

UNIVERSIDADE FEDERAL DO AMAZONAS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DEPARTAMENTO DE APOIO À PESQUISA
PROGRAMA DE INICIAÇÃO CIENTÍFICA

APLICAÇÃO DA OTIMIZAÇÃO QUADRÁTICA À ANÁLISE DE DADOS
MULTIVARIADOS

Bolsista: Thiago Parente da Silveira, FAPEAM

Manaus - Amazonas
2013

UNIVERSIDADE FEDERAL DO AMAZONAS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DEPARTAMENTO DE APOIO À PESQUISA
PROGRAMA DE INICIAÇÃO CIENTÍFICA

RELATÓRIO FINAL
PIB - E / 0010 / 2012-2013
APLICAÇÃO DA OTIMIZAÇÃO QUADRÁTICA À ANÁLISE DE DADOS
MULTIVARIADOS

Bolsista: Thiago Parente da Silveira, FAPEAM
Orientador: Prof^o Dr. Sandro Dimy Barbosa Bitar

Manaus - Amazonas
2013

Resumo

Neste relatório são apresentados resultados relacionados à Condições de Otimalidade com restrições de igualdade e também as principais propriedades com respeito ao Quociente de Rayleigh. Para o desenvolvimento teórico são explorados e apresentados as definições necessárias para a compreensão de resultados sobre Autovalores e Autovetores, Diagonalização, Diagonalização Ortogonal, Covariância e Matriz de Covariância e Formas Quadráticas.

Os principais Teoremas e Corolários são demonstrados. Após a apresentação dos principais resultados relacionados aos assuntos mencionados, é apresentada uma ferramenta da Análise de Dados Multivariados ADM, a Análise de Componentes Principais.

Abstract

In this report we present the results related to Optimality Conditions with equality constraints and also the main properties with respect to the Quotient Rayleigh. For the theoretical development are explored and presented the definitions necessary for the understanding of results on Eigenvalues and Eigenvectors, Diagonalization, Orthogonal Diagonalization, Covariance and Covariance Matrix and Quadratic Forms.

The main theorems and corollaries are stated. After the presentation of the main results related to the subjects mentioned, is presented a tool Multivariate Data Analysis ADM, Principal Components Analysis.

Conteúdo

1	Introdução	7
2	Desenvolvimento	8
2.1	Apresentação do problema	8
2.2	Autovalores e Autovetores	8
2.3	Diagonalização	9
2.4	Diagonalização Ortogonal	12
3	Covariância e Matriz de Covariância	14
3.1	Tipos de Variáveis	14
3.2	Valor Médio de uma Variável Aleatória	15
3.3	Covariância entre duas Variáveis Aleatórias	16
3.4	Matriz de Covariância	18
4	Condições de Otimalidade	20
4.1	Condições Necessárias de Primeira Ordem	23
4.2	Condições de Segunda Ordem	24
5	Formas Quadráticas	27
5.1	Formas Quadráticas Definidas	29
6	Quociente de Rayleigh	30
7	Análise de Componentes Principais	33
	Conclusão	38
	Referências Bibliográficas	39
	Apêndice	40

Lista de Figuras

4.1	Uma restrição de igualdade em \mathbb{R}^2	20
4.2	Uma restrição de igualdade em \mathbb{R}^3	21
4.3	Plano tangente.	21
4.4	Curvatura das restrições.	26
7.1	Matriz de Correlações extraída do software Tanagra.	35
7.2	Variância explicada por cada componente extraída do software Tanagra.	35
7.3	Correlação entre as variáveis e as componentes extraída do software Tanagra. . .	36
7.4	Projeção dos indivíduos no hiperplano gerado pelas componentes extraído do software Tanagra.	36
7.5	Círculo de Correlações extraído do software Tanagra.	37

Capítulo 1

Introdução

Este relatório expõe assuntos relacionados aos problemas de Otimização Quadrática. O projeto teve como objetivo principal a capacitação com modelos de Otimização Quadrática, com ou sem restrições de factibilidade. Apresentam-se os principais Teoremas e Corolários relacionados ao Quociente de Rayleigh, Autovalores e Autovetores, Diagonalização, Diagonalização Ortogonal, Covariância e Matriz de Covariância e Formas Quadráticas.

Inicialmente discute-se um problema clássico relacionado a conjunto de dados coletados e as possíveis variáveis latentes que poderão ser suficientes para a explicação do fenômeno relacionado ao contexto. Ao longo do relatório, é apresentado o arcabouço teórico necessário para a resolução do problema, com ênfase nas Condições de Otimalidade, onde os principais Teoremas são explanados juntamente com as definições necessárias para entendimento do assunto.

Como aplicação alcançou-se o entendimento de uma das ferramentas da Análise de Dados Multivariados (ADM): a Análise de Componentes Principais (PCA). O projeto foi dividido inicialmente em três etapas. A primeira delas consistiu no amadurecimento teórico para a compreensão de modelos de otimização quadrática; a segunda consistiu na coleta de dados para simulação em softwares para confirmação dos resultados teóricos, e na última tratou-se da interpretação dos resultados obtidos.

Para uso da ferramenta, utilizaram-se as notas das Escolas de Samba do Grupo Especial do Rio de Janeiro de 2013 e aplicamos Componentes Principais para realizar interpretações sobre as notas dos jurados.

O conjunto de dados assim como o ferramental teórico estudado encontram-se neste relatório.

Capítulo 2

Desenvolvimento

2.1 Apresentação do problema

Na análise estatística de dados descritos por variáveis aleatórias x_1, \dots, x_n , é frequentemente importante tentar achar um pequeno conjunto de fatores que explicarão os resultados experimentais. Se as médias das variáveis x_i são, por hipótese, nulas, e se S é a matriz de ordem n chamada matriz de covariância, cujo (i-j)-ésimo elemento s_{ij} mede a média de $x_i x_j$, um método usual é procurar uma nova variável y_1 como combinação linear:

$$y_1 = a_1 x_1 + \dots + a_n x_n$$

das variáveis antigas, de modo que a variância:

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j s_{ij}$$

de y_1 seja tão grande quanto possível, estando sujeita à restrição técnica de que $a_1^2 + \dots + a_n^2 = 1$ (Ben Noble, 1988). Se a representar um vetor coluna com componentes a_1, \dots, a_n então a afirmativa matemática deste problema de estatística se torna:

$$\text{Maximize } \langle a, Sa \rangle = a^T S a \tag{2.1}$$

$$\text{Sujeito à } \langle a, a \rangle = 1$$

Para que se possa encontrar a solução deste problema, serão investigadas as condições de otimalidade para um problema de otimização. A solução pretendida foi estudada usando propriedades da Análise Matemática, Programação Não-Linear, Álgebra Linear e Estatística Multivariada.

2.2 Autovalores e Autovetores

Nesta seção descreve-se sobre autovalores e autovetores objetivando utilizar alguns resultados fundamentais à construção da solução para o problema descrito na Seção 2.1.

Definição 2.2.1 *Se A é uma matriz de ordem n , então um vetor não nulo $x \in \mathbb{R}^n$ é chamado um autovetor de A se Ax é um múltiplo escalar de x , ou seja,*

$$Ax = \lambda x$$

para algum escalar $\lambda \in \mathbb{R}$. O escalar λ é chamado autovalor de A e dizemos que x é um autovetor associado a λ .

Em \mathbb{R}^2 e \mathbb{R}^3 , a multiplicação pelo operador A associa cada autovetor x de A (se houver) a um vetor de mesma direção. Dependendo do sinal e da magnitude do autovalor associado λ a x , o operador linear $Ax = \lambda x$ "comprime" ou "estica" x por um fator λ , invertendo o sentido no caso em que λ for negativo. Para encontrar os autovalores de uma matriz A de ordem n , reescreve-se $Ax = \lambda x$ como:

$$Ax = \lambda Ix,$$

ou equivalentemente,

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

Nesse sentido para λ ser um autovalor é necessário que o sistema acima admita uma solução não nula. Esse sistema tem solução se, e somente se, $\det(A - \lambda I) = 0$ (Anton e Rorres, 2001). Esta equação é a *equação característica* de A ; os escalares que satisfazem esta equação são os autovalores de A . Quando expandido, o determinante $\det(A - \lambda I)$ é um polinômio em λ e é definido como *polinômio característico* de A .

Se A é uma matriz de ordem n , então o polinômio característico de A tem grau n e coeficiente de λ^n é $(-1)^n$ (Anton e Rorres, 2001), ou seja, o polinômio característico de uma matriz de ordem n é da forma:

$$p(\lambda) = (-1)^n \lambda^n + k_2 \lambda^{n-1} + \dots + k_n \lambda^1 + k_{n+1}$$

Pelo Teorema Fundamental da Álgebra segue que esta equação tem no máximo n soluções distintas, de modo que uma matriz de ordem n tem no máximo n autovalores distintos.

Agora que já foi mostrado como encontrar autovalores, passa-se ao problema de encontrar autovetores. Os autovetores de A associados a um autovalor λ são os vetores não nulos que satisfazem $Ax = \lambda x$. Equivalentemente, os autovetores associados a λ são os vetores não nulos no espaço solução de $(A - \lambda I)x = 0$. Chamamos este espaço de *auto-espaço* de A associado a λ . Portanto as soluções não nulas desse sistema são os autovetores de A associados ao autovalor λ .

2.3 Diagonalização

O objetivo desta seção, é mostrar que os dois problemas a seguir são equivalentes.

1. Dada uma matriz A de ordem n , existe uma base de \mathbb{R}^n consistindo de autovetores de A ?
2. Dada uma matriz A de ordem n , existe uma matriz invertível P tal que $P^{-1}AP = D$ onde D é uma matriz diagonal?

Para o problema 2 necessita-se da definição seguinte:

Definição 2.3.1 *Uma matriz quadrada A é dita diagonalizável quando existir uma matriz invertível P tal que $P^{-1}AP = D$ onde D é uma matriz diagonal; dizemos então que P "diagonaliza" A .*

O seguinte teorema mostra que o problema de conseguir uma base de autovetores e o problema da diagonalização são equivalentes.

Teorema 2.3.1 *Se A é uma matriz de ordem n , então as seguintes afirmações são equivalentes:*

- a) A é diagonalizável.
- b) A tem n autovetores linearmente independentes.

Demonstração: " a " \Rightarrow " b ". Como por hipótese A é diagonalizável, então existe uma matriz:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

invertível, tal que $P^{-1}AP$ é diagonal, $P^{-1}AP = D$, onde :

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

Segue da fórmula $P^{-1}AP = D$ que $AP = PD$, ou seja,

$$AP = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1 p_{11} & \lambda_2 p_{12} & \cdots & \lambda_n p_{1n} \\ \lambda_1 p_{21} & \lambda_2 p_{22} & \cdots & \lambda_n p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 p_{n1} & \lambda_2 p_{n2} & \cdots & \lambda_n p_{nn} \end{bmatrix}$$

Se os vetores-coluna de P forem denotados por P_1, P_2, \dots, P_n , se tem que as sucessivas colunas de AP são $\lambda_1 P_1, \lambda_2 P_2, \dots, \lambda_n P_n$. No entanto, as sucessivas colunas de AP podem ser representadas por AP_1, AP_2, \dots, AP_n . Assim:

$$AP_1 = \lambda_1 P_1, AP_2 = \lambda_2 P_2, \dots, AP_n = \lambda_n P_n$$

Como P é invertível, seus vetores-coluna são todos não nulos, assim sendo temos que $\lambda_1, \lambda_2, \dots, \lambda_n$ são autovalores de A e que P_1, P_2, \dots, P_n são os autovetores associados. Como P é invertível, temos que P_1, P_2, \dots, P_n são linearmente independentes (Anton e Rorres, 2001). Portanto, A tem n autovetores linearmente independentes.

" b " \Rightarrow " a ". Suponha que A tem n autovetores linearmente independentes P_1, P_2, \dots, P_n com autovalores associados $\lambda_1, \lambda_2, \dots, \lambda_n$ e seja

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

a matriz cujos vetores-coluna são P_1, P_2, \dots, P_n . Os vetores-coluna de AP são AP_1, AP_2, \dots, AP_n , mas $AP_1 = \lambda_1 P_1, AP_2 = \lambda_2 P_2, \dots, AP_n = \lambda_n P_n$, de modo que:

$$AP = \begin{bmatrix} \lambda_1 p_{11} & \lambda_2 p_{12} & \cdots & \lambda_n p_{1n} \\ \lambda_1 p_{21} & \lambda_2 p_{22} & \cdots & \lambda_n p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 p_{n1} & \lambda_2 p_{n2} & \cdots & \lambda_n p_{nn} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = PD$$

onde D é uma matriz diagonal com autovalores $\lambda_1, \lambda_2, \dots, \lambda_n$ na diagonal principal. Como os vetores-coluna de P são linearmente independentes, P é invertível (Anton e Rorres, 2001), assim sendo pode-se reescrever $AP = PD$ como sendo $P^{-1}AP = D$, ou seja, A é diagonalizável. ■

Teorema 2.3.2 *Se v_1, v_2, \dots, v_k são autovetores de A associados a autovalores distintos $\lambda_1, \lambda_2, \dots, \lambda_k$, então $\{v_1, v_2, \dots, v_k\}$ é um conjunto linearmente independente.*

Demonstração: Sejam v_1, v_2, \dots, v_k autovetores de A associados a autovalores de A distintos $\lambda_1, \lambda_2, \dots, \lambda_k$. Será suposto que v_1, v_2, \dots, v_k são linearmente dependentes e obter uma contradição. Assim pode-se concluir que v_1, v_2, \dots, v_k são linearmente independentes.

