gradientes conjugados.

Sobre a convergência global do algoritmo A2, ela é garantida pelo fato que passos "Steepest Descente" são realizados toda vez que k/n é um inteiro [33].

2.3.3 Implementação dos Algoritmos A1 e A2

Os algoritmos A1 e A2 foram implementados num computador e experiências com algumas matrizes testes foram realizadas. Alguns pontos práticos importantes para um bom desempenho são a seguir apontados.

2.3.3.a - Escolha do Ponto Inicial

De um modo geral, nos algoritmos iterativos, como é o caso de A1 e A2, a escolha do ponto inicial, a menos de alguns casos particulares, desempenha um regra muito importante. No caso particular da função quociente de Rayleigh, R , em que o número de pontos críticos é igual a ordem da matriz, torna-se necessário que a escolha do ponto x_1 seja feita obedecendo determinados critérios. Para isto sugerimos duas formas distintas de escolher um bom ponto inicial.

A primeira consiste em extrair toda informação possível sobre a estrutura do problema a ser resolvido, seja ela de ordem física, econômica, etc. (ver exemplo 3)

A segunda forma, baseada mais nos elementos da matriz do problema, consiste no seguinte procedimento (quando o P.E.A. é de encontrar o par $(\lambda_{\min}, x_{\min})$):

Use como ponto inicial o ponto associado ao vetor da base canônica, cuja componente diferente de zero é aquela que tem o mesmo índice do menor elemento da diagonal da matriz do problema (exem. 4).

Se existir mais que um elemento na diagonal satisfazendo esta condição, a escolha será pelo ponto que fornece a maior norma do gradiente.

2.3.3.b - Uso da Homogeneidade da Função R.

Fazendo uso da propriedade de homogeneidade da função quociente de Rayleigh, a cada ciclo do algoritmo A1, o novo ponto encontrado é normalizado. No algoritmo A2, a normalização é realizada dentro do ciclo de gradientes conjugados.

O processo de normalização, consiste na divisão de todos os componentes do ponto (vetor) pelo componente de maior valor absoluto.

2.3.3.c - Escolha de h no Algoritmo A1

No algoritmo A1, o vetor s é calculado através de um quociente de diferenças que aproxima $G_1 p_k$. Teóricamente, quando $h \rightarrow 0$, temos o método de Newton como limite do algoritmo A1. Na prática, devido a possíveis erros de arredondamentos, não podemos permitir a h tender a zero. Sugerimos usar o valor $h = 10^{-4}$, quando as operações forem realizadas em dupla precisão.

2.3.3.d - Busca Unidimensional em A2.

O problema de otimização unidimensional que aparece no algoritmo A2, passo 2.a, em geral é resolvido por uma das muitas rotinas especializadas para este tipo de problema. No caso do P.E.A., como dispomos da expressão analítica do gradiente, é perfeitamente razoável

tentar encontrar alguma forma de resolver o problema unidimensional analiticamente. Para isto, vamos usar a condição $g'_{k+1} p_k = 0$.

Sendo o gradiente de R no ponto x dado por

$$\nabla g(x) = 2.[Ax - R(x)x]/(x'x) \quad x \neq 0$$

e $x = x_k + tp_k$, através de $g'_{k+1} p_k = 0$, podemos determinar o valor $t = a_k$ que minimize a função $f_k(t) = R(x_k + tp_k)$. Este valor de t é uma das raízes da equação do segundo grau

$$at^2 + bt + c = 0 \quad (2.2.27)$$

onde os coeficientes a, b e c são dados por :

$$a = (p_k' A p_k) (x_k' p_k) - (x_k' A p_k) (p_k' p_k)$$

$$b = (p_k' A p_k) (x_k' x_k) - (x_k' A x_k) (p_k' p_k)$$

$$c = (x_k' A p_k) (x_k' x_k) - (x_k' A x_k) (x_k' p_k)$$

A raiz positiva de (2.2.27) é usada para determinação do par $(\lambda_{\min}, x_{\min})$ e a raiz negativa para o par $(\lambda_{\max}, x_{\max})$.

3.3.● - Reinicialização Prematura em A2.

Experiências numéricas realizadas com A2 mostraram que: quando a maior raiz da equação do segundo grau (2.2.27) é negativa a direção p_k encontrada para busca, é uma direção de subida e não de descida

como é o desejo para o caso de minimização. Isto faz com que a sequência de valores funcionais (R_k) deixe de ser monótona decrescente, passando a oscilar. A correção para esta dificuldade foi feita através do que denominamos reinicialização prematura. Tomamos como ponto inicial, para o novo ciclo de gradientes conjugados, o ponto gerado pelo algoritmo na iteração anterior ao cálculo da raiz negativa. A direção de busca p será o negativo do vetor gradiente neste ponto. Esta implementação pode ser efetuada através do teste dos valores funcionais de R .

2.3.4 Exemplos

Exemplo 3 (Ponto inicial)

$$A = \begin{bmatrix} 1.7 & -1. & 0. \\ -1. & 2. & -1. \\ 0. & -1. & 2. \end{bmatrix}$$

Através do exemplo 3, procuramos ilustrar a importância de um bom ponto inicial. Este exemplo provém de um problema de mecânica vibratória [36]. Pela física do problema, sabemos que o autovetor associado ao autovalor mínimo deve ter todas as componentes com o mesmo sinal. Escolhendo para ponto inicial, $x_1 = (1., 1., 1.)$, a primeira aproximação para o autovalor mínimo de A é igual a $\lambda_1 = .566667$. Com duas iterações do algoritmo A1, obtivemos $\lambda_{\min} = .5$ e $x'_{\min} = [-.833333, -1., -.666667]$. Para uma escolha arbitrária do ponto inicial $x_1 = (1., -1., -2.)$ com $R(x_1) = 1.616667$ encontramos o mesmo resultado em três iterações.

Exemplo 4 (Reinicialização Prematura)

$$A = \begin{bmatrix} 10. & 9. & 8. & 7. & 6. & 5. & 4. & 3. & 2. & 1. \\ & 10. & 9. & 8. & 7. & 6. & 5. & 4. & 3. & 2. \\ & & 10. & 9. & 8. & 7. & 6. & 5. & 4. & 3. \\ & & & 10. & 9. & 8. & 7. & 6. & 5. & 4. \\ & & & & 10. & 9. & 8. & 7. & 6. & 5. \\ & & & & & 10. & 9. & 8. & 7. & 6. \\ & & & & & & 10. & 9. & 8. & 7. \\ & & & & & & & 10. & 9. & 8. \\ & & & & & & & & 10. & 9. \\ & & & & & & & & & 10. \end{bmatrix}$$

simétrica

O exemplo 4 ilustra o fenômeno de reinicialização prematura. Aplicando o algoritmo A2, com ponto inicial $x_1 = (0,0,\dots,1)$ encontramos $R(x_1) = 10.0$. No passo 2 do algoritmo, os valores funcionais decresceram até o valor .840934, quando subitamente passou a 13.312128 antes de completar o ciclo de gradientes conjugados. Em virtude disto, assumimos como ponto inicial, para a próxima iteração de A2 o ponto correspondente ao valor funcional .840934. Após 10 iterações encontramos a solução :

$$\lambda_{\min}(A) = 0.512254284$$

$$x'_{\min}(A) = [-0.069862, 0.202875, -0.316027, 0.398190, \\ -0.441555, 0.441785, -0.398697, 0.316527, \\ -0.203298, 0.700062]$$

Capítulo 3

Problemas de Otimização com Restrições Matriciais Semidefinidas

3.1. Introdução

Alguns problemas da vida real, quando modelados como problemas de otimização, apresentam em suas formulações algumas restrições do tipo semidefinidas.

Estes problemas de natureza matricial podem ser, formalmente, representados por

$$\begin{aligned} \min f(P) \\ \text{s.a } g(P) \leq 0 \\ P \geq 0 \end{aligned} \quad (3.1.1)$$

onde $f: \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$, $g: \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{m \times m}$ e P uma matriz real simétrica.

A classe de problemas definida por (3.1.1), aparentemente restritiva, engloba um número significativo de problemas de otimização matricial. Uma ótica, para efeito de solução, sob a qual podemos abordá-los, é quanto ao número e a posição dos elementos da matriz P que podem variar. Por exemplo, sendo P uma matriz simétrica, o número máximo de elementos variáveis será sempre igual a $n(n+1)/2$. Outro grupo de problemas do tipo (3.1.1), de real interesse para nossos propósitos, é aquele em que somente os elementos localizados na diagonal da matriz P são variáveis, permanecendo os demais fixos.

Problemas do tipo (3.1.1), com alguma estrutura especial, serão o tema principal deste capítulo. Aqui, vamos analisar as restrições matriciais semidefinidas, propor uma metodologia baseada no método de hiperplanos de cortes para resolver esses problemas, e aplicar esta metodologia, através da elaboração de algoritmos numéricos, com objetivo de resolver dois problemas existentes na literatura.

O primeiro aparece em estatística, e é denominado problema do Teste Educacional [14]. O segundo tem sua origem no estudo da estabilidade de sistemas lineares econômicos [28] e sistemas de controles descentralizados, através da solução da equação matricial de Lyapunov [19].

Expressões para os vetores que caracterizam os hiperplanos de cortes, serão fornecidas. Análise de convergência e alguns exemplos numéricos comprovando o desempenho dos algoritmos propostos, também serão fornecidos.

Especificamente no caso do problema do Teste Educacional, determinamos um coeficiente que mede o grau de confiabilidade de um exame ou teste, constituído de vários subtestes.

Para o caso de sistemas dinâmicos lineares invariantes no tempo, contínuo ou discreto, além de encontrar a solução diagonal positiva da equação de Lyapunov, se ela existir, um parametro que fornece uma estimativa do grau de estabilidade ótimo do sistema é encontrado.

3.2. A Restrição Matricial Semidefinida

O aparecimento de restrições matriciais semidefinidas em problemas de otimização conduz a um tratamento especial desses problemas, seja pelo significado atribuído às restrições, seja pela dimensão que

apresenta quando transformado em um problema na forma padrão.

Como definido anteriormente, uma restrição matricial como aquela de (3.1.1) significa que para todo vetor não-nulo $y \in \mathbb{R}^n$, temos

$$y'(g(P))y \leq 0.$$

Na hipótese de g ser uma função linear em P , podemos sempre escrever (3.1.1) na forma

$$\begin{aligned} \min f(P) \\ \text{s.a } Q \leq 0 \\ P \geq 0 \end{aligned} \quad (3.2.1)$$

onde, $Q \in \mathbb{R}^{m \times m}$.

Uma função que está intimamente relacionada ao fato da matriz Q ser semidefinida é a função autovalor. Através do critério de autovalores, sendo Q uma matriz semidefinida negativa, todos os autovalores de Q serão não-positivos, isto é,

$$\lambda_i(Q) \leq 0 \quad i=1,2,\dots,n.$$

Supondo os autovalores da matriz Q ordenados segundo

$$\lambda_{\max} = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n = \lambda_{\min},$$

podemos escrever (3.2.1) da seguinte maneira:

$$\begin{aligned} \min f(P) \\ \text{s.a } \lambda_{\max}(Q) \leq 0 \\ P \geq 0. \end{aligned} \quad (3.2.2)$$

Como foi ilustrado no exemplo 7 (cap.1), o conjunto das matrizes simétricas negativas semidefinidas é um cone convexo. Vamos mostrar agora que este resultado pode ser obtido como consequência do seguinte teorema.

