



Universidade de São Paulo
B R A S I L

Análise estatística multivariada aplicada a processos químicos

- **Escola Politécnica**
- **Departamento de Engenharia Química**
- Prof. Dr. Cláudio Augusto Oller do Nascimento
- Prof. Dr. Roberto Guardani

■ 2007

Parte 3. ANÁLISE DE COMPONENTES PRINCIPAIS

3.1. Introdução

Componentes principais são combinações lineares das variáveis originais de processo, obtidas a partir da matriz de covariância ou de correlação dos dados. Essas novas variáveis são independentes entre si e capazes de descrever o comportamento de um dado sistema multivariado. Técnicas baseadas em análise de componentes principais (será usada a abreviatura PCA: “principal component analysis”) são utilizadas visando dois objetivos principais:

- a) Redução da dimensão de modelos do sistema, obtida com a redução do número de variáveis consideradas. Em sistemas complexos e com grande número de variáveis, essa redução é importante, porque possibilita o desenvolvimento de modelos mais simples, com menor número de variáveis, para simulação e análise

do sistema. Para tanto, os resultados da PCA são utilizados para selecionar um subconjunto de m variáveis (com $m < p$) capaz de descrever a maior parte da variabilidade dos dados de processo.

- b) Detecção de correlações entre variáveis, ou grupos de variáveis, implícitas nos dados. A detecção baseia-se na estrutura implícita da matriz de covariâncias, para identificação e interpretação das relações entre grupos de variáveis. Essas análises normalmente visam identificar processos implícitos na estrutura dos dados, não evidentes a partir do exame inicial das correlações entre pares de variáveis.

Para facilitar a compreensão do conceito envolvendo componentes principais, é apresentada a seguir uma descrição geométrica, com um exemplo simples, baseado em Sharma (1996). Na segunda parte do texto é apresentada a formulação matemática da técnica e sua aplicação.

3.2. Interpretação geométrica dos componentes principais

O significado e algumas implicações da técnica de PCA podem ser mais facilmente compreendidos com uma interpretação geométrica dos componentes, associando-os a projeções das variáveis originais. Para o caso de duas variáveis, é possível visualizar essas projeções, como no exemplo seguinte, o qual é baseado em Sharma (1996). Na Tabela 3.1 há 12 observações de duas variáveis, x_1 e x_2 . A tabela apresenta os valores originais, média e variância das variáveis, bem como os valores das variáveis centradas na média.

A variância total do sistema equivale à soma:

$$s^2 = s_1^2 + s_2^2 = 23,091 + 21,091 = 44,182$$

As variáveis x_1 e x_2 representam, respectivamente, 52,3% e 47,7% da variância total do sistema. Portanto, sua importância para a variabilidade total dos dados é aproximadamente igual. As matrizes de quadrados e produtos cruzados, covariância e de correlação são, respectivamente:

$$\text{SSCP} = \begin{bmatrix} 254 & 181 \\ 181 & 232 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 23,091 & 16,455 \\ 16,455 & 21,091 \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} 1,000 & 0,746 \\ 0,746 & 1,000 \end{bmatrix}$$

Tabela 3.1. Exemplo de dados bivariados: variáveis originais e centradas na média.

| | x_1 | X_1 | x_2 | X_2 |
|------------|----------|-------------------|----------|-------------------|
| Observação | Original | Centrada na média | Original | Centrada na média |
| 1 | 16 | 8 | 8 | 5 |
| 2 | 12 | 4 | 10 | 7 |
| 3 | 13 | 5 | 6 | 3 |
| 4 | 11 | 3 | 2 | -1 |
| 5 | 10 | 2 | 8 | 5 |
| 6 | 9 | 1 | -1 | -4 |
| 7 | 8 | 0 | 4 | 1 |
| 8 | 7 | -1 | 6 | 3 |
| 9 | 5 | -3 | -3 | -6 |
| 10 | 3 | -5 | -1 | -4 |
| 11 | 2 | -6 | -3 | -6 |
| 12 | 0 | -8 | 0 | -3 |
| média | 8 | 0 | 3 | 0 |
| variância | 23,091 | 23,091 | 21,091 | 21,091 |