Como um autovetor é por definição não nulo, tem-se que $\{v_1\}$ é linearmente independente. Seja r o maior inteiro tal que $\{v_1, v_2, \dots, v_r\}$ é linearmente independente. Como por hipótese $\{v_1, v_2, \dots, v_k\}$ é linearmente dependente, r satisfaz $1 \leq r < k$. Além disto, pela definição de r , $\{v_1, v_2, \dots, v_r, v_{r+1}\}$ é linearmente dependente. Assim, existem escalares $c_1, c_2, \dots, c_r, c_{r+1}$ não todos nulos tais que:

$$c_1v_1 + c_2v_2 + \dots + c_{r+1}v_{r+1} = 0 \quad (2.2)$$

multiplicando ambos os lados de (2.2) por A e usando

$$Av_1 = \lambda_1v_1, Av_2 = \lambda_2v_2, \dots, Av_{r+1} = \lambda_{r+1}v_{r+1}$$

é obtido:

$$c_1\lambda_1v_1 + c_2\lambda_2v_2 + \dots + c_{r+1}\lambda_{r+1}v_{r+1} = 0 \quad (2.3)$$

multiplicando ambos os lados de (2.2) por λ_{r+1} e subtraindo a equação resultante de (2.3), tem-se:

$$c_1(\lambda_1 - \lambda_{r+1})v_1 + c_2(\lambda_2 - \lambda_{r+1})v_2 + \dots + c_r(\lambda_r - \lambda_{r+1})v_r = 0$$

Como $\{v_1, v_2, \dots, v_r\}$ é linearmente independente, esta equação implica que:

$$c_1(\lambda_1 - \lambda_{r+1}) = c_2(\lambda_2 - \lambda_{r+1}) = \dots = c_r(\lambda_r - \lambda_{r+1}) = 0$$

e como $\lambda_1, \lambda_2, \dots, \lambda_k$ são distintos, segue-se que:

$$c_1 = c_2 = \dots = c_r = 0 \quad (2.4)$$

substituindo estes valores em (2.2) teremos:

$$c_{r+1}v_{r+1} = 0$$

como o autovetor v_{r+1} é não-nulo, segue-se que:

$$c_{r+1} = 0 \quad (2.5)$$

As equações (2.4) e (2.5) contradizem o que foi suposto a respeito destas constantes, a saber, que os escalares c_1, c_2, \dots, c_{r+1} não são todos nulos. ■

Teorema 2.3.3 *Se uma matriz A de ordem n tem n autovalores distintos, então A é diagonalizável.*

Demonstração: Se $\{v_1, v_2, \dots, v_n\}$ são autovetores associados aos autovalores distintos $\lambda_1, \lambda_2, \dots, \lambda_n$, então pelo Teorema 2.3.2 o conjunto $\{v_1, v_2, \dots, v_n\}$ é linearmente independente. Assim sendo A é diagonalizável pelo Teorema 2.3.1. ■

2.4 Diagonalização Ortogonal

Nesta seção, será feito um estudo sobre diagonalização ortogonal, tendo em vista o uso dos resultados para a elaboração da resposta ao problema apresentado na primeira seção.

Definição 2.4.1 *Seja A uma matriz de ordem n e P uma matriz de ordem n e ortogonal. Se $P^{-1}AP = P^TAP$ é uma matriz diagonal, então dizemos que A é ortogonalmente diagonalizável e que P diagonaliza A ortogonalmente.*

Para a Definição 2.4.1, tem-se duas questões a considerar:

- a) Quais matrizes são ortogonalmente diagonalizáveis?
- b) Como encontrar uma matriz ortogonal que efetue tal diagonalização?

Com relação à primeira questão, observa-se que não existe maneira de diagonalizar A ortogonalmente a menos que A seja simétrica (ou seja $A = A^T$) (Anton e Rorres, 2001). Para verificação desta afirmação, suponha que:

$$P^TAP = D$$

onde P é uma matriz ortogonal e D é uma matriz diagonal. Como P é ortogonal $PP^T = P^TP = I$ e,

$$A = PDP^T$$

e como D é diagonal, tem-se que $D = D^T$ e transpondo ambos os lados

$$A^T = (PDP^T)^T = (P^T)^T D^T P^T = PDP^T = A$$

e portanto A é simétrica.

O próximo teorema mostra que qualquer matriz simétrica, é de fato, ortogonalmente diagonalizável.

Teorema 2.4.1 *Se A é uma matriz de ordem n , então as seguintes afirmações são equivalentes:*

- a) A é ortogonalmente diagonalizável;
- b) A tem um conjunto ortonormal de n vetores;
- c) A é simétrica.

Demonstração: "a" \Rightarrow "b". Como A é ortogonalmente diagonalizável, existe uma matriz ortogonal P tal que P^TAP é diagonal. Como foi mostrado no Teorema 2.3.1, os n vetores-coluna de P são autovetores de A . Como P é ortogonal, estes vetores-coluna são ortonormais de modo que A tem n autovetores ortonormais.

"b" \Rightarrow "a". Suponha agora que A tenha um conjunto ortonormal de n autovetores $\{p_1, p_2, \dots, p_n\}$. Como foi mostrado na demonstração do Teorema 2.3.1, a matriz P que tem estes autovetores

como colunas, diagonaliza A , como estes vetores são ortonormais, P é ortogonal e, portanto diagonaliza A ortogonalmente.

"a" \Rightarrow "c". Na prova em que "b" \Rightarrow "a", foi mostrado que uma matriz A de ordem n que é ortogonalmente diagonalizável, é ortogonalmente diagonalizável por uma matriz P de ordem n cujas colunas formam um conjunto ortonormal de autovetores de A . Seja D a matriz diagonal tal que:

$$D = P^{-1}AP$$

assim,

$$A = PDP^{-1} \text{ ou } A = PDP^T$$

já que P é ortogonal. Segue-se que,

$$A^T = (PDP^T)^T = PDP^T = A$$

o que mostra que A é simétrica. ■

O próximo teorema, mostra uma característica das matrizes simétricas, com respeito aos autovalores.

Teorema 2.4.2 *Se A é uma matriz simétrica, então:*

- a) *Os autovalores de A são reais;*
- b) *Autovetores associados a auto-espacos diferentes são ortogonais.*

Demonstração: "a". Para a demonstração deste teorema, vamos considerar V um espaço vetorial sobre \mathbb{F} , onde $\mathbb{F} = \mathbb{R}$ ou $\mathbb{F} = \mathbb{C}$. Assim sendo, o produto interno é uma aplicação $\varphi : V \times V \rightarrow \mathbb{F}$ que satisfaz:

1. $\varphi(\alpha, \alpha) > 0; \alpha \neq 0$
2. $\varphi(\alpha, \beta) = \overline{\varphi(\beta, \alpha)}$
3. $\varphi(\alpha, \beta + \gamma) = \varphi(\alpha, \beta) + \varphi(\alpha, \gamma)$
4. $\varphi(\lambda\alpha, \beta) = \lambda\varphi(\alpha, \beta)$

Agora considere $Av = \lambda v$, onde $\lambda \in \mathbb{C}$ e A é simétrica. Usando a seguinte expressão, tem-se:

$$\lambda \langle v, v \rangle = \langle \lambda v, v \rangle = \langle Av, v \rangle = \langle v, A^T v \rangle = \langle v, Av \rangle = \langle v, \lambda v \rangle = \overline{\langle \lambda v, v \rangle} = \overline{\lambda \langle v, v \rangle} = \bar{\lambda} \langle v, v \rangle$$

Como um autovetor é não-nulo, tem-se que $\langle v, v \rangle > 0$ e $\lambda \langle v, v \rangle = \bar{\lambda} \langle v, v \rangle$ implica $\lambda = \bar{\lambda}$ o que acontece se, e somente se $\lambda \in \mathbb{R}$.

"b". Sejam v_1, v_2 associados a autovalores distintos λ_1, λ_2 da matriz A , respectivamente. O objetivo é mostrar que $\langle v_1, v_2 \rangle = 0$. A prova disto inclui a vantagem de começar com a expressão $\langle Av_1, v_2 \rangle$, usando a simetria de A tem-se:

$$\langle Av_1, v_2 \rangle = \langle v_1, Av_2 \rangle$$

como v_1 é autovetor de A associado a λ_1 e v_2 é autovetor de A associado a λ_2 , segue-se que a relação:

$$\langle \lambda_1 v_1, v_2 \rangle = \langle v_1, \lambda_2 v_2 \rangle$$

que pode ser reescrita como:

$$(\lambda_1 - \lambda_2) \langle v_1, v_2 \rangle = 0$$

como por hipótese λ_1 e λ_2 são distintos, temos que $\lambda_1 - \lambda_2 \neq 0$ e portanto $\langle v_1, v_2 \rangle = 0$. ■

Capítulo 3

Covariância e Matriz de Covariância

Neste capítulo será feito um estudo sobre Variância e Matriz de Covariância, com o intuito de melhor compreensão para a resolução de problemas que envolvem análise de dados.

3.1 Tipos de Variáveis

De modo geral, para cada elemento investigado numa pesquisa, tem-se associado um (ou mais de um) resultado correspondendo à realização de uma característica (ou características). Algumas variáveis, como estado civil, grau de escolaridade e nacionalidade, apresentam como possíveis realizações uma quantidade (ou atributo) do indivíduo pesquisado, ao passo que outras, como número de filhos, salário, idade, apresentam como possíveis realizações números resultantes de uma contagem ou mensuração. As variáveis do primeiro tipo são chamadas *qualitativas*, e as do segundo tipo, *quantitativas*.

Dentre as variáveis qualitativas, ainda pode-se fazer uma distinção entre dois tipos: variável qualitativa *nominal*, para qual não existe nenhuma ordenação nas possíveis realizações, e variável qualitativa *ordinal*, para o qual existe uma ordem para os resultados. A nacionalidade por exemplo, é um caso de variável qualitativa nominal, enquanto que grau de escolaridade é um caso de variável qualitativa ordinal, visto que ensinos fundamental, médio e superior correspondem a uma ordenação baseada no número de anos de escolaridade completos.

De modo análogo, as variáveis quantitativas podem sofrer uma classificação:

- a) Variáveis quantitativas *discretas*, cujos possíveis valores formam um conjunto finito ou enumerável de números, e que resultam, frequentemente, de uma contagem, como por exemplo o número de filhos $(0, 1, 2, \dots)$;
- b) Variáveis quantitativas *contínuas*, cujos possíveis valores pertencem a um intervalo de números reais e que resultam de uma mensuração, como por exemplo estatura e massa de um indivíduo.

A distribuição de frequências é um instrumento importante para avaliarmos a variabilidade das observações de um fenômeno aleatório. A partir dessas frequências observadas podemos calcular medidas de posição e variabilidade, como média, variância, e etc. Essas frequências e medidas calculadas a partir dos dados são *estimativas* de quantidades desconhecidas, associadas em geral a populações das quais os dados foram extraídos na forma de *amostras*. Em particular as frequências são estimativas de *probabilidades* de ocorrências de certos eventos de interesse.

Com suposições adequadas, e sem observarmos diretamente o fenômeno aleatório de interesse, podemos criar um *modelo teórico* que reproduza de maneira razoável a distribuição das frequências, quando o fenômeno é observado diretamente. Tais modelos são chamados *modelos probabilísticos*.

Definição 3.1.1 Para cada experimento ε , o espaço amostral é o conjunto de todos os resultados possíveis para ε que será denotado por Ω

Definição 3.1.2 Um evento é um subconjunto do espaço amostral Ω . Isto significa dizer que o próprio espaço amostral constitui um evento, bem como o é o conjunto vazio \emptyset . Qualquer resultado individual também pode ser tomado como um evento.

Definição 3.1.3 Dois eventos A e B são denominados mutuamente excludentes, se eles não puderem ocorrer juntos, ou seja, $A \cap B = \emptyset$.

Definição 3.1.4 Seja ε um experimento. Seja Ω o espaço amostral associado à ε . A cada evento ω será associado um número real representado por $P(\omega)$ e denominado probabilidade de ω , que satisfaça as seguintes propriedades:

1. $0 \leq P(\omega) \leq 1$
2. $P(\Omega) = 1$
3. Se $\omega_1, \omega_2, \dots, \omega_n, \dots$ forem, dois a dois, eventos mutuamente excludentes, então:

$$P(\cup_{i=1}^{\infty} \omega_i) = \sum P(\omega_i)$$

De modo geral, se A for qualquer evento de Ω , então

$$P(A) = \sum_{\omega_j \in A} P(\omega_j),$$

Definição 3.1.5 Chama-se função de probabilidade da variável aleatória discreta X , que assume os valores $x_1, x_2, \dots, x_n, \dots$ a função $\{(x_i, p(x_i)), i = 1, 2, \dots\}$, que a cada valor de x_i associa a sua probabilidade de ocorrência, isto é,

$$p(x_i) = P(X = x_i) = p_i, i = 1, 2, \dots$$

3.2 Valor Médio de uma Variável Aleatória

Será introduzido agora o conceito de valor médio, variância e desvio padrão.