Teorema 3.1 : A função autovalor máximo, $\lambda_{\max}(Q)$, definida no conjunto das matrizes reais simétricas é uma função convexa.

Prova : Através da definição de norma espectral temos que $\|Q\| = \sqrt{\lambda_{\max}(Q'Q)}$. Sendo Q uma matriz simétrica, $\|Q\| = \lambda_{\max}(Q)$. Segundo o exemplo 9 (cap.1), a função norma $\|Q\|$, é uma função convexa. Portanto, podemos concluir que a função autovalor máximo de uma matriz simétrica é convexa.

Usando o teo.3.1, podemos facilmente provar que a função autvalor mínimo, $\lambda_{\min}(Q)$, de uma matriz real simétrica é côncava.

Esses resultados são muito significativos para solução de problemas do tipo (3.2.1) ou (3.2.2), pois, no caso da função objetivo ser uma função convexa teremos então um problema de otimização convexa. Este é o caso dos problemas aqui abordados.

3.3. Método dos Hiperplanos de Corte

A abordagem feita aos problemas de otimização convexa através do método de hiperplanos de corte, consiste fundamentalmente em resolver uma sequência de problemas, mais simples, que são aproximações do problema original.

A idéia de hiperplano de corte aparece em virtude de que, a medida que o método vai se desenvolvendo, pontos soluções ótimas dos proble-

mas aproximados vão sendo separados do conjunto de factibilidade do problema original através de hiperplanos adequadamente escolhidos. Esta estratégia permite obter a solução de problemas de otimização convexa - solução única - mesmo que o conjunto de factibilidade apresente algumas dificuldades não tratáveis por outros métodos.

Rigorosamente relacionada à eficiência dos algoritmos que usam esta metodologia, para resolver problemas de otimização, está a escolha do hiperplano que efetua os cortes. Isto é, desta escolha depende o processo de convergência do algoritmo, assim como seu maior ou menor grau de complexibilidade para implementação.

Restritos aos problemas de otimização com restrições matriciais semidefinidas do tipo (3.2.2), vamos apresentar uma maneira de escolher os hiperplanos de cortes.

3.3.1 Vetor Que Caracteriza o Corte

Para o caso de problemas do tipo (3.2.2), uma das maneiras pela qual podemos caracterizar os hiperplanos de cortes depende de um vetor parametrizado $c(\cdot)$. Em virtude da natureza matricial do conjunto de factibilidade, a parametrização será feita através dos elementos de um conjunto de vetores unitários do \mathbb{R}^n . Para isto, vamos definir o conjunto

$$Y = \{y \in \mathbb{R}^n : \|y\| = 1\}$$

A determinação de uma expressão para o vetor $c(\cdot)$ depende da forma como se apresenta a função $g(P)$ em (3.1.1). Nas referências [19] e [45], uma expressão definindo este vetor é fornecida, sem maiores

comentários. Aqui, como tratamos de diferentes problemas, para obter as expressões de definição de $c(\cdot)$ além de mostrar como elas surgem, propomos uma regra muito simples que evita seu cálculo direto.

Inicialmente, vamos supor que a matriz Q no problema (3.2.1) pode ser colocada na forma $Q = L(P)$ onde P é uma matriz diagonal, $P = \text{diag}(x_1, x_2, \dots, x_n)$ e L uma função linear de P .

Através do conhecimento explícito da expressão que define $L(P)$, podemos calcular o vetor $c(y)$. Por exemplo em [45], para

$$L(P) = A'P + PA$$

onde A é uma matriz real constante de ordem n , $c(y)$ foi definido como

$$c(y) = 2 \text{diag}(y_i) Ay \quad i=1,2,\dots,n$$

com $y \in Y$.

Para chegar a esta expressão de $c(y)$, podemos fazer uso da função autovalor máximo, ou seja

$$L(P) = A'P + PA \leq 0$$

que pode ser substituída por

$$\lambda_{\max}(A'P + PA) \leq 0$$

onde

$$\lambda_{\max}(A'P + PA) = \max_{y \in Y} y'(A'P + PA)y$$

$$= \max_{y \in Y} h(x, y)$$

com $P = \text{diag}(x_1, x_2, \dots, x_n)$ e $x = [x_1, x_2, \dots, x_n]'$.

Calculando a derivada da função h em relação a x , obtemos a expressão que define o vetor $c(y)$.

A regra que propomos para determinar a expressão do vetor $c(y)$, evita o cálculo explícito de derivadas. Através de um simples arranjo na expressão que define a função h , é possível determinar o vetor $c(y)$. A regra é a seguinte:

Tomamos a função

$$h(P, y) = y' L(P) y \quad (3.3.1)$$

e isolamos a matriz P em cada termo do segundo membro de (3.3.1). Isto é possível porque estamos supondo que $L(P)$ é linear em P . Desse modo, a derivada de h em relação a P ou x será obtida, derivando cada termo no lado direito de (3.3.1). A derivada de cada um destes termos será o produto de uma matriz diagonal, cujos elementos são as componentes do vetor a direita da matriz P , por um vetor cujas componentes são as mesmas do vetor a esquerda da matriz P .

Para ilustrar o uso desta regra, vamos tomar

$$L(P) = A'P + PA.$$

A função h será definida por:

$$h(P, y) = y' (A'P + PA) y$$

arranjando os termos do segundo membro, de modo que a matriz P fique isolada, temos

$$h(P, y) = (Ay)'(P)y + y'(P)Ay.$$

Com isso, a expressão do vetor $c(y)$ será

$$c(y) = \text{diag}(y_i)Ay + \text{diag}((Ay)_i)y.$$

Como é simples verificar, os dois termos no segundo membro são iguais. Logo, $c(y)$ será dado por

$$c(y) = 2 \text{diag}(y_i)Ay \quad i=1, 2, \dots, n.$$

Outro exemplo seria para o caso da função L ser definida por

$$L(P) = AP - B$$

com A e B matrizes reais fixas de ordem n .

A função h será

$$\begin{aligned} h(P, y) &= y'(AP - B)y \\ &= (A'y)'(P)y - y'By \end{aligned}$$

e o vetor $c(y)$ será dado por

$$c(y) = \text{diag}(y_i)A'y \quad i=1, 2, \dots, n.$$

Note que o segundo termo do segundo membro é uma constante em relação a matriz P .

3.4. Problema do Teste Educacional

Um exemplo de problemas de otimização com restrição matricial, se encontra em psicometria (Estatística). O problema, na sua forma original, chamado problema do Teste Educacional [14], procura responder a seguinte questão:

Suponha N estudantes submetidos a um exame ou teste, composto de n subtestes. Através dos resultados obtidos pelos estudantes, como estimar o grau de confiabilidade desses resultados?

Para determinar um coeficiente, ρ , que estima o grau de confiabilidade, precisamos resolver um problema do tipo (3.2.2), em que a função objetivo é definida como o negativo do traço de uma matriz diagonal, $f(P) = -\text{tr}(P)$, e as restrições definidas como $-Q = A - P \geq 0$, $P \geq 0$ sendo $P = \text{diag}(x_1, x_2, \dots, x_n)$. Portanto, para determinar uma estimativa do coeficiente ρ , precisamos resolver o seguinte problema:

$$\begin{aligned} \min f(P) &= -\text{tr}(P) \\ \text{s.a. } A - P &\geq 0 \\ P &\geq 0 \end{aligned} \tag{3.4.1}$$

No caso deste problema, a matriz constante A é simétrica e positiva definida, $A > 0$, e seus elementos, a_{ik} , são obtidos de uma tabela de dados de ordem $N \times n$, onde N é o número de estudantes submetidos ao exame e n o número de subtestes que constituem o exame [14].

Representando esta tabela pela matriz $B(N \times n)$, os elementos a_{ik} serão dados por :

$$a_{ik} = (1/(N-1)) \sum_l (b_{lj} - \bar{b}_j)(b_{lk} - \bar{b}_k) \quad (3.4.2)$$

sendo $\bar{b}_j = (1/R) \sum_l b_{lj}$ e b_{lj} o resultado obtido pelo estudante l no subteste j .

Um limite inferior para o coeficiente de confiabilidade, ρ , dos resultados, é obtido através de

$$\rho^2 \geq 1 - (f_{\max} / d) \quad (3.4.3)$$

onde d é um escalar obtido somando todos os elementos da matriz Λ .

Independente de se originar de uma aplicação na área do Estatística, do ponto de vista matemático, podemos pensar em (3.4.1) como decorrente da seguinte questão:

Seja A uma matriz real simétrica e positiva definida. Quanto pode ser subtraído de sua diagonal e ainda resultar uma matriz positiva semidefinida?

Seja P uma matriz diagonal, $P = \text{diag}(x_1, x_2, \dots, x_n)$, a função objetivo de (3.4.1), $f(P) = -\text{tr}(P)$, pode ser escrita como

$$f(x) = - \sum_{i=1}^n x_i,$$

uma função linear das variáveis x_i , $i=1,2,\dots,n$. As restrições, no caso, serão $A - \text{diag}(x_i) \geq 0$ e $x_i \geq 0$ $i=1,2,\dots,n$.

Desse modo, o problema (3.4.1) pode ser escrito como:

$$\begin{aligned} \min f(x) &= - \sum_{i=1}^n x_i \\ \text{s.a. } A - \text{diag}(x_i) &\geq 0 \\ x_i &\geq 0 \quad i=1,2,\dots,n \end{aligned} \quad (3.4.4)$$

Alguns resultados notáveis, relacionados a (3.4.1) ou (3.4.4) podem ser deduzidos de imediato. Sendo f uma função linear nas variáveis x_i , é uma função convexa. Assim, (3.4.4) é um problema convexo, além de que, sendo A uma matriz positiva definida, necessariamente $a_{ii} - x_i \geq 0, i=1,2,\dots,n$. Portanto, o conjunto de restrições de (3.4.4) é um conjunto compacto. Com isto, podemos concluir que (3.4.4) tem sempre uma única solução. Note também que, sendo f uma função linear, a solução ótima de (3.4.4) se localiza sobre o contorno do conjunto de soluções factíveis.

Como já dissemos anteriormente, a metodologia proposta para resolver (3.4.1) ou (3.4.4), faz uso do método de hiperplanos de cortes. Assim, antes de enunciar o algoritmo, vamos determinar a expressão do vetor $c(y)$ que caracteriza o hiperplano de corte para (3.4.4) e demonstrar algumas de suas propriedades.

3.4.1 Expressão do Vetor $c(y)$ e suas Propriedades

Sendo $L(P) = A - P$, a função h (seq.3.3.1) será definida por:

$$\begin{aligned} h(P,y) &= y'(A - P)y \\ &= y'Ay - y'(P)y. \end{aligned}$$

Aplicando a regra estabelecida anteriormente aos dois termos do segundo membro da igualdade, obtemos a expressão para o vetor

$$c(y) = - \text{diag}(y_i) y \quad i=1,2,\dots,n \quad (3.4.5)$$

Duas propriedades do vetor $c(y)$, de interesse são:

P1 - Para todo $x, y \in \mathbb{R}^n$

$$c(y)'x = - y'Py. \quad (3.4.6)$$

Prova : Trivial.