Observa-se que os dados são correlacionados positivamente, com coeficiente de correlação $r_{12} = r_{21} = 0,746$. Os valores dessas duas variáveis centradas na média podem ser representados graficamente em um plano, como mostra a Figura 3.1. Assim, qualquer ponto no gráfico é representado pelas coordenadas nos eixos X_1 e X_2 . No caso de haver p variáveis, seus valores seriam pontos no espaço de dimensão p . É possível fazer uma mudança nessa base ortogonal, fazendo-se uma rotação dos eixos X_1 e X_2 em um ângulo θ , obtendo-se uma nova base ortogonal, representada pelos novos eixos de coordenadas e_1 e e_2 , conforme mostrado na Figura 3.1.

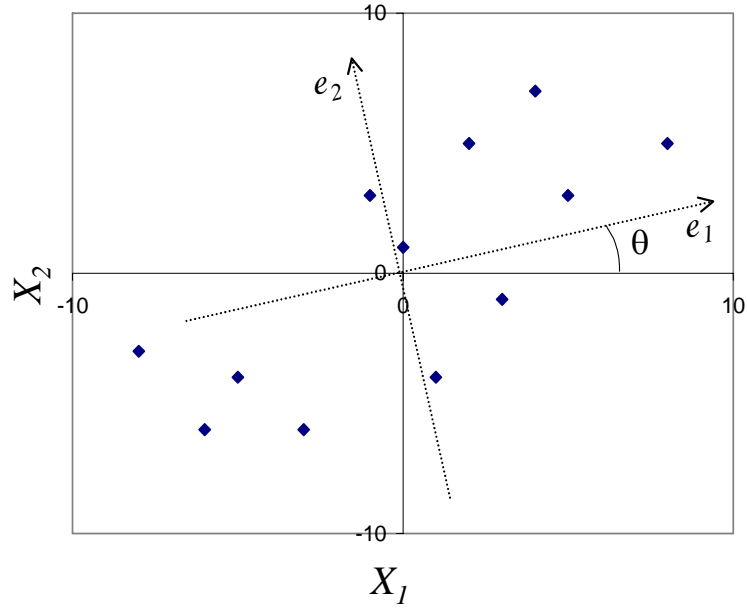


Figura 3.1. Representação gráfica das variáveis centradas na média, da Tabela 3.1, e de um novo sistema de eixos ortogonais, e_1 e e_2 , deslocados em relação a X_1 e X_2 de um ângulo θ .

Como resultado dessa mudança de base, pode-se demonstrar que uma observação i qualquer, de coordenadas $[X_{1,i}, X_{2,i}]$ no gráfico, pode ser representado nas novas coordenadas e_1 e e_2 da seguinte forma:

$$e_{1,i} = X_{1,i} \cos \theta + X_{2,i} \sin \theta \quad (3.1)$$

$$e_{2,i} = -X_{1,i} \sin \theta + X_{2,i} \cos \theta \quad (3.2)$$

Dependendo do ângulo de rotação dos eixos (θ), pode-se alinhar um dos eixos (no caso, e_1) mais ou menos em relação à direção de maior variação dos valores de X_1 e X_2 , o que corresponde à maior variância dos dados. Por exemplo, para θ igual a 10° , os valores das variáveis na nova base e as respectivas variâncias são apresentadas na Tabela 3.2. Observa-se que as duas novas variáveis, obtidas por combinações lineares das variáveis originais segundo as Equações 3.1 e 3.2, também têm média igual a zero. No entanto, embora a variância total das novas variáveis permaneça a mesma (pois o sistema é o mesmo) a fração da variância total representada por cada uma delas mudou: a variância contida em e_1 passa a representar 64,9 % e a contida em e_2 , 35,1 % da variância total. Isto é consequência da nova orientação dos eixos: e_1 está mais alinhado com a direção de maior variância dos dados.