Definição 3.2.1 Dada uma variável aleatória X discreta, assumindo os valores x_1, x_2, \dots, x_n , chamamos o valor médio ou esperança matemática de X ao valor:

$$E(X) = \sum_{i=1}^n x_i P(X = x_i) = \sum_{i=1}^n x_i p_i \quad (3.1)$$

Definição 3.2.2 Chama-se de variância da variável aleatória X o valor:

$$Var(X) = \sum_{i=1}^n [x_i - E(X)]^2 p_i$$

O desvio padrão de X , $dp(X)$, é definido como o módulo da raiz quadrada da variância.

Uma variável aleatória X pode assumir um número infinito de valores, porém enumerável, de valores $x_1, x_2, \dots, x_n, \dots$ com probabilidades $p_1, p_2, \dots, p_n, \dots$ tal que cada $p_i > 0$ e a soma de todos os p_i seja 1. Nesse caso, a definição de esperança deve ser modificada. A soma definida anteriormente é uma série, que tem que ser convergente.

Definição 3.2.3 *Dada a variável aleatória discreta X e a respectiva função de probabilidade $p(x)$, a esperança matemática da função $h(x)$ é dada por:*

$$E[h(X)] = \sum h(x_i)p(x_i)$$

Segue agora algumas propriedades:

a) Se $h(X) = aX + b$, onde a e b são constantes, então:

$$\begin{aligned} E(aX + b) &= aE(X) + b, \\ \text{Var}(aX + b) &= a^2\text{Var}(X). \end{aligned}$$

b)

$$\text{Var}(X) = E(X^2) - [E(X)]^2 \tag{3.2}$$

Definição 3.2.4 *Uma função X definida sobre um espaço amostral Ω e assumindo valores num intervalo de números reais, é dita variável aleatória contínua.*

Deduz-se intuitivamente que qualquer função $f(x)$, não-negativa tal que:

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

define uma variável aleatória contínua X , ou seja, cria um modelo teórico para as frequências relativas de uma variável aleatória contínua. A área compreendida entre dois valores a e b , da abscissa x , sob a curva representativa de $f(x)$, dá a probabilidade (proporção teórica) da variável pertencer ao intervalo limitado pelos dois valores. Usando o conceito de integral, podemos escrever:

$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$

Definição 3.2.5 *Dada uma variável aleatória X contínua, chamamos o valor médio ou esperança matemática de X ao valor:*

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx.$$

onde a função $f(x)$ é chamada função de densidade de probabilidade (f.d.p.) da variável aleatória X

3.3 Covariância entre duas Variáveis Aleatórias

Será introduzido agora uma medida de relação linear entre duas variáveis aleatórias.

Definição 3.3.1 *Se X e Y são duas variáveis aleatórias, a "covariância" entre elas é definida por:*

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))],$$

ou seja, o valor médio do produto dos desvios de X e Y em relação às suas respectivas médias.

Suponha que X assuma os valores x_1, x_2, \dots, x_n , e Y os valores y_1, y_2, \dots, y_m e que $P(X = x_i, Y = y_j) = p(x_i, y_j)$. Então a Definição 3.3.1 pode ser escrita como:

$$\text{Cov}(X, Y) = \sum_{i=1}^n \sum_{j=1}^m [x_i - E(X)][y_j - E(Y)]p(x_i, y_j)$$

ou então de maneira mais simples. Note que:

$$\begin{aligned} \text{Cov}(X, Y) &= E[XY - XE(Y) - YE(X) + E(X)E(Y)] \\ \text{Cov}(X, Y) &= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) \end{aligned}$$

ou seja,

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) \quad (3.3)$$

Definição 3.3.2 Quando $\text{Cov}(X, Y) = 0$, dizemos que as variáveis aleatórias X e Y são não correlacionadas.

Proposição 3.3.1 Se X e Y são duas variáveis aleatórias independentes, então $\text{Cov}(X, Y) = 0$.

Demonstração: De fato, se X e Y são duas variáveis aleatórias independentes, então $E(XY) = E(X)E(Y)$, então $\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$. ■

Teorema 3.3.1 Sejam X e Y duas variáveis aleatórias quaisquer, então:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

Demonstração: $\text{Var}(X + Y) = E[(X + Y) - E(X + Y)]^2$
 $\text{Var}(X + Y) = E[X - E(X) + Y - E(Y)]^2 = E[X - E(X)]^2 + E[Y - E(Y)]^2 + 2E[(X - E(X))(Y - E(Y))]$
 $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$ ■

Corolário: Se X e Y forem independentes, então:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Demonstração: De fato, para X e Y duas variáveis aleatórias quaisquer, então:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y),$$

como X e Y são independentes, tem-se:

$$\text{Cov}(X, Y) = 0.$$

Portanto $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$. ■

Definição 3.3.3 O coeficiente de relação entre X e Y é definido por:

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma(X)\sigma(Y)}$$

onde $\sigma(X)$ é o desvio padrão de X e $\sigma(Y)$ é o desvio padrão de Y .

Teorema 3.3.2 O coeficiente de correlação entre X e Y satisfaz a seguinte desigualdade:

$$-1 \leq \rho(X, Y) \leq 1.$$

Demonstração: Considere a seguinte função da variável real t :

$$q(t) = E[V + tW]^2,$$

onde

$$V = X - E[X] \text{ e } W = Y - E[Y].$$

Visto que $[V + tW]^2 \geq 0$, temos que $q(t) \geq 0$ para todo t . Desenvolvendo obteremos

$$q(t) = E[V^2 + 2tVW + t^2W^2] = E(V^2) + 2tE(VW) + t^2E(W^2).$$

Deste modo, $q(t)$ é uma expressão quadrática de t . Em geral, se uma expressão quadrática $q(t) = at^2 + bt + c$ tem a propriedade de que $q(t) \geq 0$ para todo t , isto significa que seu gráfico toca o eixo t em apenas um ponto, ou então não toca. Isto por sua vez, significa que seu discriminante $b^2 - 4ac$ deve ser ≤ 0 , porque $b^2 - 4ac > 0$ significaria que $q(t)$ teria duas raízes reais *distintas*. Aplicando-se a essa conclusão à função $q(t)$, obtem-se

$$4[E(VW)]^2 - 4E(V^2)E(W^2) \leq 0.$$

Isto acarreta

$$\frac{[E(VW)]^2}{E(V^2)E(W^2)} \leq 1, \text{ e daí } \frac{\{E[X - E(X)][Y - E(Y)]\}^2}{Var(X)Var(Y)} = \rho^2 \leq 1.$$

Portanto, $-1 \leq \rho(X, Y) \leq 1$

O coeficiente de correlação é uma medida de relação *linear* entre X e Y . Quando $\rho(X, Y) = \pm 1$, existe uma correlação perfeita entre X e Y , pois $Y = aX + b$. Se $\rho(X, Y) = 1$, então $a > 0$, e se $\rho(X, Y) = -1$ então $a < 0$. O grau de associação linear entre X e Y varia a medida que $\rho(X, Y)$ varia entre -1 e $+1$.

3.4 Matriz de Covariância

A Matriz de Covariância das amostras $S = (s_{ij})$ é a matriz:

$$\begin{bmatrix} Cov(X_1, X_1) & Cov(X_1, X_2) & \dots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & Cov(X_2, X_2) & \dots & Cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \dots & Cov(X_n, X_n) \end{bmatrix}$$

Serão dadas duas maneiras de se obter S . A primeira delas é simplesmente calcular cada entrada da matriz s_{ij} individualmente. A variância da j -ésima variável $s_{ii} = s_i^2$ pode ser calculada da seguinte maneira, usando a j -ésima coluna da matriz de dados Y :

$$s_{ii} = s_i^2 = \frac{1}{n-1} \sum_{i=1}^n (y_{ij} - \bar{y}_j)^2 \quad (3.4)$$

$$s_{ii} = s_i^2 = \frac{1}{n-1} \left(\sum_i y_{ij}^2 - n\bar{y}_j^2 \right) \quad (3.5)$$

onde \bar{y}_j é a média da j -ésima variável. A covariância entre a j -ésima e a k -ésima variáveis, s_{jk} também pode ser calculada como segue, usando a j -ésima e a k -ésima colunas de Y :

$$s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (y_{ij} - \bar{y}_j)(y_{ik} - \bar{y}_k) \quad (3.6)$$

$$s_{jk} = \frac{1}{n-1} \left(\sum_i y_{ij}y_{ik} - n\bar{y}_j\bar{y}_k \right). \quad (3.7)$$

Note que em (3.4) a variância s_{jj} é expressa como s_j^2 , o quadrado do desvio padrão s_j e que S é simétrica porque $s_{jk} = s_{kj}$. Outros nomes são usados para a matriz de covariância, são eles: *matriz de variância*, *matriz de variância-covariância* e *matriz de dispersão*.

A matriz de covariância das amostras S também pode ser expressa por meio dos vetores da observação:

$$S = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})(y_i - \bar{y})^t \quad (3.8)$$

$$S = \frac{1}{n-1} \left(\sum_{i=1}^n y_i y_i^T - n\bar{y}\bar{y}^T \right). \quad (3.9)$$

Desde $(y_i - \bar{y})^T = (y_{i1} - \bar{y}_1, y_{i2} - \bar{y}_2, \dots, y_{ip} - \bar{y}_p)$, o elemento na posição (1, 1) de $(y_i - \bar{y})(y_i - \bar{y})^T$ é $(y_{i1} - \bar{y}_1)^2$ e quando é somado sobre i em (3.8) o resultado é o numerador de s_{11} em 3.4. Similarmente, o elemento da posição (1, 2) de $(y_i - \bar{y})(y_i - \bar{y})^T$ é $(y_{i1} - \bar{y}_1)(y_{i2} - \bar{y}_2)$, que somado ao numerador de s_{12} em (10). Então (3.8) é equivalente a (3.4) e (3.6) e também (3.9) produz (3.5) e (3.7) (Rencher, 2002).

Capítulo 4

Condições de Otimalidade

Neste capítulo serão investigadas as principais condições de otimalidade, com o intuito de investigar as soluções de problemas da seguinte forma:

$$\begin{aligned} &\text{Minimizar } f(x) \\ &\text{Sujeita a } x \in S \end{aligned}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e $S \subset \mathbb{R}^n$, S é chamado *conjunto factível* e este problema é a forma genérica dos problemas de otimização. Serão considerados dois tipos de soluções deste problema:

Definição 4.0.1 Um ponto $x^* \in S$ é um *minimizador local* de f em S se, e somente se existe $\varepsilon > 0$ tal que $f(x) \geq f(x^*)$ para todo $x \in S$ tal que $\|x - x^*\| < \varepsilon$. Se $f(x) > f(x^*)$ para todo $x \in S$ tal que $x \neq x^*$ e $\|x - x^*\| < \varepsilon$ diremos que se trata de um *minimizador local estrito* em S . Um ponto $x^* \in S$ é um *minimizador global* de f em S se, e somente se $f(x) \geq f(x^*)$ para todo $x \in S$. Se $f(x) > f(x^*)$ para todo $x \in S$ tal que $x \neq x^*$, diremos que se trata de um *minimizador global estrito* em S . De forma análoga são definidos *maximizadores locais e globais*.

Observe que "Maximizar f " é equivalente a "Minimizar $-f$ " (Friedlander, 1994), a razão pela qual pode-se sem perda de generalidade, falar apenas de "Minimização" ao longo desta seção.

Considere o seguinte problema:

$$\text{Minimizar } f(x) \tag{4.1}$$

$$\text{Sujeita a } h(x) = 0$$

onde $f, h \in C^1$, $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $m < n$ e a função h é não-linear. O estudo da região de factibilidade no caso de restrições lineares é feito utilizando exclusivamente os conceitos da Álgebra Linear. A presença de restrições não-lineares exige o uso de conceitos mais complicados. Na Figura 4.1 mostra-se o caso em que há uma única restrição de igualdade em \mathbb{R}^2 .

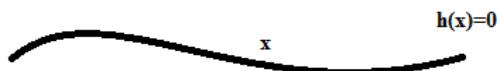


Figura 4.1: Uma restrição de igualdade em \mathbb{R}^2 .
Fonte: Friedlander, 1994.

Neste caso a região factível $S \equiv \{x \in \mathbb{R}^2/h(x) = 0\}$ é uma curva. Para permanecer em S , os movimentos a partir de $x \in S$ devem ser curvilíneos. É claro que dado um ponto $x \in S$ a dificuldade em determinar outro ponto $\bar{x} \in S$ depende da função h .