Esta propriedade será útil na demonstração da propriedade P2, a seguir, e no teorema de convergência do algoritmo.

P2 - Para todo $x \in \mathbb{R}^n$ e $y \in Y$

$$c(y)'x + y'Ay \geq 0. \quad (3.4.7)$$

Prova : De P1, temos que

$$c(y)'x = - y'Py$$

para todo $x, y \in \mathbb{R}^n$. Logo (3.4.6) se verifica para todo $y \in Y$. Adicionando $y'Ay$ aos dois membros desta igualdade, temos

$$\begin{aligned} c(y)'x + y'Ay &= -y'Py + y'Ay \\ &= y'(A - P)y \\ &\geq 0 \end{aligned}$$

o que prova P2.

3.4.2. Algoritmo

Utilizando o vetor $c(y)$ e suas propriedades podemos enunciar o seguinte algoritmo:

Passo 1 - Faça $k=n$ e o vetor $y^j \in Y$ $j=1,2,\dots,k$ igual a j -ésima coluna da matriz identidade I .

Passo 2 - Resolver o seguinte prob. de Programação Linear (P.L.):

$$\begin{aligned} \min f(x) &= - \sum_{i=1}^n x_i \\ \text{s.a. } c(y^j)'x + y^{j'}Ay^j &\geq 0 \quad j=1,2,\dots,k \quad (3.4.8) \\ x_i &\geq 0 \quad i=1,2,\dots,n \end{aligned}$$

Seja x^* , solução ótima do P.L.

Passo 3 - Para $x = x^*$, determine a matriz $(A - P)$ e seu autovalor mínimo, $\lambda_{\min}(A - P)$.

Passo 4 - Se $\lambda_{\min}(A - P) \geq 0$, $P = \text{diag}(x_1, x_2, \dots, x_n)$ é a solução ótima do problema, pare. Se não, ir para o passo 5.

Passo 5 - Determine o autovetor $v \in Y$ associado ao autovalor $\lambda_{\min}(A - P)$. Faça $k \leftarrow k+1$, $y^k = v$ e retorne ao passo 2.

Comentários ; 1) Inicialmente, geramos um politópo que contém o conjunto de factibilidade do problema a ser resolvido.

2) No passo 2, a cada iteração, precisamos resolver um problema de Programação Linear (P.L.). Vamos nos deter um pouco sobre as peculiaridades desses P.L.'s. Para isso, vamos considerar uma iteração k do algoritmo. Imediatamente, verificamos que o conjunto de factibilidade do P.L. gerado na iteração $(k+1)$, está contido naquele da iteração k . Isto decorre do corte efetuada no politópo da iteração k , pelo hiperplano

$$c(y^{k+1})'x + y^{k+1}'Ay^{k+1} = 0, \quad (3.4.9)$$

separando a solução ótima x do conjunto de factibilidade do problema original.

Com relação ao politópo gerado na iteração $(k+1)$, ele é formado pela interseção do politópo da iteração k com o semi-espaco não-negativo gerado por (3.4.9).

Deve também ser destacado o fato de que o P.L. da iteração k difere do P.L. da iteração $(k+1)$, somente de uma restrição. Esta particularidade do método de hiperplanos de corte, é muito útil para efeito de implementação do algoritmo num computador digital.

Tudo que foi comentado até agora sobre os P.L.'s são características gerais do método de hiperplanos de cortes. No caso restrito à solução do problema (3.4.1), acrescentamos uma característica notável dos subproblemas lineares (3.4.8). Quando aplicamos o método Simplex, para resolver os P.L.'s, não se faz necessário o uso da Fase I, isto é, temos sempre disponível uma solução básica inicial. Além disso, na passagem de uma iteração para outra, podemos usar como solução básica inicial a solução ótima da iteração anterior. Isto permite uma grande economia de tempo e espaço em memória de computadores.

Uma questão que resta é quanto ao número de restrições com as quais devemos trabalhar em cada iteração. Dado que este número cresce com o número de iterações, várias tentativas para resolver essa questão foram feitas no sentido de trabalhar com o menor número possível de restrições. Estratégias heurísticas são propostas por vários autores mas nenhuma garante convergência. Neste trabalho usamos o método de hiperplanos de cortes padrão como aparece em [33].

3) Nos passos 3 e 5, temos o cálculo de um elemento específico do autosistema da matriz simétrica $(A-P)$, no caso autovalor mínimo e autovetor associado. Esta questão foi abordada, com profundidade, no capítulo 2 deste trabalho.

4) No passo 4, temos o critério de otimalidade do algoritmo. No exemplo 2, Fig.3.2, traçamos o gráfico dos autovalores versus iteração, mostrando o comportamento típico da sequência de autovalores mínimos gerada pelo algoritmo.

5) No passo 5, usamos o autovetor associado a $\lambda_{\min}(A-P)$ como parâmetro no vetor $c(\cdot)$.

3.4.3 Exemplos

Para testar a realizabilidade do algoritmo, ele foi implementado num computador digital usando a linguagem Fortran. Alguns exemplos numéricos foram resolvidos, tendo sido comprovado que o algoritmo tem um bom desempenho para os propósitos que ele foi projetado.

Exemplo 1 - (Geometria do Método) Com este exemplo, procuramos mostrar através da fig.3.1, as noções geométricas do método dos hiperplanos de corte aplicado ao problema (3.4.1). Suponhamos que a

matriz A seja dada por:

$$A = \begin{bmatrix} 5 & 3 \\ 3 & 4 \end{bmatrix}$$

Solução:

Iteração 1.

$$x^{*1} = (5.0, 4.0)$$

$$\lambda_{\min}(A-P) = -3.0$$

$$c(y) = [0.5, 0.5]'$$

Iteração 2.

$$x^{*2} = (3.0, 0.0)$$

$$\lambda_{\min}(A-P) = -0.162278$$

$$c(y) = [0.658114, 0.341886]'$$

Iteração 3.

$$x^{*3} = (2.753420, 0.0)$$

$$\lambda_{\min}(A-P) = -0.002189$$

$$c(y) = [0.640252, 0.359748]'$$

Iteração 4.

$$x^{*4} = (2.750000, 0.0)$$

$$\lambda_{\min}(A-P) = 0.0$$

A matriz solução é $P = \text{diag}(2.75, 0.0)$. O coeficiente de confiabilidade tem como limite inferior $\rho = 0.904$.

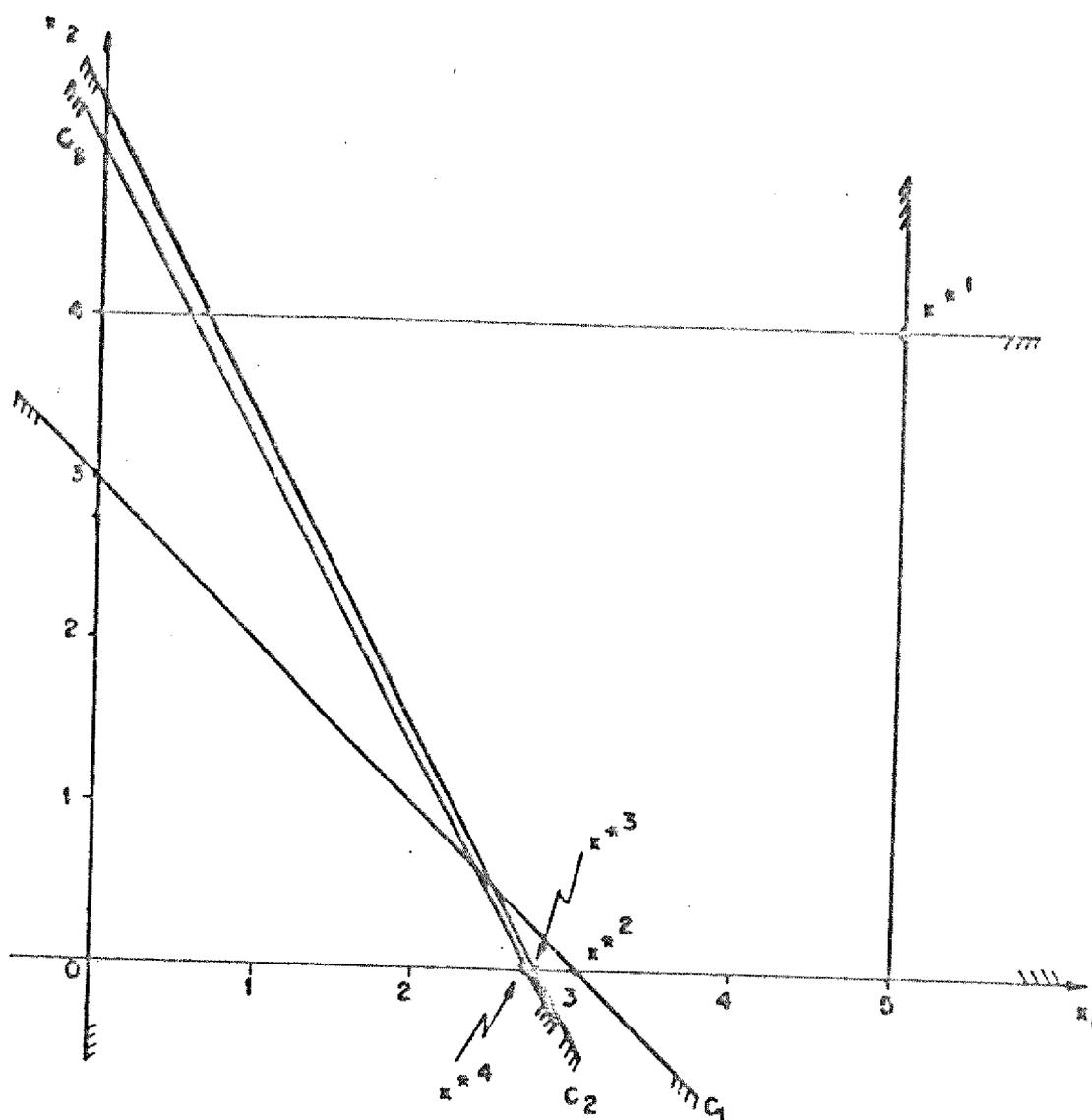


fig.3.1

Exemplo 2. (Comportamento de $\lambda_{\min}(A-P)$). Neste exemplo procuramos, através da fig 3.2, mostrar que a sequência de autovalores mínimos é monotônica crescente. Suponha que a matriz A é dada por:

$$A = \begin{bmatrix} 3.5 & 2.0 & 3.0 \\ 2.0 & 6.0 & 3.0 \\ 3.0 & 3.0 & 8.0 \end{bmatrix}$$

Solução :