Tabela 3.2. Efeito da mudança de base: rotação em $\theta = 10^\circ$.

| Observação | X_1 | X_2 | e_1 | e_2 |
|------------|--------|--------|--------|--------|
| 1 | 8 | 5 | 8,747 | 3,535 |
| 2 | 4 | 7 | 5,155 | 6,199 |
| 3 | 5 | 3 | 5,445 | 2,086 |
| 4 | 3 | -1 | 2,781 | -1,506 |
| 5 | 2 | 5 | 2,838 | 4,577 |
| 6 | 1 | -4 | 0,290 | -4,113 |
| 7 | 0 | 1 | 0,174 | 0,985 |
| 8 | -1 | 3 | -0,464 | 3,128 |
| 9 | -3 | -6 | -3,996 | -5,388 |
| 10 | -5 | -4 | -5,619 | -3,071 |
| 11 | -6 | -6 | -6,951 | -4,867 |
| 12 | -8 | -3 | -8,399 | -1,565 |
| média | 0 | 0 | 0 | 0 |
| variância | 23,091 | 21,091 | 28,659 | 15,523 |

Pode-se fazer um gráfico da fração da variância total representada pela nova variável e_1 em função do ângulo de rotação θ , conforme mostrado na Figura 3.2. O valor máximo da curva ocorre para θ igual a $43,26^\circ$, com a variância de e_1 igual a 38,576 (87,3% do total de 44,182). Para esse valor, as Equações 3.1 e 3.2 ficam:

$$e_{1,i} = 0,728X_{1,i} + 0,685X_{2,i} \quad (3.3)$$

$$e_{2,i} = -0,685X_{1,i} + 0,728X_{2,i} \quad (3.4)$$

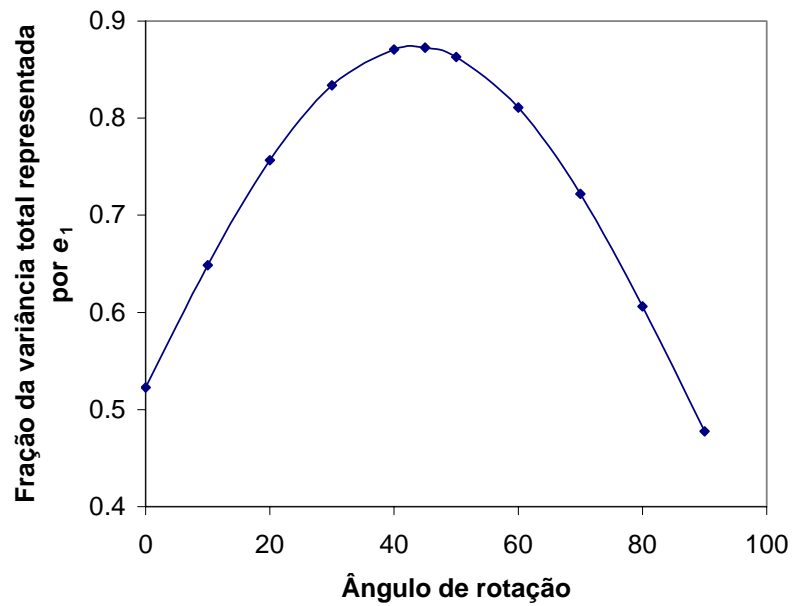


Figura 3.2. Fração da variância total representada por e_1 em função de θ .

As variâncias representadas por e_1 e e_2 são, respectivamente, 38,576 (87,3 % do total) e 5,606 (12,7 % do total). As novas matrizes **SSCP**, Σ e **R**, baseadas nas novas variáveis e_1 e e_2 , são:

$$\mathbf{SSCP} = \begin{bmatrix} 424,334 & 0 \\ 0 & 61,666 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 38,576 & 0 \\ 0 & 5,606 \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} 1,000 & 0 \\ 0 & 1,000 \end{bmatrix}$$

Nas novas matrizes, a covariância entre e_1 e e_2 vale zero, refletindo o fato de que as novas variáveis são ortogonais (independentes entre si). A variância total permanece a mesma das variáveis originais. No entanto, a nova variável e_1 representa quase 90 % da variância total e a outra variável incorpora o restante da variância total dos dados experimentais. Assim, a nova variável e_1 pode ser usada para representar as duas variáveis originais, com a vantagem de que se reduz a dimensão do problema de 2 para 1. A desvantagem é que cerca de 13% das informações contidas nos dados originais é confundida, ou perdida, quando se usa apenas e_1 para representar o sistema.