Na Figura 4.2, ilustra-se o caso em que há uma restrição de igualdade em \mathbb{R}^3 . A região $S \equiv \{x \in \mathbb{R}^3/h(x) = 0\}$ é uma superfície em \mathbb{R}^3 .

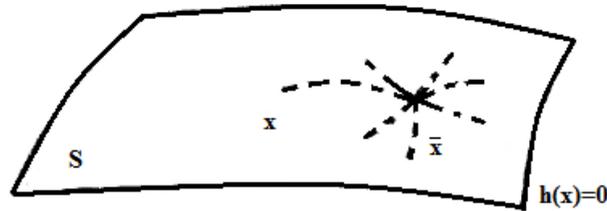


Figura 4.2: Uma restrição de igualdade em \mathbb{R}^3 .
Fonte: Friedlander, 1994.

Dado um ponto $\bar{x} \in S$, observa-se que há infinitos caminhos que passam por \bar{x} contidos em S , todos eles curvilíneos. Podem-se unir dois pontos x e \bar{x} por arcos de curva contidos em S , que chamamos de *arcos factíveis*. Em geral, m equações (não-lineares) em \mathbb{R}^n , $m < n$, determinam uma superfície de dimensão " $n - m$ " (Friedlander, 1994).

Na Figura 4.2, note que por um ponto $\bar{x} \in S$ passa uma família de curvas contidas em S . Cada curva é uma superfície de dimensão 1 e sob certas condições é possível expressar esta curva, numa vizinhança do ponto \bar{x} , mediante uma parametrização contínua $x : \mathbb{R} \rightarrow S$, tal que $x(t) \in S$ para todo $t \in (a, b)$ e, ademais existe $\bar{t} \in (a, b)$ tal que $x(\bar{t}) = \bar{x}$ (Friedlander, 1994).

O arco $x(t)$, assim definido, é diferenciável se $x'(t)$ existe para todo $t \in (a, b)$. Por exemplo, para $x : \mathbb{R} \rightarrow \mathbb{R}^2$,

$$x(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} \text{ e } x'(t) = \begin{pmatrix} x'_1(t) \\ x'_2(t) \end{pmatrix}$$

O vetor $x'(t)$ é tangente ao arco da curva no ponto \bar{x} , portanto tangente à superfície. Se consideramos todas as curvas diferenciáveis que passam por \bar{x} , intuitivamente observa-se que seus vetores tangentes em \bar{x} definem um "plano tangente" T , como ilustra a Figura 4.3:

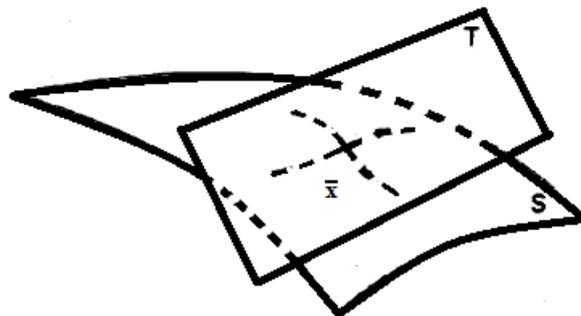


Figura 4.3: Plano tangente.
Fonte: Friedlander, 1994.

Se a superfície S for $(n - m)$ -dimensional em \mathbb{R}^n , generalizando essas ideias, nota-se que o plano tangente será gerado por vetores da forma:

$$x'(t) = (x'_1(t), x'_2(t), \dots, x'_n(t))^T$$

A dimensão do plano também é $n - m$ (Friedlander, 1994).

Definição 4.0.2 *O plano tangente a uma superfície $S \subseteq \mathbb{R}^n$ em um ponto $\bar{x} \in \mathbb{R}^n$ é o conjunto de vetores do \mathbb{R}^n , que são tangentes em \bar{x} a alguma curva diferenciável contida em S e que passa por \bar{x} .*

Agora se tornou possível a caracterização dos arcos factíveis que passam por um ponto factível \bar{x} .

Se: $S = \{x \in \mathbb{R}^n / h(x) = 0\}$ e $x(t) : (a, b) \rightarrow S$ é a parametrização de um arco factível, tem-se:

$$h(x(t)) = 0, \forall t \in (a, b),$$

derivando a equação acima em relação a t :

$$J_h(x(t))x'(t) = 0, \forall t \in (a, b) \quad (4.2)$$

ou seja,

$$\nabla^T h_i(x(t))x'(t) = 0, \forall t \in (a, b) \quad (4.3)$$

em particular, para $x(\bar{t}) = \bar{x}$:

$$\nabla^T h_i(\bar{x})x'(\bar{t}) = 0, 1 \leq i \leq m,$$

O que significa que dado um arco factível diferenciável, é necessário que o vetor tangente ao arco em \bar{x} seja ortogonal ao gradiente das restrições avaliados em \bar{x} .

Para que (4.2) seja uma caracterização dos arcos factíveis diferenciáveis, é necessário que para todo $p \in \mathbb{R}^n / J_h(\bar{x})p = 0$ exista um arco factível diferenciável $z(t)$ tal que $z'(t) = p$.

Infelizmente isso nem sempre acontece, como mostra o seguinte exemplo devido a Kuhn e Tucker.

$$h_1(x) = (1 - x_1)^3 - x_2, h_2(x) = x_2 \\ S = \{x \in \mathbb{R}^2 / h_1(x) = 0 \text{ e } h_2(x) = 0\} = \{(1, 0)^T\},$$

como S consiste em um único ponto, não existe arcos factíveis, mas:

$$\nabla^T h_i(\bar{x}) = (0, -1),$$

e portanto, para todo vetor $p \in \mathbb{R}^2$ da forma $(\gamma, 0)^T$, $\gamma \in \mathbb{R}$ verifica:

$$J_h(\bar{x})p = 0.$$

Assim, caracterizar o conjunto dos arcos factíveis diferenciáveis através do plano tangente mediante a equação (4.2) não é possível sem alguma hipótese adicional. A mais simples é a regularidade do ponto \bar{x} em relação às restrições (Friedlander, 1994).

Definição 4.0.3 *Um ponto \bar{x} que satisfaz as equações $h(x) = 0$ é regular em relação às restrições se o conjunto de vetores $\{\nabla h_1(\bar{x}), \nabla h_2(\bar{x}), \dots, \nabla h_m(\bar{x})\}$ for linearmente independente.*

Com esta hipótese sobre \bar{x} é possível caracterizar os arcos factíveis diferenciáveis.

Teorema 4.0.1 *Seja x^* um ponto regular da superfície S definida por $S \equiv \{x \in \mathbb{R}^n / h(x) = 0\}$, o plano tangente é igual a:*

$$M \equiv \{y \in \mathbb{R}^n / J_h(x^*)y = 0\} \quad (4.4)$$

Demonstração: Seja T um plano tangente em x^* . É claro que $T \subset M$ se x^* é regular ou não, para qualquer curva $x(t)$ passando por x^* em $t = t^*$ existindo derivada $x'(t^*)$ tal que $\nabla h(x^*)x'(t^*) = 0$ (Friedlander, 1994).

Para provar que $M \subset T$, é necessário e suficiente mostrar que, se $y \in M$ existe uma curva em S passando por x^* com derivada y . Para construir tal curva, vamos considerar as equações:

$$h(x^* + ty + \nabla h(x^*)^T u(t)) = 0 \quad (4.5)$$

onde t é fixo e $u(t)$ desconhecido.

Este é um sistema não-linear de m equações e m incógnitas, continuamente parametrizadas por t . Em $t = 0$, note que $u(0) = 0$ é uma solução para $h(x) = 0$. A matriz do Jacobiano do sistema com respeito a " u " em $t = 0$ é uma matriz de ordem m :

$$\nabla h(x^*) \nabla h(x^*)^T$$

que é não-singular (Friedlander, 1994), desde que $\nabla h(x^*)$ seja de posto completo se x^* é um ponto regular. Assim, pelo Teorema da Função Implícita (Elon Lima, 2012), é continuamente diferenciável a solução $u(t)$ em uma região $-a \leq t \leq a$.

A curva $x(t) = x^* + tx + \nabla h(x^*)^T u(t)$ é assim, por construção uma curva em S . Diferenciando o sistema (4.5) com respeito a t em $t = 0$:

$$0 = \left. \frac{dh(x(t))}{dt} \right|_{t=0} = \nabla h(x^*)y + \nabla h(x^*) \nabla h(x^*)^T u(0)$$

Por definição de y temos $\nabla h(x^*)y = 0$ e assim, novamente Desde que $\nabla h(x^*) \nabla h(x^*)^T$ seja não-singular, concluímos que $u(0) = 0$. Portanto:

$$x'(0) = y + \nabla h(x^*)x(0) = y$$

e assim sendo, a curva construída tem derivada y em x^* . ■

4.1 Condições Necessárias de Primeira Ordem

Teorema 4.1.1 *Seja x^* um minimizador local. Suponha que x^* seja um ponto regular das restrições. Então existe $\lambda^* \in \mathbb{R}^m$ tal que:*

$$\nabla f(x^*) = \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*)$$

ou, equivalentemente

$$Z^T(x^*) \nabla f(x^*) = 0$$

onde $Z(x^*) \in \mathbb{R}^{n(n-m)}$ e suas colunas formam uma base de $Nu(J_h(x^*))$

Demonstração: Seja x^* um ponto regular de $S \equiv \{x \in \mathbb{R}^n / h(x) = 0\}$ minimizador local de (4.1). Então, para qualquer parametrização:

$$x : (a, b) \longrightarrow S, x(t^*) = x^*, t^* \in (a, b)$$

temos que t^* é solução do problema (Friedlander, 1994).

$$\text{Minimizar : } \varphi(x) = f(x(t)) \quad (4.6)$$

$$t \in (a, b)$$

A condição necessária de primeira ordem para (4.6) é $\varphi'(x^*) = 0$, portanto:

$$\varphi'(t^*) = J_f(x(t^*))x'(t^*) = \nabla^T f(x^*)x'(t^*) = 0 \quad (4.7)$$

ou seja, $\nabla f(x^*)$ deve ser ortogonal a qualquer vetor do plano tangente T à superfície S em x^* .

Como x^* é regular, vale a caracterização de T dada no Teorema 4.0.1 e se deduz que existe $\lambda^* \in \mathbb{R}^m$ tal que:

$$\nabla f(x^*) = J_h^T(x^*)\lambda^*, \quad (4.8)$$

ou,

$$\nabla f(x^*) = \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*)$$

e

$$Z^T(x^*)\nabla f(x^*) = 0 \quad (4.9)$$

■

4.2 Condições de Segunda Ordem

Teorema 4.2.1 *Sejam x^* um ponto regular, minimizador local de (4.1) e T como no Teorema 4.0.1. Suponha $f, h \in C^2$. Então existe $\lambda^* \in \mathbb{R}^m$ tal que:*

$$\nabla f(x^*) + \sum_{i=1}^n \lambda_i^* \nabla h_i(x^*) = 0 \quad (4.10)$$

e

$$y^T \nabla_x^T L_x(x^*, \lambda^*) y \geq 0, \forall y \in T \quad (4.11)$$

onde $L(x, \lambda) = f(x) + \lambda^T h(x)$, $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^m$ é a função Lagrangeana.

Demonstração: Primeiramente note que (4.10) é a Condição Necessária de Primeira Ordem. Suponha agora que $x(t) \in C^2$.

A Condição Necessária de Segunda Ordem para (4.6) é $\varphi''(t^*) \geq 0$. Agora,

$$\varphi'(t) = \nabla^T f(x(t))x'(t) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x(t))x'_i(t),$$

portanto,

$$\varphi''(t) = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}(x(t))x'_i(t) \right)' \quad (4.12)$$

Mas,

$$\left(\frac{\partial f}{\partial x_i}(x(t))x'_i(t) \right)' = \nabla^T \left[\frac{\partial f}{\partial x_i}(x(t)) \right] x'(t)x'_i(t) + \frac{\partial f}{\partial x_i}(x(t))x''_i(t). \quad (4.13)$$

De (4.12) e (4.13) tem-se:

$$\varphi''(t) = x'(t)^T \nabla^2 f(x(t)) x'(t) + \nabla^T f(x(t)) x''(t). \quad (4.14)$$

Por outro lado, para qualquer $\lambda \in \mathbb{R}^m$, $j \in \{1, 2, \dots, m\}$ e $t \in (a, b)$,

$$\rho_j(t) \equiv \lambda_j h_j(x(t)) = 0$$

Portanto, para todo $j \in \{1, 2, \dots, m\}$ e $t \in (a, b)$,

$$\rho_j'(t) = \lambda_j \nabla^T h_j(x(t)) x'(t) = 0$$

e

$$\rho_j''(t) = \lambda_j [x'(t)^T \nabla^2 h_j(x(t)) x'(t) + \nabla^T h_j(x(t)) x''(t)] = 0$$

Então,

$$x'(t)^T \left[\sum_{j=1}^m \lambda_j \nabla^2 h_j(x(t)) \right] x'(t) + \left[\sum_{j=1}^m \lambda_j \nabla h_j(x(t)) \right]^T x''(t) = 0. \quad (4.15)$$

De (4.14) e (4.15) resulta, para $t = t^*$,

$$\varphi''(t^*) = x'(t^*) \left[\nabla^2 f(x^*) + \sum_{j=1}^m \lambda_j \nabla^2 h_j(x^*) \right] x'(t^*) + \left[\nabla f(x^*) + \sum_{j=1}^m \lambda_j \nabla h_j(x^*) \right]^T x''(t^*). \quad (4.16)$$

Se x^* é minimizador local de (4.1), sabe-se que existe $\lambda^* \in \mathbb{R}^m$ tal que $\nabla f(x^*) + \sum_{j=1}^m \lambda_j^* \nabla h_j(x^*) = 0$, portanto de (4.16) obtem-se,

$$\varphi''(t^*) = x'(t^*) \left[\nabla^2 f(x^*) + \sum_{j=1}^m \lambda_j^* \nabla^2 h_j(x^*) \right] x'(t^*) \geq 0, \quad (4.17)$$

onde $x'(t^*)$ é qualquer vetor de T . Dado que

$$\nabla_x^2 L(x, \lambda) = \nabla^2 f(x) + \sum_{j=1}^m \lambda_j \nabla^2 h_j(x),$$

(4.11) segue de (4.15). ■

É importante tentar entender o significado de (4.11). As condições de segunda ordem expressam sempre informações sobre a curvatura das funções. No caso de restrições lineares, nas condições de segunda ordem aparece somente a função objetivo. Se considerar restrições não-lineares, (4.11) significa que as curvaturas, tanto da função objetivo como das restrições, devem ser levadas em conta para caracterizar um minimizador local.