ITER	x			$\lambda_{\min} (A-P)$
1	(3.50000,	6.00000,	8.00000)	- 3.35889
2	(3.50000,	6.00000,	2.53578)	- 2.00000
3	(3.50000,	2.00000,	3.78923)	- 1.59925
4	(1.51962,	3.98038,	3.78936)	- 0.14003
5	(1.58560,	3.91440,	3.47985)	- 0.58014
6	(1.49886,	4.00114,	3.50529)	- 0.00274
7	(1.50031,	3.99969,	3.50029)	- 0.00030

tab.3.1

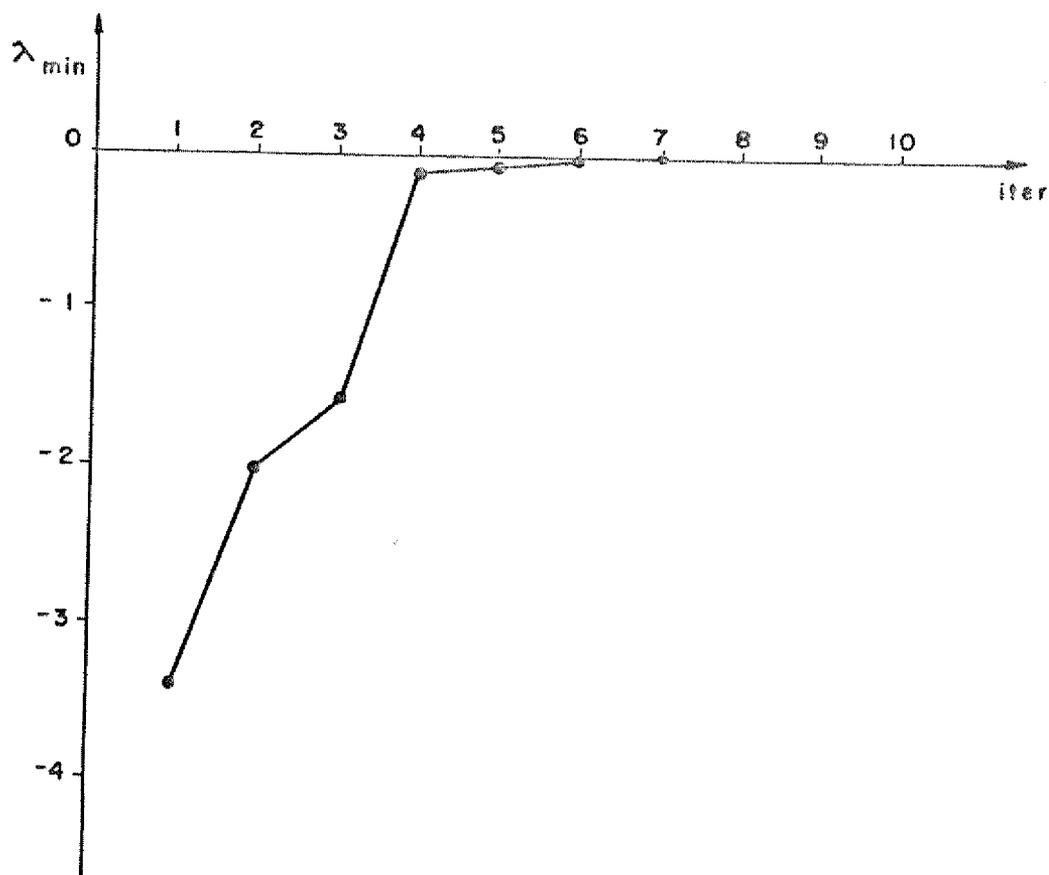


fig.3.2

Exemplo 3. (Ilustração do problema (3.4.4)) Para o caso de uma matriz A positiva definida, quanto podemos subtrair de sua diagonal e ainda reter uma matriz positiva semidefinida ?

Suponha que A seja dada por:

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\ 1 & 2 & 3 & 4 & 4 & 4 & 4 & 4 & 4 & 4 \\ 1 & 2 & 3 & 4 & 5 & 5 & 5 & 5 & 5 & 5 \\ 1 & 2 & 3 & 4 & 5 & 6 & 6 & 6 & 6 & 6 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 7 & 7 & 7 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 8 & 8 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 9 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{bmatrix}$$

Solução:

$$P = \text{diag}(0.35776, 0.30805, 0.0000, 0.00000, 0.50042, \\ 0.00000, 0.50012, 0.0000, 0.50000, 0.00000)$$

3.4.4 - Convergência

A convergência global do algoritmo será estabelecida pelo seguinte teorema.

Teorema 3.2 - Seja $(x^{(i)})$ uma sequência de pontos do R^n , gerada pelo algoritmo. Então, se A é uma matriz positiva definida, qualquer

ponto limite desta sequência é uma solução do seguinte problema

$$\max_{x \in S} F(x) \quad (3.4.10)$$

onde F é definida por $F(x) = -f(x)$ do problema (3.4.4) e

$$S = \{x \in \mathbb{R}^n : c(y)'x \geq -y' Ay \forall y \in Y\}.$$

Prova: Note que, pela propriedade P2 do vetor $c(y)$, o problema (3.4.10) é equivalente ao problema (3.4.1). Vamos supor que x seja um ponto limite da sequência $(x^{(i)})$ (ou qualquer subsequência de $(x^{(i)})$). Assim, podemos obter a matriz $(A-P)$, com $P = \text{diag}(x_1, x_2, \dots, x_n)$. Vamos agora definir, o conjunto de restrições lineares

$$S_k = \{x \in \mathbb{R}^n : c(y^i)'x \geq -y^i' Ay^i, i = 1, 2, \dots, k\} \quad (3.4.11)$$

e tomar o problema de

$$\max_{x \in S_k} F(x) \quad (3.4.12)$$

Por construção, temos

$$S_i \supset S_{i+1} \supset \dots \supset S \quad (3.4.13)$$

e

$$F_i^* \geq F_{i+1}^* \geq \dots \geq F^* \quad (3.4.14)$$

Além disso, a última restrição gerada pelo algoritmo para $i = k$, é dada por

$$c(y^k)'x > -y^k' Ay^k \quad (3.4.15)$$

Pela propriedade P1 do vetor $c(y)$, temos

$$c(y^k) \cdot x^k = -y^{k'} P_k y^k \quad (3.4.16)$$

De (3.4.15) e (3.4.16) obtemos

$$\begin{aligned} c(y^k)(x - x^k) &\geq y^{k'} (P_k - A)y^k \\ &\geq \lambda_{\min}(P_k - A) \end{aligned} \quad (3.4.17)$$

Naturalmente, (3.4.17) deve ser satisfeita para qualquer interação subsequente $j > k$, implicando que

$$\begin{aligned} \lambda_{\min}(A - P_k) &> c(y^k) \cdot (x^k - x^j) \\ &\geq \|c(y^k)\| \cdot \|x^k - x^j\|. \end{aligned} \quad (3.4.18)$$

Desde que $\|c(y)\|$ é limitado para todo $y \in Y$, o lado direito de (3.4.18) irá para zero quando j e k tendem ao infinito. O lado esquerdo tende para $\lambda_{\min}(A - P)$. Assim $\lambda_{\min}(A - P) \geq 0$ e x é factível para (3.4.10), isto é $x \in S$.

Se F^* é o valor ótimo de (3.4.10) e para cada k temos $F_k^* = F(x^k)$, então, através de (3.4.14) e devido a continuidade da função F

$$F(x) \geq F^*.$$

Portanto, x é uma solução ótima, o que conclui a prova do teorema.

3.5. Solução Diagonal Positiva de Equação de Lyapunov

No estudo de condições de estabilidade em Economia [28], no estudo de estabilidade de sistemas interconectados [59] e algumas classes de sistemas não-lineares [19], uma questão que frequentemente aparece, é aquela de investigar sobre a existência de uma solução diagonal positiva, $P = \text{diag}(x_1, x_2, \dots, x_n)$, da equação matricial de Lyapunov, seja para o caso discreto

$$A'PA - P + Q = 0, \quad Q' = Q > 0, \quad (3.5.1)$$

seja para o caso contínuo

$$A'P + PA + Q = 0, \quad Q' = Q > 0. \quad (3.5.2)$$

Nas referências [28] e [19], essa questão é modelada como um problema de otimização do tipo (3.1.1), onde a restrição matricial semidefinida é dada por

$$L(P) = A'P + PA.$$

Aqui, usaremos o mesmo procedimento aplicado ao problema do Teste Educacional para resolver este problema. Além disso, determinamos uma estimativa do grau de estabilidade ótimo do sistema dinâmico.

Do ponto de vista formal, temos dois problemas para resolver. Um, relativo a equação (3.5.1), para sistemas dinâmicos discretos e o outro referente a equação (3.5.2), associada aos sistemas dinâmicos contínuos. Como os resultados para ambos são muito semelhantes,

deduziremos aqueles referentes aos sistemas dinâmicos discretos e, simplesmente citaremos os seus correspondentes para o caso contínuo.

3.5.1 - Expressão do Vetor $c(y)$ e suas Propriedades

Antes de obter uma expressão para o vetor $c(y)$, vamos definir o conjunto das matrizes que admitem uma solução diagonal positiva da equação de Lyapunov.

Def.3.1 - Seja A uma matriz real $n \times n$, $x \in \mathbb{R}^n$ e $P = \text{diag}(x_1, x_2, \dots, x_n)$ com x_i , $i=1,2,\dots,n$ componentes do vetor x . Definimos o conjunto

$$D = \{A \in \mathbb{R}^{n \times n}; \text{ existe } P, \text{ sol. da eq. } A'PA - P + Q = 0, Q' = Q > 0 \}.$$

Uma definição análoga, existe para eq. $A'P + PA + Q = 0$, $Q' = Q > 0$.

A expressão de definição do vetor $c(y)$ para o caso de

$$L(P) = A'PA - P$$

é facilmente obtida fazendo

$$h(P,y) = y'(A'PA - P)y$$

$$= (Ay)'(P)Ay - y'(P)y$$

Aplicando a regra da seq.3.1, temos

$$c(y) = \text{diag}((Ay)_i)Ay - \text{diag}(y_i)y \quad i=1,2,\dots,n. \quad (3.5.3)$$

Para o caso contínuo, $L(Q) = A'P + PA$, $c(y)$ é dado por:

$$c(y) = 2 \operatorname{diag}(y_i) \lambda y \quad i=1,2,\dots,n \quad (3.5.4)$$

Este último resultado foi obtido anteriormente, quando da apresentação da regra para o cálculo do vetor $c(\cdot)$.

Com base em (3.5.3) ou (3.5.4) podemos demonstrar duas propriedades do vetor $c(y)$.

P3 - Para todo $x, y \in R^n$

$$c(y)'x = y'(A'PA - P)y. \quad (3.5.5)$$

Prova : Trivial.

P3.a - Para todo $x, y \in R^n$

$$c(y)'x = y'(A'P + PA)y. \quad (3.5.5a)$$

P4 - Existe uma matriz P , com $Q = -(A'PA - P) > 0$, se e somente se existir $x \in R^n$ factível, com $c(y)'x \leq -1$, para todo $y \in Y$.

Prova :=>) Por hipótese, existe uma matriz P , com $Q > 0$. Assim, para todo $y \in Y$ temos $y'(-Q)y < 0$. Pela propriedade P3, $y'(-Q)y = c(y)'x < 0$. Portanto, existe um $x_0 \in R^n$ tal que $c(y)'x_0 < -\epsilon, \epsilon > 0$. Fazendo $x = (x_0/\epsilon)$, temos que $c(y)'x \leq -1$, o que prova a condição =>).

<=) Por hipótese, existe $x \in R^n$ com $c(y)'x \leq -1$, para todo $y \in Y$. Consequentemente existe x tal que $c(y)'x < 0$. Pela propriedade P3, $c(y)'x = y'(-Q)y < 0$ para todo y não-nulo pertencente ao R^n e, em particular para todo $y \in Y$. Sendo $-Q = A'PA - P$ com A uma matriz fixa,

a existência de P é prontamente verificada o que conclui a prova.

P4.a - Existe uma matriz P , com $Q = -(A'P + PA) > 0$, se e somente se existe $x \in \mathbb{R}^n$ com $c(y)'x \leq -1$, para todo $y \in Y$.

3.5.2 - Grau de Estabilidade

Utilizando as propriedades P3 (P3.a) e P4 (P4.a) do vetor $c(y)$ e definindo um critério para encontrar P diagonal, se existe uma, solução da equação de Lyapunov, chegaremos a um problema de otimização do tipo

$$\min_{x \in S} f(x) \quad (3.5.6)$$

com $f: \mathbb{R}^n \rightarrow \mathbb{R}$ e

$$S \doteq \{x \in \mathbb{R}^n : c(y)'x \leq -1 \text{ para todo } y \in Y\}.$$

Uma função critério pode ser determinada através da tentativa de se obter o grau de estabilidade do sistema.

Seja o sistema linear invariante no tempo, autônomo

$$w(t+1) = Aw(t) \quad t = 0, 1, 2, \dots \quad (3.5.7)$$

assintoticamente estável. Tomando $V(w(t)) = w(t)'Pw(t)$, $w \in \mathbb{R}^n$ e P diagonal, como uma função de Lyapunov para (3.5.7), temos

$$\begin{aligned} V(w(t+1)) &= w(t+1)'Pw(t+1) \\ &= w(t)'A'P Aw(t) \quad t = 0, 1, 2, \dots \quad (3.5.8) \end{aligned}$$

Subtraindo (3.5.7) de (3.5.8) temos

$$\begin{aligned} V(w(t+1)) - V(w(t)) &= w(t)'A'PAw(t) - w(t)'Pw(t) \\ &= w(t)'(A'PA - P)w(t) \quad t = 0, 1, 2, \dots \end{aligned} \quad (3.5.9)$$

Se $P = \text{diag}(x_1, x_2, \dots, x_n)$ está restrita a ser calculada tal que $x \in S$ como em (3.5.6), então

$$\begin{aligned} V(w(t+1)) - V(w(t)) &= [w(t)'(A'PA - P)w(t)]V(w(t))/[w(t)'Pw(t)] \\ &\leq - (1/\lambda_{\max}(P)) \cdot V(w(t)) \end{aligned} \quad (3.5.10)$$

sendo P uma matriz diagonal

$$\begin{aligned} \lambda_{\max}(P) &= \max_{1 \leq i \leq n} (x_i) \\ &= \min_z (z : z \geq x_i, i = 1, 2, \dots, n). \end{aligned} \quad (3.5.11)$$

Nossa meta será minimizar $\lambda_{\max}(P)$ sobre o conjunto S . Portanto, através de (3.5.11), isto é equivalente a

$$\min_{x \in S, z} (z : z \geq x_i, i = 1, 2, \dots, n). \quad (3.5.12)$$

Resolvendo o problema de otimização (3.5.12), obtemos $P = \text{diag}(x_1, x_2, \dots, x_n)$ solução da equação de Lyapunov (3.5.1).

Para estimar o grau de estabilidade ótimo do sistema (3.5.7), seja z^* o valor ótimo da função objetivo em (3.5.12). De (3.5.10) temos

$$V(w(t+1)) - V(w(t)) \leq - (1./z^*) V(w(t))$$

logo

$$V(w(t+1)) \leq (1. - (1./z^*)) V(w(t)).$$

Assim,

$$V(w(t)) \leq (1 - (1./z^*))^t V(w(0)). \quad (3.5.13)$$

Mas,

$$\lambda_{\min}(P) \|w(t)\|^2 \leq V(w(t)) \quad \text{e} \quad V(w(0)) < \lambda_{\max}(P) \|w(0)\|^2 \quad (3.5.14)$$

Logo de (3.5.13) e das relações (3.5.14) temos

$$\|w(t)\|^2 \leq (\lambda_{\max}(P)/\lambda_{\min}(P)) (1. - (1./z^*))^t \|w(0)\|^2$$

ou

$$\|w(t)\| \leq \sqrt{(\lambda_{\max}(P)/\lambda_{\min}(P))} \cdot \|w(0)\| \cdot \sqrt{(1 - (1./z^*))^t} \quad t = 0, 1, \dots$$

Concluimos assim, que a melhor estimativa para o grau de estabilidade do sistema (3.5.7) será:

$$\rho = \sqrt{1 - (1./z^*)}. \quad (3.5.15)$$

Para o caso de um sistema linear contínuo no tempo, autônomo

$$\dot{x}(t) = Ax(t) \quad t > 0 \quad (3.5.16)$$

assintoticamente estável, a função de Lyapunov terá a mesma forma daquela para o caso discreto (ver ex.10 - cap 1). Os passos para determinação do grau de estabilidade ótima, σ , serão os mesmos. Sua expressão será dada por:

$$\sigma = 1./2.z .$$

3.5.3 - Algoritmo

Passo 1 : Faça $k = n$ e os vetores $y^j \in Y$, $j = 1, 2, \dots, k$, iguais a j -ésima coluna da matriz identidade.

Passo 2 : Resolver o seguinte P.L.

$$\begin{aligned} \min \quad & z \\ \text{s.a} \quad & c(y^j)'x \leq -1 \quad j = 1, 2, \dots, k \quad (3.5.18) \\ & z \geq x_i \quad i = 1, 2, \dots, n \end{aligned}$$

Se o conjunto de factibilidade é vazio, então $A \notin D$, pare. Se não, seja x^* a solução ótima do P.L., ir para o próximo passo.

Passo 3 : Para $x = x^*$, determine a matriz $Q = -(A'PA - P)$ ou $Q = -(A'P + PA)$, de acordo com o caso, e seu autovalor mínimo, $\lambda_{\min}(Q)$.

Passo 4 : Se $\lambda_{\min}(Q) \geq 1$ então $A \in D$. Calcule a estimativa ótima do grau de estabilidade do sistema, ρ , (caso discreto) ou σ , (caso contínuo), pare. Se não, vá para o passo 5.

Passo 5 : Para $v \in Y$, com v autovetor associado a $\lambda_{\min}(Q)$, faça $k \leftarrow k+1$, $v = y^k$ e volte ao passo 2.

Comentários. Este algoritmo difere levemente daquele proposto para resolver o problema do Teste Educacional. Vamos tecer alguns comentários sobre essas pequenas diferenças.

No passo 2, o problema de Programação Linear (3.5.18), se apresenta numa forma muito adequada para ser resolvido por uma rotina dual-simplex [45].

Ainda no passo 2, temos a verificação quanto a existência ou não da solução diagonal da equação de Lyapunov, o que não era preciso no caso do problema do Teste Educacional, como foi mostrado. Essa verificação é levada a efeito através da solução do P.L.. No caso do conjunto de factibilidade deste problema ser vazio, o algoritmo pára e a conclusão é que o sistema dinâmico não é diagonalmente estável.

No critério de parada, passo 4, a mudança para $\lambda_{\min}(Q) \geq 1$, é uma consequência de que para todo $y \in Y$, temos $y'Qy \geq 1$. Portanto, $Q \geq 1$.

3.5.4 Exemplos

Exemplo 4. Para o sistema discreto

$$x(t+1) = \begin{bmatrix} .577 & 0.0 \\ .577 & .577 \end{bmatrix} x(t)$$

a matriz sol. diagonal da equação de Lyapunov, foi encontrada após 6 iterações:

$$P = \text{diag}(4.287 \quad 2.309)$$

O grau de estabilidade ótimo estimado é: $\rho = .876$

A figura 3.3 mostra o comportamento de z e $\lambda_{\min}(Q)$ em função das iterações.

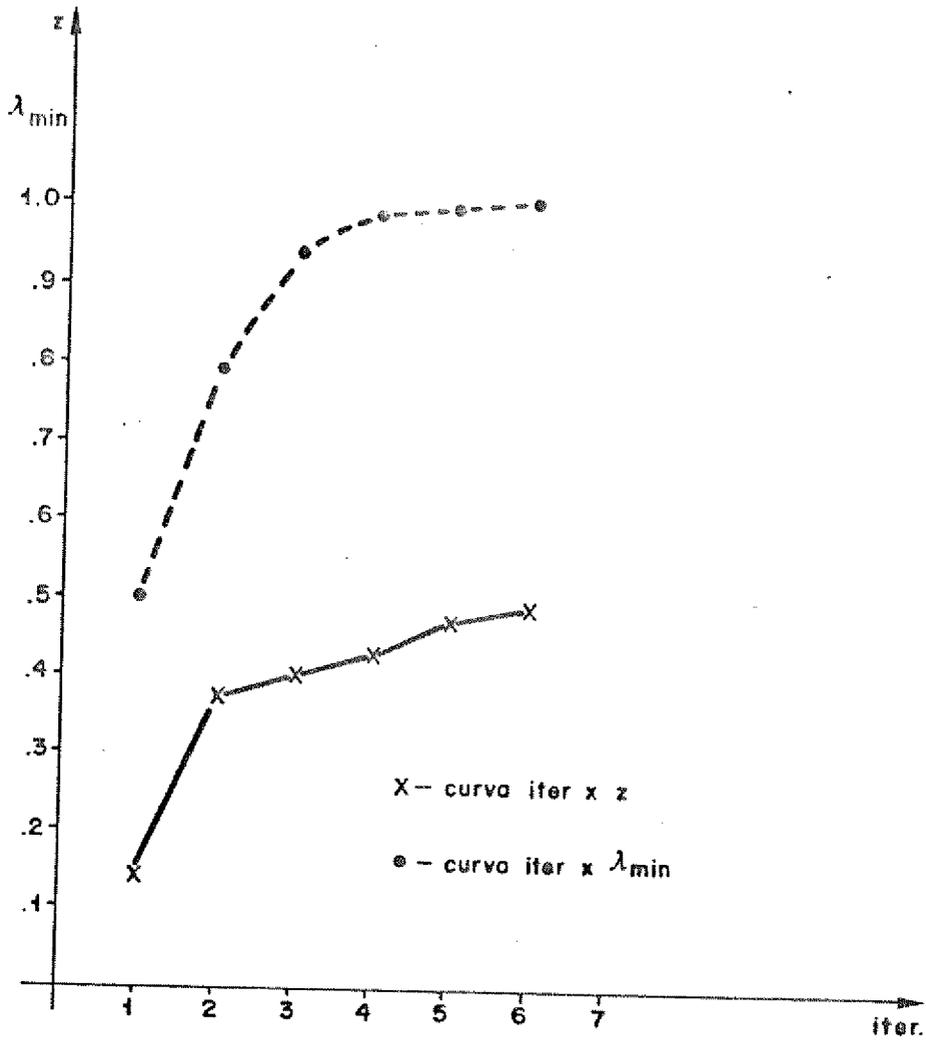


fig 3.3

Exemplo 5. Para o caso do sistema contínuo

$$\dot{x}(t) = Ax(t)$$

onde a matriz A é dada por:

$$A = (-1) \begin{bmatrix} 1.4 & 1.0 & 1.0 & 2.0 & 0.0 & 0.0 \\ 1.0 & 1.8 & 1.0 & 0.0 & 1.0 & 0.0 \\ 1.0 & 1.0 & 2.0 & 0.0 & 0.0 & 2.0 \\ 0.0 & 0.0 & 1.0 & 6.0 & 0.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 9.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 12. \end{bmatrix}$$

a matriz sol. diagonal da equação de Lyapunov foi obtida após 13 iterações:

$$P = \text{diag} (1.17 \quad 1.17 \quad 1.17 \quad 1.17 \quad .669 \quad 1.17)$$

O grau de estabilidade ótimo estimado é: $\sigma = .427$

Exemplo 6. Para o sistema contínuo, com a matriz do exemplo anterior tendo seus três primeiros elementos da diagonal principal alterados para

$$a_{11} = -1.04 \quad a_{22} = -1.08 \quad a_{33} = -1.10$$

após 17 iterações, verificou-se que o sistema não admite solução diagonal da equação de Lyapunov.