Na Figura 4.4, ilustra-se a importância da curvatura das restrições na caracterização de um minimizador e observa-se que \bar{x} é minimizador local de f sujeita a $h(x) = 0$, mas é maximizador de f sujeita a $g(x) = 0$.

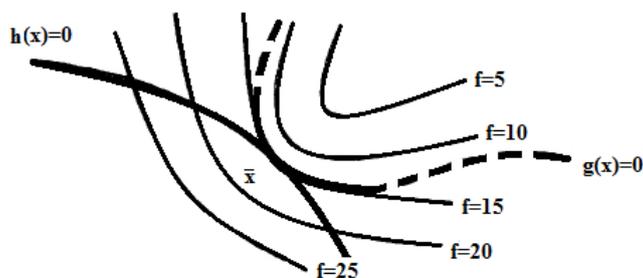


Figura 4.4: Curvatura das restrições.

Fonte: Friedlander, 1994.

No seguinte Teorema, são mostradas as condições suficientes de segunda ordem para que um ponto regular seja minimizador estrito de f com restrições não-lineares de igualdade.

Teorema 4.2.2 *Seja x^* um ponto regular tal que $h(x^*) = 0$ e T como no Teorema 4.0.1. Se $\lambda^* \in \mathbb{R}^m$ é tal que:*

$$\nabla f(x^*) + \sum_{j=1}^m \lambda_j^* \nabla h_j(x^*) = 0 \quad (4.18)$$

e

$$y^T \nabla_x^2 L(x^*, \lambda^*) y > 0, \forall y \in T - \{0\}$$

Então x^* é um minimizador local estrito.

Demonstração: Se x^* não é um ponto mínimo estrito, existe uma sequência de pontos viáveis (y_k) convergente para x^* tal que para cada k , $f(y_k) \leq f(x^*)$. Escrevendo cada y_k na forma $y_k = x^* + \delta_k s_k$ onde $s_k \in \mathbb{R}^n$, $\|s_k\| = 1$ e $\delta_k > 0$ para cada k . Claramente $\delta_k \rightarrow 0$ e a sequência s_k sendo limitada, tendo então uma subsequência convergente para algum s^* . Temos também que $h(y_k) - h(x^*) = 0$, e dividindo por δ_k e fazendo $k \rightarrow \infty$ vemos que $\nabla h(x^*) s^* = 0$.

Agora, pelo Teorema de Taylor, temos que para cada j ,

$$0 = h_j(y_k) = h_j(x^*) + \delta_k \nabla h_j(x^*) s_k + \frac{\delta_k^2}{2} s_k^T \nabla^2 h_j(\eta_j) s_k \quad (4.19)$$

e,

$$0 \geq f(y_k) - f(x^*) = \delta_k \nabla f(x^*) s_k + \frac{\delta_k^2}{2} s_k^T \nabla^2 f(\eta_0) s_k, \quad (4.20)$$

onde cada η_j é um ponto no segmento da reta que une x^* e y_k , multiplicando (31) por λ_j e adicionando este para (32), obtem-se, em soma para (30),

$$0 \leq \frac{\delta_k^2}{2} s_k^T \left\{ \nabla^2 f(\eta_0) + \sum_{j=1}^m \lambda_j \nabla^2 h_j(\eta_j) \right\} s_k$$

o que resulta em uma contradição quando $k \rightarrow \infty$.

■

Capítulo 5

Formas Quadráticas

Neste capítulo, será feita uma análise sobre formas quadráticas, serão verificadas suas propriedades, será dada ênfase quando a matriz da forma quadrática for simétrica. Uma forma quadrática em n variáveis reais x_1, x_2, \dots, x_n associadas a uma matriz real A é uma quantidade escalar que consiste em uma soma dos múltiplos dos produtos e quadrados das variáveis (Ben Noble, 1988):

$$F = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j = a_{11} x_1^2 + (a_{12} + a_{21}) x_1 x_2 + a_{22} x_2^2 + \dots + a_{nn} x_n^2$$

Se forem acrescentadas as matrizes $x = [x_1]$, $A = a_{ij}$, a forma quadrática pode ser escrita, usando a notação de produto interno, como:

$$F = \sum_{i=1}^n a_i \sum_{j=1}^n a_{ij} x_j = \sum_{i=1}^n x_i (Ax)_i = \langle x, Ax \rangle$$

Segundo a definição que foi dada, pode-se associar a uma forma quadrática qualquer matriz A , mas é claro que, quando a forma quadrática é desenvolvida, os termos envolvendo a_{ij} e a_{ji} contribuirão com:

$$(a_{ij} + a_{ji}) x_i x_j$$

Ou seja, a contribuição depende da soma, e não separadamente. Isso significa que, se, dada uma matriz geral A e a partir da mesma obter uma segunda matriz $A' = \frac{1}{2}(A + A^T)$, cujos (i-j), e (j-i)-ésimos coeficientes são ambos $\frac{1}{2}(a_{ij} + a_{ji})$ as formas quadráticas $\langle x, Ax \rangle$ e $\langle x, A'x \rangle$ são idênticas (Ben Noble, 1988). Portanto, sem perda de generalidade, pode-se supor que uma forma quadrática está associada a uma matriz simétrica desde que a_{ij} e x_i sejam reais.

Definição 5.0.1 *A forma quadrática associada à uma matriz simétrica A é a quantidade real $\langle x, Ax \rangle$.*

Observe que, em muitos textos, a forma quadrática $\langle x, Ax \rangle$ é escrita como $x^T Ax$. A chave do estudo de formas quadráticas é a ideia de mudança de variáveis coordenadas x para coordenada y , por meio de uma transformação:

$$x = Sy$$

Em que S é uma matriz não-singular. Exige-se o fato das matrizes S serem não-singulares pois são somente nesses casos que podemos escrever $y = S^{-1}x$, de maneira que um x dado determina em um único y , e um y dado determina um único x (Ben Noble, 1988). Isso corresponde a escolher uma nova base (o conjunto das colunas de S) para representar os vetores; se interesse fosse em transformações lineares por uma matriz A , isso conduz ao estudo de semelhança $S^{-1}AS$ e ao problema de até que ponto uma matriz pode ser simplificada por uma transformação de semelhança. No momento, contudo, o interesse é em formas quadráticas, em vez de transformações lineares, e assim deve-se perguntar como a forma quadrática se altera sob a mudança de variáveis $x = Sy$. A substituição de x por Sy em $F = \langle x, Ax \rangle$ fornece:

$$F = \langle x, Ax \rangle = \langle Sy, ASy \rangle = (Sy)^T ASy = y^T S^T ASy = \langle y, S^T ASy \rangle = \langle y, By \rangle$$

onde $B = S^T AS$. Se puder achar S de maneira que B seja uma matriz diagonal com elementos β_i , então será obtido uma representação para F , do tipo:

$$F = \beta_1 y_1^2 + \beta_2 y_2^2 + \dots + \beta_n y_n^2$$

E a forma de F teria sido consideravelmente simplificada. Até agora, nenhuma hipótese foi feita sobre S , exceto que deve ser não-singular. Em muitas aplicações das formas quadráticas deseja-se comparar os tamanhos de várias quantidades que estão envolvidas, de maneira que, se mudam-se variáveis por uma fórmula do tipo $x = Sy$, é importante que os comprimentos sejam preservados, isto é, que $\langle x, x \rangle = \langle y, y \rangle$. Então $\langle x, x \rangle = \langle Sy, Sy \rangle = \langle y, S^T Sy \rangle = \langle y, y \rangle$ para todo x , de maneira que $S^T S = I$. Isso significa que, a fim de preservar distâncias, S deve ser ortogonal (Ben Noble, 1988), e então teremos uma transformação ortogonal. Se A é simétrica, existe então uma matriz ortogonal P tal que $P^T AP$ é uma matriz diagonal cujos elementos da diagonal são os autovalores de A .

Teorema 5.0.3 *Se A é simétrica real e P é uma matriz ortogonal cujas colunas são autovetores normalizados de A , então se $x = Py$,*

$$\langle x, Ax \rangle = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_n y_n^2$$

Demonstração: Fazendo a seguinte substituição $x = Py$, e sabendo que $P^T AP = D$, tem-se:

$$\langle x, Ax \rangle = \langle Py, APy \rangle = (Py)^T APy = y^T P^T APy = \langle y, P^T APy \rangle = \langle y, Dy \rangle$$

onde D é uma matriz diagonal cujos elementos são autovalores de A .

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, Dy = \begin{bmatrix} \lambda_1 y_1 \\ \lambda_2 y_2 \\ \vdots \\ \lambda_n y_n \end{bmatrix}$$

assim sendo,

$$\langle y, Dy \rangle = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_n y_n^2$$

■

Uma forma quadrática que não contem produtos cruzados é chamada uma forma diagonal. Diz-se que a transformação $x = Py$ no Teorema acima diagonalizou a forma quadrática. O Teorema acima mostra que qualquer forma quadrática pode ser diagonalizada por uma transformação ortogonal.

A superfície no espaço n-dimensional definida pela equação:

$$\langle x, Ax \rangle = 1$$

em que A é uma matriz real simétrica de ordem n , é conhecida como *quádrlica central*. No caso $n = 2$ a palavra "superfície" deve ser substituída por "curva".

No caso n -dimensional temos a forma:

$$\lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_n y_n^2 = 1$$

Veremos agora algumas "curiosidades" para o caso de dimensão 2:

$$\lambda_1 y_1^2 + \lambda_2 y_2^2 = 1$$

Temos três possibilidades:

1. Elipse: λ_1, λ_2 tem o mesmo sinal;
2. Hipérbole: λ_1, λ_2 tem sinais opostos;
3. Duas retas paralelas: Um dos autovalores é nulo.

5.1 Formas Quadráticas Definidas

Algumas formas quadráticas tem a propriedade de que $\langle x, Ax \rangle > 0, \forall x \neq 0$. Um exemplo simples é:

$$F = x_1^2 + x_2^2 + 2x_3^2$$

Outro tipo importante de forma quadrática tem a propriedade $\langle x, Ax \rangle \geq 0, \forall x$, embora para esta classe de forma não seja necessariamente verdadeiro que $\langle x, Ax \rangle = 0$ implique $x = 0$. Um exemplo simples é:

$$F = (x_1 - x_2)^2 + 2x_3^2$$

Que é zero desde que os dois primeiros valores sejam iguais e diferente de zero e o terceiro valor seja zero. Em geral, naturalmente, uma forma quadrática pode ter valores positivos, nulos e negativos. Os casos particulares cujos valores são todos do mesmo sinal são importantes, e se introduz a seguinte terminologia.

Definição 5.1.1 *(A matriz A nesta definição é sempre simétrica). A forma quadrática $\langle x, Ax \rangle$ é dita "definida positiva" se $\langle x, Ax \rangle > 0, \forall x \neq 0$ e "semi definida positiva" quando $\langle x, Ax \rangle \geq 0, \forall x$. De forma análoga definem-se matrizes e formas quadráticas "definida negativa" e "semi definida negativa". Uma forma quadrática é dita "indefinida" quando for positiva para alguns valores de x e negativa para outros.*

Segundo esta definição, matrizes definidas positivamente também são semi definidas positivamente. Se for necessário dizer que $\langle x, Ax \rangle \geq 0$ e existe um $x \neq 0$ tal que $\langle x, Ax \rangle = 0$, diremos que A é semi definida positiva e singular.