Capítulo 4

Robustez de Sistemas Dinâmicos

Lineares Contínuos no Tempo

4.1. Introdução

Uma propriedade muito importante e desejada nos sistemas dinâmicos é a propriedade de robustez, no sentido que o sistema permaneça estável mesmo quando sujeito à perturbações.

Este capítulo é dedicado ao problema de investigar a robustez de sistemas dinâmicos lineares contínuos no tempo. A questão a ser abordada diz respeito à existência e à determinação de uma matriz constante de ganho de realimentação para um dado sistema do tipo $\dot{x}(t) = Ax(t) + Bu(t)$.

Um resultado importante para o estudo da propriedade de robustez de sistemas dinâmicos é devido a Persidskii [39], o qual estabelece que, se uma matriz $A \in \mathbb{R}^{n \times n}$ admite uma solução diagonal positiva da equação de Lyapunov

$$A^T P + PA + Q = 0 \quad (4.1.1)$$

então o sistema $\dot{x}(t) = Ax(t)$ é robusto, isto é, sob hipóteses adicionais o sistema perturbado $\dot{x}(t) = \Lambda g(x)$ é assintoticamente estável para toda não-linearidade $g_i(x)$, $i = 1, 2, \dots, n$ no setor positivo. Vale recordar que no capítulo anterior já tratamos a questão de determinar uma solução diagonal positiva da equação de Lyapunov.

Ainda sobre o resultado devido a Persidskii, podemos concluir que mesmo sob influência de perturbação não lineares dependentes do estado o sistema dinâmico permanece assintoticamente estável. Isto é importante, pois caracteriza uma propriedade estrutural da matriz do sistema [27,35].

Depois dessas considerações podemos dar um enunciado mais formal para a questão aqui abordada.

Para um dado sistema dinâmico linear contínuo no tempo $\dot{x}(t) = Ax(t) + Bu(t)$ com $x \in R^n$ e $u \in R^m$, determinar uma matriz constante de ganho de realimentação $K \in R^{m \times n}$ (fig. 4.1) tal que com $u = -Kx$, a matriz do sistema de malha fechada $(A - BK)$ satisfaça o teorema de Persidskii.

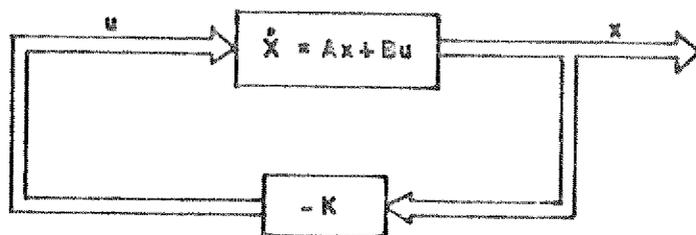


fig.4.1

Para resolver esta questão, fornecemos condições suficientes para existência de uma solução e um algoritmo numérico para sua determinação. O desenvolvimento deste capítulo segue a mesma linha da referência [21].

4.2. Condições Suficientes

Para estabelecer as condições suficientes para existência da matriz de ganho de realimentação, vamos considerar um sistema dinâmico

linear contínuo no tempo

$$\dot{x}(t) = Ax(t) + Bu(t) \quad (4.2.1)$$

onde $x \in R^n$ é o vetor das variáveis de estado e $u \in R^m$ é o vetor das variáveis de controle. Supomos que o par (A,B) é completamente controlável [32] e B é a matriz de posto completo, isto é, $\text{Rank}(B) = m \leq n$. Vamos também, definir os seguintes conjuntos:

$$X = \{x \in R^n : \|x\| = 1\}$$

$$\Gamma = \{W \in R^{n \times n} : W = \text{diag}(w_1, w_2, \dots, w_n); w_i > 0 \quad i=1,2,\dots,n\}. \quad (4.2.2)$$

e

$$D(B) = \{A \in R^{n \times n} : x'S(W)x < 0 \quad \forall x \in \text{Ker}(B') \cap X \text{ para algum } W \in \Gamma\} \quad (4.2.3)$$

onde $S(W) = AW + WA'$. O conjunto $D(B)$ tem algumas interpretações interessantes. Por exemplo, se $B=0$, então $\text{Ker}(B') \cap X = X$, como um resultado, $D(0)$ é o conjunto de todas as matrizes de ordem $n \times n$ que admitem uma solução diagonal positiva da equação de Lyapunov. [Note que $A \in D(0)$ se e somente se $A' \in D(0)$]. Neste caso se $A' \in D(0)$, então (4.2.1) satisfaz o teorema de Persidskii.

Para resolver o problema enunciado anteriormente podemos utilizar o conjunto $D(B)$ da seguinte maneira: encontrar uma matriz constante de ganho de realimentação K de modo que

$$(A - BK)' \in D(0) \quad (4.2.4)$$

Note que este enunciado é absolutamente equivalente a encontrar uma matriz de realimentação K tal que o sistema de malha fechada (4.2.1) com $u = -Kx$ admite uma solução positiva diagonal da equação de Lyapunov.

A consequência de (4.2.4) é realmente importante para robustez do sistema de malha fechada, com respeito a perturbações dependentes do estado. Realmente, se (4.2.4) é verdade, então o sistema

$$\dot{x}(t) = (A - BK)g(x)$$

é assintoticamente estável. Para provar este fato, vamos considerar a função de Lyapunov [27],[39].

$$V(x) = \sum_{i=1}^n w_i \int_0^{x_i} g_i(z) dz \quad (4.2.5)$$

onde $g_i(\cdot)$ é a i -ésima componente da função $g: \mathbb{R}^n \rightarrow \mathbb{R}^n$, que por hipótese pertence ao setor positivo. Assim, $V(x)$ é positiva definida e

$$\begin{aligned} \dot{V}(x) &= \dot{x}' W g(x) \\ &= (1/2) g(x)' S(W) g(x) \\ &< 0 \end{aligned} \quad (4.2.6)$$

onde a última desigualdade é uma consequência de (4.2.4) e (4.2.3). Antes de prosseguir, é necessário o seguinte resultado.

Lema 1 : Se $A \in D(B)$, então existe uma matriz positiva definida $R = R' \in \mathbb{R}^{m \times m}$, tal que $x' S(W)x < x' B R^{-1} B' x \forall x \in X$.

Prova : Vamos escolher a matriz $R = \alpha^{-1}(B'B)$ onde $\alpha \in \mathbb{R}_+$ deve ser determinado. Está claro que, temos que voltar nossa atenção somente para $x \in \mathbb{R}^n$ que pertence também ao cone

$$C = \{x \in \mathbb{R}^n : x'S(W)x \geq 0\}. \quad (4.2.7)$$

Definindo

$$\gamma = \min_{x \in C \cap X} x'B(B'B)^{-1}B'x \quad (4.2.8)$$

é fácil ver que $\gamma > 0$, desde que $\gamma = 0$ se e somente se a solução ótima de (4.2.8) $x^* \in \text{Ker}(B')$, o que é impossível, pois $\lambda \in D(B)$ implica que $C \cap \text{Ker}(B') = \{0\} \notin X$. Em adição vamos também definir

$$\mu = \max_{x \in C \cap X} x'S(W)x \quad (4.2.9)$$

que satisfaz $0 \leq \mu < \lambda_{\max}(S(W))$. Note que se $\mu = 0$, então $S(W)$ será negativa semidefinida e para qualquer $\alpha > 0$ o resultado do lema 1 se verifica. Portanto, consideramos $\mu > 0$. Para concluir a prova, é suficiente notar que a matriz $S(W) - BR^{-1}B'$ é negativa definida se e somente se $x'(S(W) - BR^{-1}B')x < 0 \forall x \in C \cap X$ que com (4.2.8) e (4.2.9) produz

$$x'S(W)x < \mu < (\alpha\gamma) < \alpha x'B(B'B)^{-1}B'x, \quad \forall x \in C \cap X \quad (4.2.10)$$

Impondo que $\alpha > (\mu/\gamma)$. Então, com $0 < R < (\gamma/\mu)B'B$ o lema 1 está provado.

Nota: Em primeira instância, parece ser difícil obter uma matriz R tal que $S(W) - BR^{-1}B' < 0$ desde que os parâmetros γ e μ são fornecidos pela solução ótima de dois problemas não-convexos. Contudo, se λ_+ denota qualquer autovalor não-negativo de $S(W) - BR^{-1}B'$ associado ao autovetor $x_+ \in X$, temos [6]

$$d\lambda_+/d\alpha = -x_+'B(B'B)^{-1}B'x_+ \quad (4.2.11)$$

e $\lambda \in D(B)$ implica que $x_+ \in C \cap X$. Então através de (4.2.8) concluímos que

$$-1 \leq d\lambda_+/d\alpha \leq -\gamma.$$

Isto nos permite concluir que aumentando α obtemos

$$S(W) - \alpha B(B'B)^{-1}B' < 0.$$

Estamos aptos agora a estabelecer o principal resultado deste capítulo.

Teorema 4.1 : Se $\lambda \in D(B)$ e R é uma matriz simétrica positiva definida tal que $S(W) - BR^{-1}B' < 0$, então com $K = R^{-1}B'W^{-1}$, temos

$$(\lambda - BK)' \in D(0).$$

Prova : Desde que $\lambda \in D(B)$, existe $W \in \mathbb{T}$ e $R > 0$ tal que $S(W) - BR^{-1}B' < 0$. Temos que provar a existência de uma matriz P tal que :

$$(\lambda - BR^{-1}B'W^{-1})'P + P(\lambda - BR^{-1}B'W^{-1}) < 0 \quad (4.2.12)$$

Escolhendo $P = W^{-1} \in \Gamma$, o lado esquerdo de (4.2.12) pode ser reescrito como

$$\begin{aligned}
 & (\Lambda - BR^{-1}B'W^{-1})'W^{-1} + W^{-1}(\Lambda - BR^{-1}B'W^{-1}) \\
 & = W^{-1}(S(W) - 2BR^{-1}B')W^{-1} \\
 & < -W^{-1}BR^{-1}B'W^{-1} \\
 & = -K'RK \\
 & < 0
 \end{aligned} \tag{4.2.13}$$

o que completa a prova.

Além de dar uma descrição exata da condição sob a qual existe uma matriz de realimentação tal que o sistema de malha fechada admita uma solução da Equação de Lyapunov, o teo.4.1 mostra também que, tomando

$$Q = W^{-1}(BR^{-1}B' - S(W))W^{-1} > 0 \tag{4.2.14}$$

a matriz $P = W^{-1} \in \Gamma$ resolve a equação de Riccati

$$\Lambda'P + PA - PBR^{-1}B'P + Q = 0. \tag{4.2.15}$$

Conseqüentemente o controle $u = -Kx = -R^{-1}B'Px$ é solução ótima do problema linear-quadrático

$$\min_u \int_0^{\infty} (x'Qx + u'Ru) dt \tag{4.2.16}$$

A existência de $P \in \Gamma$, solução da equação de Riccati (4.2.15), é suficiente para garantir que o sistema de malha fechada apresenta uma propriedade de robustez adicional, com relação a inserção no modelo do

perturbações não-lineares, dependentes do estado.