Em transformações ortogonais, os comprimentos originais são inalterados. Ao estudar se uma forma quadrática é definida positiva ou não, é conveniente enfraquecer o fato de ela ser ortogonal e considerar que qualquer tipo de transformação $B = S^T A S$ em que S é não-singular, pois então B escreve a nossa forma quadrática relativamente a uma nova base (Ben Noble, 1988).

Definição 5.1.2 *Dizemos que A e B são semelhantes, se existe uma matriz não-singular S tal que $B = S^{-1} A S$*

É importante observar que o fato de ser definida positiva não é afetada por uma transformação de semelhança; isto é intuitivamente óbvio, pois A e B representam a "mesma" forma quadrática, mas relativas a bases diferentes.

Capítulo 6

Quociente de Rayleigh

Uma das motivações iniciais do estudo de formas quadráticas foi a consideração de um problema estatístico de análise de dados que exige o cálculo de extremos (em verdade a maximização) de uma forma quadrática submetida a uma normalização de suas variáveis. Agora que a compreensão necessária das formas quadráticas e condições de otimalidade foi abrangida, pode-se "olhar" melhor para achar extremos do seguinte formato:

$$\text{Maximize } \langle a, Sa \rangle = a^T Sa \quad (6.1)$$

$$\text{Sujeita à } \langle a, a \rangle = 1$$

Teorema 6.0.1 *Seja A uma matriz simétrica. Então x_0 maximiza $\frac{\langle x, Ax \rangle}{\langle x, x \rangle}$ com a restrição $x \neq 0$ e fornece um valor máximo M se e somente se $x_1 = \frac{x_0}{\|x_0\|}$ maximiza $\langle x, Ax \rangle$ com a restrição $\langle x, x \rangle = 1$ e fornece um valor máximo igual a M .*

Demonstração: Suponha que x_0 maximiza $\frac{\langle x, Ax \rangle}{\langle x, x \rangle}$ com a restrição $x \neq 0$. Então x_1 é tal que $\langle x_1, x_1 \rangle = 1$ e para qualquer x com $\langle x, x \rangle = 1$, teremos:

$$\langle x, Ax \rangle = \frac{\langle x, Ax \rangle}{\langle x, x \rangle} \leq \frac{\langle x_0, Ax_0 \rangle}{\langle x_0, x_0 \rangle} = \langle x_1, Ax_1 \rangle = M$$

como desejado. Reciprocamente, se x_1 com $\langle x_1, x_1 \rangle = 1$ maximiza $\langle x, Ax \rangle$ sujeita à restrição $\langle x, x \rangle = 1$, então para qualquer escalar não-nulo α e qualquer vetor não-nulo x , o vetor $y = \frac{x}{\|x\|}$ tem $\|y\| = 1$, e portanto $x_0 = \alpha x_1$ e temos:

$$\frac{\langle x_0, Ax_0 \rangle}{\langle x_0, x_0 \rangle} = \langle x_1, Ax_1 \rangle \geq \langle y, Ay \rangle = \frac{\langle x, Ax \rangle}{\langle x, x \rangle}$$

■

Definição 6.0.3 *O Quociente de Rayleigh de uma matriz simétrica A é a expressão:*

$$\rho(x) = \frac{\langle x, Ax \rangle}{\langle x, x \rangle}$$

onde $x \neq 0$

Teorema 6.0.2 *Se A é simétrica e tem autovalores $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ e autovetores ortonormalizados associados x_1, x_2, \dots, x_n , então:*

$$\lambda_1 \leq \rho(x) \leq \lambda_n$$

Onde $\rho(x)$ é o Quociente de Rayleigh para qualquer $x \neq 0$ e

$$\lambda_1 = \min \rho(x) = \rho(x_1), \lambda_n = \max \rho(x) = \rho(x_n)$$

Demonstração: Como A é uma matriz simétrica, é possível obter uma base para \mathbb{R}^n consistindo de autovetores de A . Então, dado um vetor arbitrário x , escrevendo em termos dos autovetores de A , tem-se:

$$\begin{aligned} x &= \sum_{i=1}^n \alpha_i x_i \\ Ax &= \sum_{i=1}^n \alpha_i \lambda_i x_i \\ \rho(x) &= \frac{\langle x, Ax \rangle}{\langle x, x \rangle} = \frac{\sum_{i=1}^n \alpha_i^2 \lambda_i}{\sum_{i=1}^n \alpha_i^2} \end{aligned}$$

Subtraindo λ_1 dos dois membros da igualdade,

$$\rho(x) - \lambda_1 = \frac{\sum_{i=2}^n \alpha_i^2 (\lambda_i - \lambda_1)}{\sum_{i=1}^n \alpha_i^2}$$

Como $\lambda_i > \lambda_1, \forall i$, temos que $\rho(x) > \lambda_1$. E se tomarmos $x = x_1$, tem-se:

$$\rho(x_1) = \frac{\langle x_1, Ax_1 \rangle}{\langle x_1, x_1 \rangle} = \frac{\langle x_1, \lambda_1 x_1 \rangle}{\langle x_1, x_1 \rangle} = \lambda_1 \frac{\langle x_1, x_1 \rangle}{\langle x_1, x_1 \rangle} = \lambda_1$$

o que demonstra a primeira afirmativa do Teorema. O restante é demonstrado de forma análoga, só que considerando agora $\rho(x) - \lambda_n$. ■

Como aplicação do projeto, relembRANDO o problema estatístico na qual foi apresentado no início deste trabalho, desejava-se achar a combinação linear:

$$y_1 = \langle a, x \rangle = a_1 x_1 + a_2 x_2 + \dots + a_n x_n$$

das variáveis x_1, x_2, \dots, x_n que maximiza a variância $\langle a, Sa \rangle$ de y_1 restrito à $\langle a, a \rangle = 1$; vemos agora pelo Teorema 6.0.2, que necessita-se simplesmente escolher o vetor " a " como sendo o autovetor da Matriz de Covariância associado ao maior autovalor de S . A nova variável y_1 é chamada *componente principal*. Se esta componente sozinha não apresenta adequadamente a variação total $s_{11} + s_{22} + \dots + s_{nn}$, pode-se então razoavelmente introduzir uma segunda componente principal:

$$y_2 = \langle b, x \rangle = b_1 x_1 + b_2 x_2 + \dots + b_n x_n$$

escolhida de maneira que $\langle b, b \rangle = 1$ e tal que y_2 seja responsável por tanta variação quanto possível, independente da já devida y_1 ; sabemos da estatística, que deve-se então exigir a ortogonalidade de " b " e " a ", ou seja, $\langle a, b \rangle = 0$. Deve-se assim, procurar " b " que maximize $\langle b, Sb \rangle$ com as restrições $\langle b, b \rangle = 1$ e $\langle b, a \rangle = 0$. Seguir-se-á do Teorema 6.0.3 que será anunciado brevemente, em que " b " deve ser escolhido como sendo o autovetor de S associado ao segundo maior autovalor de S ; mais geralmente, componentes principais sucessivas deveriam ser escolhidas como autovetores associados com autovalores sucessivamente menores. Usando a linguagem da Álgebra Linear, a fim de reenunciar o novo problema de tentar achar a segunda componente principal " b ". Dado que A é uma matriz simétrica com autovalores:

$$\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$$

e autovetores ortonormalizados associados x_1, x_2, \dots, x_n , maximize $\langle x, Ax \rangle$ com as restrições:

$$\langle x, x \rangle = 1 \text{ e } \langle x, x_n \rangle = 0$$

Teorema 6.0.3 *Seja A uma matriz simétrica de ordem n com autovalores $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ e autovetores ortonormalizados associados x_1, x_2, \dots, x_n . Para cada j com $1 \leq j \leq n$ seja:*

$$S_j \equiv \text{o conjunto de todos os } x \neq 0 \text{ que satisfazem} \\ \langle x, x_n \rangle = \langle x, x_{n-1} \rangle = \dots = \langle x, x_{n-j+1} \rangle = 0,$$

$$T_j \equiv \text{o conjunto de todos os } x \neq 0 \text{ que satisfazem} \\ \langle x, x_1 \rangle = \langle x, x_2 \rangle = \dots = \langle x, x_{j-1} \rangle = 0,$$

$$\rho(x) = \frac{\langle x, Ax \rangle}{\langle x, x \rangle} \text{ para } x \neq 0. \text{ Então:}$$

$$\max_{S_j} \rho(x) = \lambda_{n-j} = \rho(x_{n-j}) \text{ e } \min_{T_j} \rho(x) = \lambda_j = \rho(x_j)$$

Demonstração: Qualquer $x \in S_j$ deve ser da forma:

$$x = \sum_{i=1}^{n-j} \alpha_i x_i$$

de maneira que:

$$\rho(x) - \lambda_{n-j} = \frac{\sum_{i=1}^{n-j-1} (\lambda_i - \lambda_{n-j}) \alpha_i^2}{\sum_{i=1}^{n-j} \alpha_i^2}$$

e portanto $\rho(x) - \lambda_{n-j} \leq 0$ com $\rho(x_{n-j}) = \lambda_{n-j}$ como desejado. Um raciocínio semelhante funciona para $x \in T_j$. ■

Para tornar este Teorema mais claro, podemos escrevê-lo de outra forma:

1. x_{n-1} maximiza $\rho(x)$ com as restrições $\langle x, x \rangle = 1$ e $\langle x, x_n \rangle = 0$ fornecendo um valor máximo $\rho(x_{n-1}) = \lambda_{n-1}$;
2. x_2 minimiza $\rho(x)$ com as restrições $\langle x, x \rangle = 1$ e $\langle x, x_1 \rangle = 0$ fornecendo um valor mínimo igual a $\rho(x_2) = \lambda_2$

Capítulo 7

Análise de Componentes Principais

Após os estudos realizados em condições de otimalidade, utilizou-se as condições necessárias e suficientes para se resolver um problema de otimização não-linear com restrições de igualdade, ou seja, as condições para resolver o problema 2.1. O estudo realizado sobre o Quociente de Rayleigh permitiu a caracterização para a solução do problema. Constatou-se que o autovetor associado ao maior autovalor da matriz de covariância, é a solução do problema 2.1. Esse autovetor está associado à *primeira componente principal*.

Se existe uma segunda componente não correlacionada com a primeira de tal sorte que as duas acumulam uma quantidade significativa da variância contida nos dados originais, acrescenta-se uma restrição ao problema original. A solução desse novo problema, também de igualdade, está relacionado com a *segunda componente principal*.

Como aplicação dos estudos realizados, buscou-se o entendimento de uma ferramenta da Análise de Dados Multivariados (ADM), a Análise de Componentes Principais (ACP). O software utilizado para a implementação da ferramenta foi o Tanagra¹.

A Análise de Componentes Principais é um método que tem por finalidade básica, a análise dos dados, fazendo a eliminação de redundâncias dos dados a partir de combinações lineares das variáveis originais.

É também uma maneira de identificar a relação entre características extraídas de dados, sendo bastante útil quando os vetores aleatórios estão imersos num espaço de dimensão relativamente grande, quando uma representação gráfica não é possível.

Para aplicação da Análise de Componentes Principais, foram coletadas as notas das Escolas de Samba do Grupo Especial do Rio de Janeiro deste ano². O grupo consiste de doze escolas de samba, avaliadas em dez quesitos, com quatro jurados para cada quesito. A nota de uma escola de samba em um determinado quesito é atribuída da seguinte maneira: cada jurado atribui uma nota de 8.0 à 10.0, podendo ter até uma casa decimal. Dadas as notas dos quatro jurados, a menor é descartada e somam-se as três restantes, e o resultado dessa soma é a nota da escola no quesito em questão. Após esse procedimento ter sido realizado em todos os quesitos, somam-se as notas das escolas em cada quesito, e a escola com mais pontos é a campeã. As notas das escolas de samba em cada quesito estão no Apêndice.

O procedimento usado para a aplicação das componentes principais foi o seguinte: dadas as notas dos quatro jurados, calculou-se a média aritmética das notas e o resultado foi atribuído como sendo a nota da escola no quesito em questão, ou seja, não foi descartada nenhuma nota.

Nas Tabelas 1 e 2, são apresentadas as médias das escolas em cada quesito.

¹Disponível em: <<http://tanagra.software.informer.com/1.4/>>

²Acessado em: <<http://g1.globo.com/rio-de-janeiro/carnaval/2013/noticia/2013/02/confira-notas-da-apuracao-do-desfile-do-grupo-especial-do-rio.html>>

Tabela 1: Médias das escolas nos cinco primeiros quesitos

Escolas	Harmonia	Conjunto	Mestre-Sala e Porta-Bandeira	Enredo	Comissão de Frente
Inocentes de Belfort Roxo	9.6	9.65	9.85	9.625	9.675
Salgueiro	9.875	9.95	9.9	9.925	9.925
Unidos da Tijuca	9.925	9.975	9.9	10	9.975
União da Ilha	9.825	9.775	9.825	9.95	9.7
Mocidade	9.725	9.8	9.8	9.65	9.725
Portela	9.975	9.85	9.875	9.8	9.725
São Clemente	9.775	9.7	9.8	9.65	9.65
Mangueira	9.85	9.8	10	9.975	9.7
Beija-Flor	9.975	9.95	10	9.95	9.9
Grande Rio	9.825	9.925	9.875	9.875	9.9
Imperatriz Leopoldinense	9.975	9.9	9.85	9.975	9.9
Vila Isabel	10	10	10	9.95	9.95

Tabela 2: Médias das escolas nos cinco últimos quesitos

Escola	Alegorias e Adereços	Bateria	Samba-Enredo	Fantasia	Evolução
Inocentes de Belfort Roxo	9.55	9.725	9.775	9.575	9.7
Salgueiro	9.25	9.9	9.8	9.925	9.925
Unidos da Tijuca	9.25	9.975	9.85	9.925	9.95
União da Ilha	9.85	9.725	9.725	9.875	9.775
Mocidade	9.725	9.825	9.675	9.775	9.725
Portela	9.7	9.975	10	9.75	9.875
São Clemente	9.6	9.875	9.75	9.725	9.825
Mangueira	9.675	9.975	9.825	9.85	9.825
Beija-Flor	9.9	9.975	9.975	9.975	9.975
Grande Rio	9.85	9.875	9.65	9.8	9.925
Imperatriz Leopoldinense	9.9	9.875	9.875	9.5	9.925
Vila Isabel	9.9	9.9	9.95	9.975	10

A análise foi feita sobre as médias das escolas. Cada escola de samba é um indivíduo do \mathbb{R}^{10} . Na Figura 7.1 mostra-se a matriz de correlações, e percebe-se que tem itens que estão fortemente correlacionados, como por exemplo os quesitos Conjunto e Evolução.

A escolha pela média é devido ao fato de que, como as notas estão entre 8.0 e 10.0, o desvio delas é pequeno, as notas estão pouco dispersas. Assim sendo todas as notas estão muito próximas da média, o que a torna uma boa representação para as notas.

Correlations

	2 - Conjunto	10 - Evolução	1 - Harmonia	9 - Fantasias	5 - Comissão de Frente	6 - Alegorias	4 - Enredo	3 - M. Sala & P. Bandeira	7 - Bateria	8 - Samba-Enredo
2 - Conjunto	1,00000	0,90094	0,80282	0,83518	0,93471	0,89918	0,71648	0,56468	0,58974	0,40866
10 - Evolução	0,90094	1,00000	0,86859	0,78120	0,84885	0,76213	0,68456	0,61185	0,68033	0,55007
1 - Harmonia	0,80282	0,86859	1,00000	0,81063	0,63848	0,70459	0,71030	0,54890	0,69799	0,72357
9 - Fantasias	0,83518	0,78120	0,81063	1,00000	0,74492	0,89902	0,82350	0,54778	0,46303	0,40438
5 - Comissão de Frente	0,93471	0,84885	0,63848	0,74492	1,00000	0,87741	0,65899	0,45899	0,43014	0,30657
6 - Alegorias	0,89918	0,76213	0,70459	0,89902	0,87741	1,00000	0,77979	0,35932	0,29984	0,22575
4 - Enredo	0,71648	0,68456	0,71030	0,82350	0,65899	0,77979	1,00000	0,64013	0,41826	0,38903
3 - M. Sala & P. Bandeira	0,56468	0,61185	0,54890	0,54778	0,45899	0,35932	0,64013	1,00000	0,60597	0,61146
7 - Bateria	0,58974	0,68033	0,69799	0,46303	0,43014	0,29984	0,41826	0,60597	1,00000	0,59346
8 - Samba-Enredo	0,40866	0,55007	0,72357	0,40438	0,30657	0,22575	0,38903	0,61146	0,59346	1,00000

Figura 7.1: Matriz de Correlações extraída do software Tanagra.

Na Figura 7.2, é apresentado a contribuição de cada componente principal, onde "Axis-i" é a i-ésima componente, e seu autovalor associado dividido pela soma dos autovalores, é a contribuição da variância explicada por esta componente.

Eigen values

Matrix trace	10,000000
Average	1,000000

Axis	Eigen value	Difference	Proportion (%)	Histogram	Cumulative (%)
1	6,872954	5,476623	68,73 %		68,73 %
2	1,396331	0,805374	13,96 %		82,69 %
3	0,590957	0,115931	5,91 %		88,60 %
4	0,475026	0,139969	4,75 %		93,35 %
5	0,335057	0,188695	3,35 %		96,70 %
6	0,146362	0,036714	1,46 %		98,17 %
7	0,109647	0,051268	1,10 %		99,26 %
8	0,058379	0,044175	0,58 %		99,85 %
9	0,014204	0,013121	0,14 %		99,99 %
10	0,001083	-	0,01 %		100,00 %
Tot.	10,000000	-	-		-

Figura 7.2: Variância explicada por cada componente extraída do software Tanagra.

Na Figura 7.3, são apresentadas as correlações de cada variável (quesito) com cada componente. A maioria das variáveis, estão fortemente correlacionadas com a primeira componente, sendo sete das dez variáveis explicada em 70% ou mais apenas pela primeira componente. Os quesitos Bateria e Samba-Enredo, também estão correlacionadas com a segunda componente, e sua explicação foi "quebrada" entre as duas componentes.

Factor Loadings

Attribute	Axis_1		Axis_2		Axis_3		Axis_4		Axis_5	
	Corr.	% (Tot. %)	Corr.	% (Tot. %)	Corr.	% (Tot. %)	Corr.	% (Tot. %)	Corr.	% (Tot. %)
2 - Conjunto	0,94206	89 % (89 %)	-0,20242	4 % (93 %)	0,16263	3 % (95 %)	0,12144	1 % (97 %)	0,07270	1 % (97 %)
10 - Evolução	0,93824	88 % (88 %)	0,01875	0 % (88 %)	0,20504	4 % (92 %)	0,06508	0 % (93 %)	0,06620	0 % (93 %)
1 - Harmonia	0,90841	83 % (83 %)	0,18298	3 % (86 %)	0,12265	2 % (87 %)	-0,30116	9 % (96 %)	-0,12006	1 % (98 %)
9 - Fantasias	0,90045	81 % (81 %)	-0,22446	5 % (86 %)	-0,14981	2 % (88 %)	-0,15201	2 % (91 %)	-0,15354	2 % (93 %)
5 - Comissão de Frente	0,85540	73 % (73 %)	-0,34273	12 % (85 %)	0,17638	3 % (88 %)	0,17793	3 % (91 %)	0,25631	7 % (98 %)
6 - Alegorias	0,85208	73 % (73 %)	-0,49783	25 % (97 %)	-0,00680	0 % (97 %)	-0,10018	1 % (98 %)	0,00334	0 % (98 %)
4 - Enredo	0,83418	70 % (70 %)	-0,14418	2 % (72 %)	-0,43430	19 % (91 %)	-0,04490	0 % (91 %)	-0,17045	3 % (94 %)
3 - M. Sala & P. Bandeira	0,70001	49 % (49 %)	0,41977	18 % (67 %)	-0,42509	18 % (85 %)	0,34414	12 % (97 %)	0,13491	2 % (98 %)
7 - Bateria	0,68083	46 % (46 %)	0,52163	27 % (74 %)	0,29057	8 % (82 %)	0,25074	6 % (88 %)	-0,32089	10 % (99 %)
8 - Samba-Enredo	0,60169	36 % (36 %)	0,66087	44 % (80 %)	-0,00859	0 % (80 %)	-0,34238	12 % (92 %)	0,26733	7 % (99 %)
Var. Expl.	6,87295	69 % (69 %)	1,39633	14 % (83 %)	0,59096	6 % (89 %)	0,47503	5 % (93 %)	0,33506	3 % (97 %)

Figura 7.3: Correlação entre as variáveis e as componentes extraída do software Tanagra.

Nota-se também que nove dos dez quesitos são explicados em 70% ou mais pelas duas primeiras componentes, sendo sete desses com ordem superior ou igual a 80%.

Na Figura 7.4, são apresentadas as projeções dos indivíduos (escolas) no hiperplano gerado pela primeira e segunda componentes. Conforme a Figura 7.2, as duas componentes juntas são responsáveis por 82,69% da variância total dos dados.

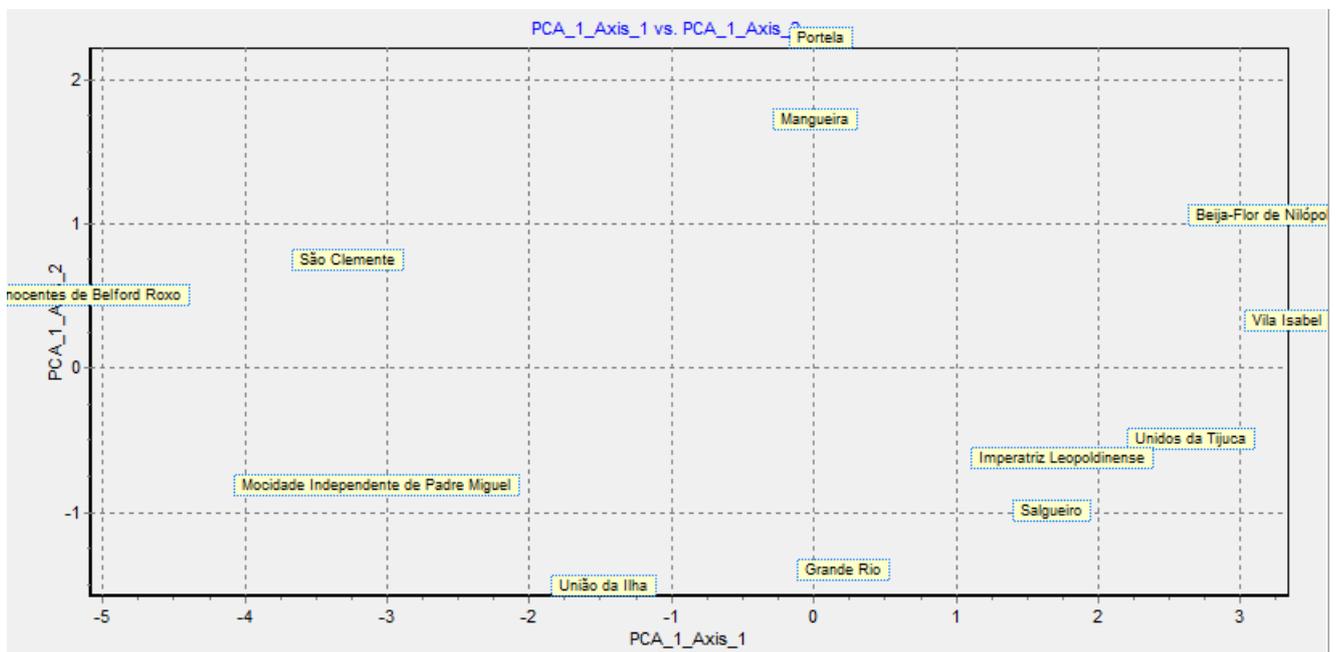


Figura 7.4: Projeção dos indivíduos no hiperplano gerado pelas componentes extraído do software Tanagra.

Na Figura 7.5, será apresentado o círculo das correlações, que mostram quanto as variáveis estão correlacionadas entre si, o que possibilita as escolas de samba fazerem uma preparação maior em determinados quesitos, para que o resultado final seja o melhor possível. O círculo das correlações mostra também o quão boa é a representação dos indivíduos no hiperplano gerado pelas componentes, quanto mais próximo do círculo, melhor é a representação no hiperplano.

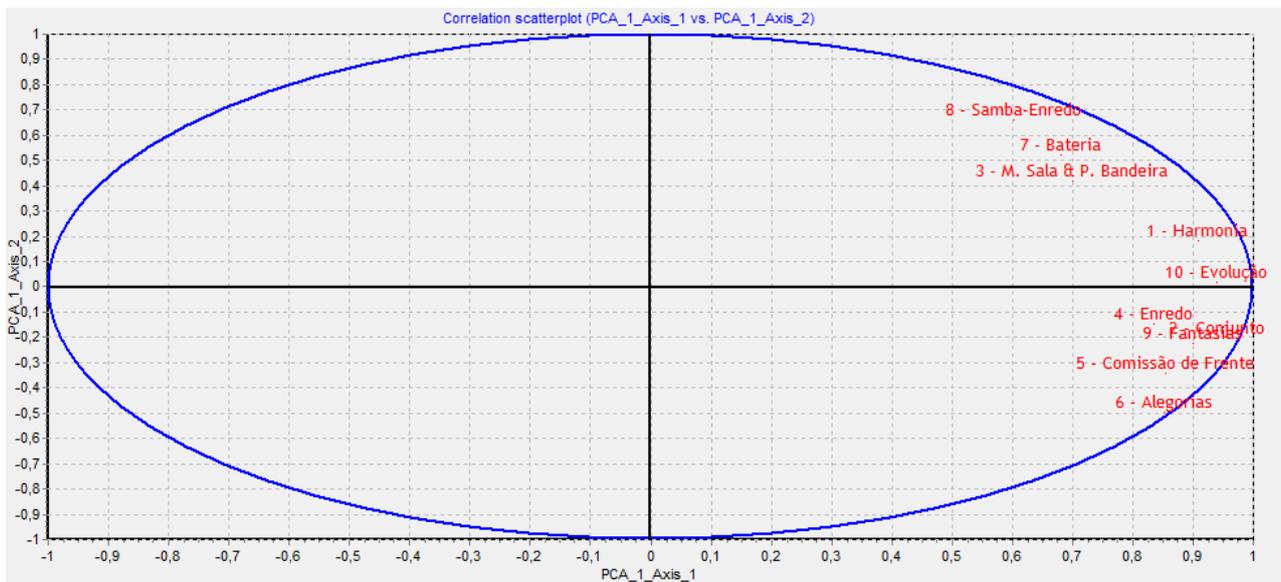


Figura 7.5: Círculo de Correlações extraído do software Tanagra.