Para um dado sistema dinâmico (4.2.1) que seja completamente caracterizado pelo par (A, B) , resta determinar se existe uma matriz $U \in \Gamma$ que nos permita saber se $A \in D(B)$. Na próxima seção analisamos este problema e propomos um algoritmo, similar aquele apresentado no capítulo anterior, para obtenção da solução diagonal positiva da equação de Lyapunov.

4.3. Análise Numérica de $D(B)$

O segundo algoritmo apresentado no cap. III, foi desenvolvido para verificar se uma dada matriz quadrada $(n \times n)$ pertence ou não ao conjunto $D(0)$. No que segue, generalizamos este algoritmo de modo a trabalhar com o conjunto $D(B)$.

Primeiro, vamos recordar que as colunas da matriz $T = [t_1, \dots, t_{n-m}]$ onde $t_j \in R^n$, $j = 1, \dots, n-m$, são os autovetores de $I_n - B(B'B)^{-1}B'$ associados aos autovalores $\lambda_j = 1$, $j=1, \dots, n-m$, definem uma base ortogonal para $\text{Ker}(B')$ e t_j pode ser normalizado tal que $T'T = I_{n-m}$. Então

$$x \in (\text{Ker}(B') \cap X) \iff x \in X_y \quad (4.3.1)$$

onde $X_y \doteq \{x \in R^n; x = Ty, \|y\| = 1\}$. Do mesmo modo que na sec. 3.5.1 para o sistema $\dot{x}(t) = Ax(t)$, temos o vetor

$$c(x) = 2 \text{diag}(x_1, x_2, \dots, x_n) A' x \quad (4.3.2)$$

o devido ao fato de $S(W)$ ser uma função linear de W , o conjunto $D(B)$ pode ser reescrito como

$$D(B) = \{\Lambda \in \mathbb{R}^{n \times n} : \sum_{i=1}^m c_i(x) w_i \leq -1, \forall x \in X_y, \text{ para algum } W \in T\} \quad (4.3.3)$$

Deve ser notado que o vetor $c(x)$ definido acima tem a importante propriedade (cap. III)

$$\sum_{i=1}^m c_i(x) w_i = x' S(W) x \quad \forall x \in \mathbb{R}^n \quad (4.3.4)$$

onde $c_i(x)$ $i = 1, \dots, m$ denota a i -ésima componente de $c(x) \in \mathbb{R}^m$.

4.3.1 Algoritmo

Passo 1 : Faça $p = n - m$, se $p = 0$, faça $W^* = I_n$ (matriz identidade) e vá para o passo 5. Caso contrário, faça $x^k = t_k$, $k = 1, 2, \dots, p$

Passo 2 : Resolva o problema de Programação Linear

$$\begin{aligned} \min \quad & z \\ \text{s.a.} \quad & 1 \leq w \leq z \quad i = 1, 2, \dots, n \\ & \sum_{i=1}^m c_i(x^k) w_i \leq -1 \quad k = 1, 2, \dots, p. \end{aligned} \quad (4.3.5)$$

Se (4.3.5) é infactível, então $\Lambda \notin D(B)$. Caso contrário, seja $W^* = \text{diag}(w_1^*, w_2^*, \dots, w_n^*)$ sua solução ótima.

Passo 3 : Para $W = W^*$, calcule o autovalor máximo de $T'S(W)T$ o y^{p+1} o correspondente autovetor com norma igual a um.

Passo 4 : Se $\lambda_{\max}(T'S(W^*)T) \leq -1$, então $\Lambda \in D(B)$, vá para o passo 5. Caso contrário, faça $p \rightarrow p+1$ $x^p = Ty^p$ e retorne ao passo 2.

Passo 5 : Determine $R > 0$ tal que $S(W^*) - BR^{-1}B' < 0$ (R sempre existe - lema 1) e a matriz de realimentação $K = R^{-1}B'(W^*)^{-1}$. Pelo teo.1 $(\Lambda - BK)' \in D(0)$.

Comentários: Este algoritmo é uma generalização daquele proposto no capítulo anterior (sec.3.5.1). Note que $D(0)$ é igual ao conjunto D da def.3.1. Para evitar repetições desnecessárias, vamos nos deter somente nessas generalizações.

1) É fácil verificar que, se $W \in \Gamma$ e $x'S(W)x \leq -1 \forall x \in X_y$, então o mesmo ocorre com $\beta W, \forall \beta \in [1, +\infty]$. Como resultado, a restrição $w_i \geq 1, i = 1, 2, \dots, n$, em (4.3.5) não causa qualquer outro efeito do que bem condicionar o problema. Em adição, a restrição $w_i \leq z, i = 1, 2, \dots, n$ quando z é minimizado, corresponde a minimizar o autovalor máximo de $W \in \Gamma$.

Se $\Lambda \in D(B)$, o algoritmo encontrará $\tilde{W}^* \in \Gamma$ que é tão próxima quanto possível da matriz identidade. Este fato contribui para ter no passo 5 uma matriz de realimentação $K \in R^{m \times n}$ com norma pequena.

2) A prova da convergência do algoritmo segue os mesmos padrões daquela realizada para o teo. de convergência do capítulo anterior.

4.4. Exemplos

Nesta seção, dois exemplos serão apresentados. Eles foram resolvidos pelo algoritmo descrito.

No passo 2, uma versão especial do método dual-simplex foi implementada para trabalhar mais eficientemente o problema de Programação Linear (4.3.5).

Exemplo 1 : Para o sistema dinâmico (4.2.1) considere $n=6$, $m=0$ e a matriz A dada por:

$$A = (-1) \begin{bmatrix} 1.0 & 1.0 & 1.0 & 0.0 & 1.0 & 0.0 \\ 1.0 & 1.0 & 1.0 & 0.0 & 0.0 & 1.0 \\ 1.0 & 1.0 & 1.0 & 1.0 & 0.0 & 0.0 \\ 2.0 & 0.0 & 0.0 & 6.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 9.0 & 0.0 \\ 0.0 & 0.0 & 2.0 & 0.0 & 0.0 & 12. \end{bmatrix}$$

A tab.4.1 fornece os resultados mais importantes de cada iteração.

Concluímos que $A \notin D(0)$.

Iter	z	$\lambda_{\max}(T'S(W)T)$
1	1.000	0.200
2	6.813	0.397
3	14.211	0.716
4	152.511	119.545
5	$+\infty$	Infacível

tab.4.1

Exemplo 2 : Considere o sistema dinâmico (4.2.1) com $n=6$, $m=2$, a matriz A como no exemplo anterior e $B \in \mathbb{R}^{6 \times 2}$, dada por:

$$B = \begin{bmatrix} 0.0 & 1.0 \\ 0.5 & 0.5 \\ 1.0 & 1.0 \\ 0.0 & 0.0 \\ -1.0 & 0.0 \\ 0.0 & 0.0 \end{bmatrix}$$

A tab. 4.2 fornece os resultados mais importantes para a convergência do algoritmo. O parâmetro de parada que satisfaz $\lambda_{\max}(T'S(W)T) < -1 + \epsilon$ foi tomado como $\epsilon = 10^{-3}$. Concluímos então que $A \in D(B)$.

Iter	α	$\lambda_{\max}(T'S(W)T)$
1	1.000	-0.316
2	1.833	-0.969
3	1.838	-0.986
4	1.851	-0.995
5	1.853	-0.998
6	1.854	-0.999

tab.4.2

A matriz $W \in \mathbb{T}$ foi encontrada ser:

$$W^* = \text{diag}(1.854, 1.854, 1.000, 1.854, 1.000, 1.515).$$

No passo 5, a matriz R foi determinada.

$$R = \begin{bmatrix} 1.125 & 0.625 \\ 0.625 & 1.125 \end{bmatrix}$$

Finalmente, a matriz ganho de realimentação $K \in \mathbb{R}^{2 \times 6}$ é dada por

$$K = \begin{bmatrix} -0.385 & 0.154 & 0.571 & 0.000 & -1.286 & 0.000 \\ 0.693 & 0.154 & 0.571 & 0.000 & 0.714 & 0.000 \end{bmatrix}$$

e pelo teorema 4.1, temos $(A - BK) \in D(0)$.

Estes exemplos e outros que foram resolvidos pelo algoritmo proposto aqui, permitem concluir que o número de iterações necessárias para atingir a solução parece ser aproximadamente linear dependendo da diferença entre o número de variáveis de estados e controles ($n-m$).

Em adição, em muitos problemas práticos, é possível mudar a matriz B mudando a localização dos atuadores do sistema em consideração.

Para tais sistemas, o algoritmo pode ser usado para determinar (por tentativa e erro) a matriz B tal que o sistema de malha fechada pertença a $D(0)$, resultando que a lei de controle de realimentação calculada, no passo 5, é tal que o sistema de malha fechada será ótimo com respeito ao critério quadrático (4.2.16) e robusto com respeito a perturbações não-lineares dependentes do estado.

Conclusões Gerais

Este trabalho foi desenvolvido com o propósito de abordar algumas questões referentes ao estudo do problema de autovalores e autovetores de uma matriz real simétrica, otimização de funções matriciais e robustez de sistemas dinâmicos.

Sobre o problema de autovalores de uma matriz real simétrica nossa contribuição refere-se a duas questões existentes neste importante ramo da análise numérica.

Com relação a primeira, sobre a determinação do espectro completo, apresentamos uma nova forma de calcular os autovetores baseada em transformações ortogonais que são numericamente simples e estáveis. Isto completa um algoritmo já existente, devido a Givens e Householder, para determinação de todos os autovalores de uma matriz simétrica, de forma individualizada. Quanto a segunda, para resolver um problema especial de calcular o autovalor mínimo ou máximo e seu respectivo autovetor, apresentamos dois algoritmos que combinam o método de Newton e gradientes conjugados, e que minimizam ou maximizam a função quociente de Rayleigh. A existência desses algoritmos justifica-se pela necessidade de calcular somente um autovalor e seu respectivo autovetor.

No capítulo sobre otimização matricial contribuimos propondo uma metodologia baseada no método de hiperplanos de corte para resolver uma classe de problemas cujos elementos variáveis estão localizados na diagonal principal da matriz incognita. Foi elaborada uma regra para determinar a expressão do vetor que caracteriza o corte e dois problemas de aplicações foram resolvidos: o problema do Teste

Educacional, que aparece em estatística, e a determinação da solução diagonal positiva da equação de Lyapunov que surge no estudo da estabilidade de sistemas dinâmicos. De acordo com os resultados obtidos nos exemplos numéricos essa metodologia além de simples mostrou-se muito eficiente para o tratamento de problemas de otimização com restrições do tipo matricial, semidefinidas.

No último capítulo um procedimento de síntese de controle foi fornecido, isto é, para um dado sistema dinâmico linear contínuo no tempo determinamos uma matriz constante de realimentação, tal que a matriz do sistema de malha fechada admite uma solução diagonal positiva da equação de Lyapunov.