Um fato que diz respeito às representações dos indivíduos neste plano, é que se todos os indivíduos forem projetados apenas na primeira componente, cria-se uma ordenação dos mesmos, devido ao isomorfismo entre a primeira componente e a reta \mathbb{R} . Fazendo uma comparação entre esta ordenação e o resultado oficial, tem-se que o resultado é praticamente o mesmo. A única diferença foi a troca de posições entre as escolas Portela e Mangueira.

Nota-se que outros quesitos estão fortemente correlacionados, ou seja, se uma escola foi bem no quesito Conjunto, as chances dessa mesma escola se dar bem no quesito Evolução são muito grandes. Do ponto de vista da escola, se torna muito mais viável investir em um desses quesitos, e naturalmente ir bem no outro.

Conclusão

Através dos estudos desenvolvidos, as metas previstas para o projeto foram todas alcançadas. Inicialmente foi feito o desenvolvimento teórico para problemas de Otimização Matemática. Mais adiante ainda na primeira etapa do projeto foram exploradas as principais propriedades relacionadas ao Quociente de Rayleigh.

O problema apresentado na Seção 2.1 foi solucionado ao longo do desenvolvimento do projeto. O estudo em otimização quadrática foi de grande importância visto que tendo em mãos as condições de otimalidade foi possível garantir a suficiência à solução do problema e através do estudo sobre o Quociente de Rayleigh uma característica foi dada a mesma. A matriz de covariância é simétrica, por isso seus autovalores são reais e é possível a ordenação desses. O maximizador da função objetivo é o autovetor associado ao maior autovalor da matriz de covariância.

Na segunda etapa, foi feito um levantamento de dados para aplicação de uma ferramenta da Análise de Dados Multivariados: a Análise de Componentes Principais.

Através do amadurecimento teórico adquirido, alguns resultados já eram esperados após o uso de softwares livres para a implementação da ferramenta, e na terceira etapa várias conclusões foram obtidas, apoiando a análise dos resultados.

Referências Bibliográficas

BUSSAB, Wilton; MORETTIN, Pedro. Estatística Básica. 6^a ed. São Paulo: Editora Saraiva, 2010.

FRIEDLANDER, A. Elementos de Programação Não-Linear. 1^aed. Campinas: UNICAMP, 1994.

G1 - Rio de Janeiro. Notas de apuração do grupo especial do Rio de Janeiro 2013. Disponível em: <<http://g1.globo.com/rio-de-janeiro/carnaval/2013/noticia/2013/02/confira-notas-da-apuracao-do-desfile-do-grupo-especial-do-rio.html>>. Acesso em: 14 de fevereiro de 2013

HOWARD, Anton; RORRES, Chris. Álgebra Linear com Aplicações. 8^a ed. Porto Alegre: Bookman, 2001.

LIMA, E. Curso de Análise vol.2. 11^a ed. Rio de Janeiro: IMPA, 2012.

MARTÍNEZ, José; SANTOS, Sandra. Métodos Computacionais de Otimização. 1^a ed. Rio de Janeiro: IMPA, 1995.

NOBLE, Ben; DANIEL, James. Applied Linear Algebra. 3^a ed. Englewood Cliffs, NJ: Prentice-Hall Inc., 1988.

RENCHER, Alvin. Methods of Multivariate Analysis. 2^a ed. New York, NY: John Wiley & Sons Inc., 2002.

Apêndice

Tabela 3: Notas do quesito Harmonia

Escola	Jurado 1	Jurado 2	Jurado 3	Jurado 4	Média
Inocentes de Belfort Roxo	9.6	9.5	9.6	9.7	9.6
Salgueiro	9.9	9.9	9.8	9.9	9.875
Unidos da Tijuca	10	9.8	10	9.9	9.925
União da Ilha	9.8	9.9	9.8	9.8	9.825
Mocidade	9.7	9.7	9.7	9.8	9.725
Portela	10	10	9.9	10	9.975
São Clemente	9.9	9.7	9.8	9.7	9.775
Mangueira	9.8	9.7	10	9.9	9.85
Beija-Flor	9.9	10	10	10	9.975
Grande Rio	10	9.6	9.8	9.9	9.9
Imperatriz Leopoldinense	9.9	10	10	10	9.975
Vila Isabel	10	10	10	10	10

Tabela 4: Notas do quesito Conjunto

Escola	Jurado 1	Jurado 2	Jurado 3	Jurado 4	Média
Inocentes de Belfort Roxo	9.6	9.7	9.6	9.7	9.65
Salgueiro	9.9	10	9.9	10	9.95
Unidos da Tijuca	10	10	10	9.9	9.975
União da Ilha	9.8	9.8	9.7	9.8	9.775
Mocidade	9.9	9.8	9.8	9.7	9.8
Portela	9.8	10	9.9	9.7	9.85
São Clemente	9.8	9.7	9.6	9.7	9.7
Mangueira	9.9	9.7	10	9.6	9.8
Beija-Flor	9.9	10	9.9	10	9.95
Grande Rio	10	10	9.8	9.9	9.925
Imperatriz Leopoldinense	9.8	9.9	10	9.9	9.9
Vila Isabel	10	10	10	10	10

Tabela 5: Notas do quesito Mestre-Sala e Porta-Bandeira

Escola	Jurado 1	Jurado 2	Jurado 3	Jurado 4	Média
Inocentes de Belfort Roxo	10	9.8	9.8	9.8	9.85
Salgueiro	9.8	9.9	9.9	10	9.9
Unidos da Tijuca	10	9.9	9.9	9.8	9.9
União da Ilha	9.8	9.8	9.9	9.8	9.825
Mocidade	9.7	9.9	9.8	9.8	9.8
Portela	9.8	9.9	9.9	9.9	9.875
São Clemente	9.7	9.8	9.8	9.9	9.8
Mangueira	10	10	10	10	10
Beija-Flor	10	10	10	10	10
Grande Rio	9.7	9.8	10	10	9.875
Imperatriz Leopoldinense	9.8	10	9.9	9.7	9.85
Vila Isabel	10	10	10	10	10

Tabela 6: Notas do quesito Enredo

Escola	Jurado 1	Jurado 2	Jurado 3	Jurado 4	Média
Inocentes de Belfort Roxo	9.6	9.7	9.7	9.7	9.675
Salgueiro	9.9	10	9.9	9.9	9.925
Unidos da Tijuca	10	10	10	10	10
União da Ilha	10	9.9	9.9	10	9.95
Mocidade	9.7	9.6	9.7	9.6	9.65
Portela	9.8	9.8	9.8	9.8	9.8
São Clemente	9.6	10	9.4	9.6	9.65
Mangueira	10	9.9	10	10	9.975
Beija-Flor	9.9	10	10	9.9	9.95
Grande Rio	9.6	10	9.9	10	9.875
Imperatriz Leopoldinense	10	9.9	10	19	9.975
Vila Isabel	10	9.8	10	10	9.95

Tabela 7: Notas do quesito Comissão de Frente

Escola	Jurado 1	Jurado 2	Jurado 3	Jurado 4	Média
Inocentes de Belfort Roxo	9.6	9.7	9.6	9.8	9.625
Salgueiro	9.9	9.9	9.9	10	9.925
Unidos da Tijuca	9.9	10	10	10	9.975
União da Ilha	9.6	9.8	9.7	9.7	9.7
Mocidade	9.5	9.9	9.8	9.7	9.725
Portela	9.5	9.8	9.9	9.7	9.725
São Clemente	9.5	9.6	9.8	9.7	9.65
Mangueira	9.9	9.6	9.8	9.5	9.7
Beija-Flor	9.9	10	10	9.7	9.9
Grande Rio	9.7	10	9.9	10	9.9
Imperatriz Leopoldinense	9.8	9.9	10	9.9	9.9
Vila Isabel	10	9.9	10	9.9	9.95

Tabela 8: Notas do quesito Alegorias e Adereços

Escola	Jurado 1	Jurado 2	Jurado 3	Jurado 4	Média
Inocentes de Belfort Roxo	9.5	9.6	9.6	9.5	9.55
Salgueiro	9.9	9.9	9.9	10	9.925
Unidos da Tijuca	10	10	10	9.7	9.925
União da Ilha	9.8	9.8	9.8	10	9.85
Mocidade	9.8	9.8	9.7	9.6	9.725
Portela	9.8	9.8	9.7	9.5	9.7
São Clemente	9.7	9.7	9.6	9.4	9.6
Mangueira	9.8	9.8	9.8	9.3	9.675
Beija-Flor	9.9	9.9	9.9	9.9	9.9
Grande Rio	9.8	9.9	9.7	10	9.85
Imperatriz Leopoldinense	9.8	9.9	9.9	10	9.9
Vila Isabel	10	10	10	9.6	9.9

Tabela 9: Notas do quesito Bateria

Escola	Jurado 1	Jurado 2	Jurado 3	Jurado 4	Média
Inocentes de Belfort Roxo	9.9	9.7	9.6	9.7	9.725
Salgueiro	9.8	10	10	9.8	9.9
Unidos da Tijuca	10	10	9.9	10	9.975
União da Ilha	9.7	9.8	9.8	9.6	9.725
Mocidade	9.8	9.9	10	9.6	9.825
Portela	10	10	10	9.9	9.975
São Clemente	10	9.9	9.7	9.9	9.875
Mangueira	10	9.9	10	10	9.975
Beija-Flor	10	9.9	10	10	9.975
Grande Rio	9.7	10	10	9.8	9.875
Imperatriz Leopoldinense	9.9	10	9.9	9.7	9.875
Vila Isabel	9.8	9.9	10	9.9	9.9

Tabela 10: Notas do quesito Samba-Enredo

Escola	Jurado 1	Jurado 2	Jurado 3	Jurado 4	Média
Inocentes de Belfort Roxo	9.7	9.8	9.8	9.8	9.775
Salgueiro	9.9	9.8	9.6	9.9	9.8
Unidos da Tijuca	10	9.9	9.7	9.8	9.85
União da Ilha	9.7	10	9.6	9.6	9.725
Mocidade	9.6	9.7	9.7	9.7	9.675
Portela	10	10	10	10	10
São Clemente	9.8	9.6	9.9	9.7	9.75
Mangueira	9.8	9.9	9.7	9.9	9.825
Beija-Flor	10	10	10	9.9	9.975
Grande Rio	9.8	9.6	9.6	9.6	9.65
Imperatriz Leopoldinense	9.9	9.9	9.8	9.9	9.875
Vila Isabel	10	9.8	10	10	9.95

Tabela 11: Notas do quesito Fantasias

Escola	Jurado 1	Jurado 2	Jurado 3	Jurado 4	Média
Inocentes de Belfort Roxo	9.6	9.6	9.6	9.5	9.575
Salgueiro	9.9	10	9.8	10	9.925
Unidos da Tijuca	10	9.9	19	9.8	9.925
União da Ilha	9.9	9.9	9.7	10	9.875
Mocidade	9.7	9.8	10	9.6	9.775
Portela	9.7	9.7	9.9	9.7	9.75
São Clemente	9.8	9.8	9.7	9.6	9.725
Mangueira	9.8	10	10	9.6	9.85
Beija-Flor	10	10	9.9	10	9.975
Grande Rio	9.6	9.9	9.7	10	9.8
Imperatriz Leopoldinense	10	9.9	10	9.9	9.95
Vila Isabel	10	10	10	9.9	9.975

Tabela 12: Notas do quesito Evolução

Escola	Jurado 1	Jurado 2	Jurado 3	Jurado 4	Média
Inocentes de Belfort Roxo	9.7	9.6	9.7	9.8	9.7
Salgueiro	9.9	9.9	9.9	10	9.925
Unidos da Tijuca	10	10	9.9	9.9	9.95
União da Ilha	9.8	9.7	9.8	9.8	9.775
Mocidade	9.8	9.7	9.8	9.6	9.725
Portela	10	10	9.8	9.7	9.875
São Clemente	10	9.8	9.8	9.7	9.825
Mangueira	9.9	9.9	9.9	9.6	9.825
Beija-Flor	9.9	10	10	10	9.975
Grande Rio	10	9.8	10	9.9	9.925
Imperatriz Leopoldinense	9.9	9.9	9.9	10	9.925
Vila Isabel	10	10	10	10	10