Sob a mesma condição, foi mostrado que é possível definir uma equação de Riccati que também admite uma solução diagonal positiva. Como resultado, o sistema de malha fechada é robusto com respeito a uma larga classe de perturbações não lineares, no sentido de permanecer assintoticamente estável.

Um procedimento numérico para síntese automática da lei de controle de realimentação foi também fornecido.

Considerando que questões muito recentes foram aqui tratadas, segundo um ponto de vista novo, acreditamos ter contribuído de alguma forma para o desenvolvimento dessa linha de pesquisa.

Para dar continuidade a este trabalho, podemos enumerar alguns problemas que podem ser focalizados seguindo a mesma linha aqui proposta:

a - Análise estrutural da matriz hessiana, que aparece nos métodos de otimização que usam condições de segunda ordem, através dos elementos de seu autosistema.

b - Aplicação dos algoritmos do capítulo dois aos problemas cuja matriz do sistema seja de dimensão grande e esparsa.

c - Aplicação da metodologia proposta no capítulo três para resolver o problema de modificação matricial que aparece nos métodos de otimização de funções sem restrições tipo Newton [15].

Referências

- [1] ATHANS, M.; FALB, P.L. - Optimal Control: An Introduction to the Theory and its Applications. McGraw-Hill. (1966).
- [2] BARNETT, S. - Matrices in Control Theory. Van Nostrand Reinhold Company, London. (1970).
- [3] BARNETT, S.; STOREY, C. - Matrix Methods in Stability Theory. Thomas Nelson and Sons LTD. (1970).
- [4] BARKER, G.P.; BERMAN, A.; PLEMMON, R.J. - "Positive diagonal solutions to the Lyapunov equations." Linear and Multilinear Algebra, vol.5, 249-256 (1978).
- [5] BARTH, W.; MARTIN, R.S.; WILKINSON, J.H. - "Calculation of the eigenvalues of a symmetric tridiagonal matrix by the method of bisection." Numerische Mathematik. 9, 386-393 (1967).
- [6] BELLMAN, R. - Introduction to Matrix Analysis. McGraw Hill. (1970).
- [7] BERNSTEIN, H.J. - "An accelerated bisection method for the calculation of eigenvalues of a symmetric tridiagonal matrix." Numerische Mathematik. 43, 153-160 (1984).

- [8] CASTRUCCI, P.B.L.; CURTI, R. - Sistemas Não-Lineares. Edgard Blucher Ltda. Vol.2. (1981).
- [9] DANILIN, Y.M.; PSHENICHNY, B.N. - "Estimates of the rate of convergence of a class of minimization algorithms." Soviet Math. Dokl. vol.14, № 6 (1973).
- [10] DANILIN, Y.M. - "Conjugate gradient methods that do not require the solution of one-dimensional minimization problems." Soviet Math. Dokl. vol.15, № 5 (1974).
- [11] EAVES, B.C.; ZANGWILL, W.I. - "Generalized cutting plane algorithms." SIAM J. Control. Vol.9, № 4, November, (1971).
- [12] EVANS, D.J.; SHANEHCHI, J.; RICK, C.C. - "A modified bisection algorithm for the determination of the eigenvalues of symmetric tridiagonal matrix." Numerische Mathematik. 417-419 (1982).
- [13] FADDEEV, D.K.; FADDEVA, V.N. - Computational Methods of Linear Algebra. (translated by Robert C. Williams) W.H. Freeman and Company (1963).
- [14] FLETCHER, R. - "A nonlinear programming problem in statistic (Educacional Testing)." SIAM J. Sci. Stat. Comput. vol.2, № 3, September, (1981).

- [15] FLETCHER, R. - "Semi-definite matrix constraints in optimization." SIAM J. Control and Optimization. vol. 23, Nº 4, July, (1985).
- [16] FLETCHER, R.; REEVES, C.M. - "Function minimization by conjugate gradients." Computer J., 7 (1964).
- [17] GANTMACHER, F.R. - Theory of Matrices. vol. 1 and 2. Chelsea, New York (1959).
- [18] GEROMEL, J.C. - Controle ótimo Determinístico. Apostila fornecida no mini-curso da IV Escola de Matemática Aplicada, R.J. (1983).
- [19] GEROMEL, J.C. - "On the determination of diagonal solution of the Lyapunov equation." IEEE Transactions on Automatic Control, vol. AC-30, 4, April, (1985).
- [20] GEROMEL, J.C.; YAMAKAMI, A. - "On the robustness of nonlinear regulators and its application to nonlinear systems stabilization." IEEE Transactions Automatic Control, vol. AC-30, 12, December, (1985).
- [21] GEROMEL, J.C.; SANTO, A.O.E. - "On the robustness of linear continuous-time dynamic systems." IEEE Transactions Automatic Control, vol. AC-31, Nº 12, December (1986).

- [22] GILL, P.E.; MURRAY, W.; WHIGHT, M.H. - Practical Optimization. Academic Press. (1981).
- [23] GREENSTAND, J. - "The determination of the characteristic roots of a matrix by the Jacobi method." Mathematical Methods for Digital Computers, vol. 1, Edited by A. Ralston. McGraw-Hill. (1971)
- [24] HESTENES, M.R. - "Iterative method for solving linear equations." Journal of Optimization Theory and Applications. 1, 322-334 (1973).
- [25] HESTENES, M.R. - Conjugate Direction Methods in Optimization. Springer-Verlag Inc., New York (1980).
- [26] KALMAN, R.E.; BERTRAM, J.E. - "Control systems analysis and design via the "second method" of Lyapunov: I continuous-time systems, II discrete-time systems. Stability Analysis." Edited by J.K. Aggarwal and M. Vidyasagar 58-87 (1977).
- [27] KASZKUREWICK, E.; HSU, L. - "Stability of nonlinear systems: A structural approach." Automatica, vol.15 609-614 (1979)
- [28] KHALIL, K.H. - "On the existence of positive diagonal P such that $PA + A'P < 0$." IEEE Transactions on Automatic Control, vol AC-30, 1, February, (1982).

- [29] LA SALLE, J.; LEFSCHETZ, S. - Stability by Lyapunov's Direct Method with Applications. Academic Press (1961).
- [30] LASDON, S.L. - Optimization Theory for Large Systems. Macmillan Publishing Co. Inc., New York. (1970).
- [31] LAUB, A.J. - "Numerical linear algebra aspects of control design computation." IEEE Transactions on Automatic Control. vol. AC-30, N°2, february, (1985).
- [32] LUENBERGER, D.G. - Introduction to Dynamic Systems : Theory, Model & Application. John Wiley & Sons. (1979).
- [33] LUENBERGER, D.G. - Linear and Nonlinear Programing. Addison-Wesley Publishing Company (1984).
- [34] MAN, F.T. - "A theorem on the Lyapunov matrix equation." IEEE Transactions Automatic Control. June, (1969).
- [35] MANSOUR, M. - " A note on the stability of linear discrete systems and Lyapunov method." IEEE Transactions on Automatic Control. vol . AC -27, N° 3, June, (1982).
- [36] NOBLE, B.; DANIEL, J.W. - Algebra Linear Aplicada. (traduzido por J.P. de Carvalho). 2ª ed., Editora Prentice-Hall do Brasil Ltda. (1968).

- [37] ORTEGA, J.M. - "On Sturm sequences for tridiagonal matrices." J. Assoc. Comp. Mach. 7 , 260-263 (1960).
- [38] ORTEGA, J. - "The Givens-Householder method for symmetric matrices." Mathematical Methods for Digital Computers, vol. II, Edited by A.Ralston and H. Wilf. Wiley, New York. (1965) 94-115.
- [39] PERSIDISKII, S.K. - "Problem of absolute stability." Automatic Remote Control, vol. AC-12, 1889-1895 (1969)
- [40] POWELL, M.J.D. - "An efficient method for finding the minimum of a function of several variables without calculating derivatives." Computer J.,7, 155-162 (1964).
- [41] POWER, H.M. -" A note on the matrix Lyapunov matrix equation." IEEE Transactions on Automatic Control. August, (1969).
- [42] PSHENICHYY, B.N.; DANILIN, Y.M. - Numerical Methods in Extremal Problems. Mir Publishers, Moscow. (1978).
- [43] RICHARD, W.C. - "On the convexity of quadratic forms over convex sets." Operations Research, 15, 170-172. (1967).
- [44] SAFANOV, M.; ATHANS, M. - "Gain and phase margin for multiloop LQG regulators. " IEEE Transactions Automatic Control, vol. AC-22 173-179 (1977).

- [45] SANTO, A.O.E.; GEROMEL, J.C. - "Determinação da solução diagonal da equação de Lyapunov." Relatório Interno, FEC, Setor Telemática, UNICAMP, Dezembro. (1984).
- [46] SANTO, A.O.E. - "Um algoritmo para calcular os autovalores e autovetores de uma matriz simétrica." Relatório Interno, FEE, Setor Telemática, UNICAMP, Junho. (1987).
- [47] SANTO, A.O.E. - "Solução de um problema especial de autovalores usando técnicas de otimização." Relatório Interno, FEE, setor Telemática, UNICAMP, Setembro. (1987).
- [48] SCHWARZ, H.R. - Numerical Analysis of Symmetric Matrices. (translated by Paul Hertelenoy) Prentice-Hall Inc. (1973).
- [49] SEZER, M.E.; SILJAK, D.D. - "Robustness of suboptimal control: gain and phase margin." IEEE Transactions Automatic Control vol.AC-12, August (1981).
- [50] STEWART, G.W. - Introduction to Matrix Computations. Academic Press Inc., New York. (1973).
- [51] TAUSSKY, O. - "A remark on theorem of Lyapunov." J. Mathematical Analysis and Applications, 2, 105-107 (1961).
- [52] WESTLAKE, J.R. - A Handbook of Numerical Matrix Inversion and Solution of Linear Equations. John Wiley & Sons Inc., New York. (1968).

- [53] WILKINSON, J.H. - "Calculation of the eigenvalues of a symmetric tridiagonal matrix by the method of bisection." Numerische Mathematik. 4, 362-367 (1962).
- [54] WILKINSON, J.H. - "Calculation of the eigenvalues of a symmetric tridiagonal matrix by inverse iteration." Numerische Mathematik. 4, 368-376 (1962).
- [55] WONG, P.K.; ATHANS, M. "Closed loop structural stability for linear quadratic optimal systems." IEEE Transactions Automatic Control vol. AC-22, 94-99 (1977).
- [56] WONHAM, W.M. - Linear Multivariable Control : A Geometric Approach. Springer-Verlag. (1979).
- [57] YAMAKAMI, A. - Contribuição ao Estudo de Sistemas Dinâmicos Interconectados: Estabilidade e Otimização. Tese de Doutorado, FEC, Setor Telemática. UNICAMP, (1985).
- [58] YAMAKAMI, A.; GEROMEL, J.C. - "Decentralized stabilization and stability region estimation for a class of nonlinear dynamic systems." Proceedings of the 20th IEEE Conference on Decision and Control, San Diego, California, December.1225-1228 (1981).
- [59] YAMAKAMI, A.; GEROMEL, J.C. - "Decentralized stabilization and minimum norm optimization of linear systems." Proceedings of 8th Trienal IFAC, World Congress, Kyoto, Japan.21-26 (1981).