O exemplo dado pode ser extrapolado para p variáveis. A análise de componentes principais é usada com freqüência na redução de dimensão em sistemas multivariados, visando concentrar em um pequeno conjunto de novas variáveis, ortogonais, a maior parte da variância do sistema.

3.3. Formulação matemática da técnica de PCA

A obtenção de componentes principais consiste em se obterem os pesos w_{kj} , k e j variando de 1 a p , para o seguinte sistema de p equações:

$$\begin{aligned} e_1 &= w_{11}X_1 + w_{12}X_2 + \dots + w_{1p}X_p \\ e_2 &= w_{21}X_1 + w_{22}X_2 + \dots + w_{2p}X_p \\ &\vdots \\ e_p &= w_{p1}X_1 + w_{p2}X_2 + \dots + w_{pp}X_p \end{aligned} \quad (3.5)$$

em que:

e_1 a e_p são os componentes principais, ordenados em ordem decrescente da sua participação na variância dos dados;

w_{kj} são os pesos da variável j sobre o componente k ("component loadings");

Os pesos são estimados de modo que as seguintes restrições sejam atendidas:

- 1) O primeiro componente, e_1 , abrange a máxima fração da variância total do sistema; o segundo componente, e_2 , abrange a máxima fração do restante da variância total do sistema não abrangido pelo primeiro componente; e assim sucessivamente, obtendo-se uma seqüência de componentes, e_1 a e_p , em ordem decrescente de importância (o número máximo de componentes é igual a p).
- 2) $w_{k1}^2 + w_{k2}^2 + \dots + w_{kp}^2 = 1$, $k = 1, \dots, p$. Esta condição, que significa que os p vetores são normalizados, é arbitrária, mas é adotada para evitar que os valores numéricos dos pesos afetem a variância representada por um dado componente em relação aos demais.
- 3) Ortogonalidade: $w_{k1} \cdot w_{j1} + w_{k2} \cdot w_{j2} + \dots + w_{kp} \cdot w_{jp} = 0$, $k \neq j$.

O objetivo dos algoritmos de cálculo de componentes principais é obter w_{kj} , respeitando as restrições 1 a 3.

Para o vetor \mathbf{X} das p variáveis X_j , centradas na média, a matriz de covariância Σ é representada pelo valor esperado do produto dos vetores:

$$\Sigma = E(\mathbf{X} \cdot \mathbf{X}^T) \quad (3.6)$$

Seja o vetor de pesos de um componente e , qualquer, representado por:

$$\mathbf{w}^T = (w_1, w_2, \dots, w_p) \quad (3.7)$$

Conforme mostra a Eq. 3.5, existem p vetores como esse, e cada componente e pode ser representado por:

$$e = \mathbf{w}^T \cdot \mathbf{X} \quad (3.8)$$

Portanto, para cada uma das n observações há um valor correspondente de cada componente. A variância do componente e é calculada pelo valor esperado do produto quadrático desse componente para todas as observações, ou seja:

$$s_e^2 = E(e \cdot e^T) = E(\mathbf{w}^T \mathbf{X} \cdot \mathbf{X}^T \mathbf{w}) = \mathbf{w}^T \Sigma \mathbf{w} \quad (3.9)$$

A obtenção dos componentes principais a partir de um conjunto de dados com n observações de p variáveis, cuja matriz de covariâncias é Σ , implica resolver o seguinte sistema para cada componente e_k , com k variando de 1 a p :

Obter os vetores \mathbf{w}_k , tais que a variância de e_k , $\mathbf{w}_k^T \Sigma \mathbf{w}_k$, seja máxima,

Sujeita à restrição: $\mathbf{w}_k^T \mathbf{w}_k = 1$ (restrição 2) e

Atendendo à condição de ortogonalidade: $\mathbf{w}_i^T \mathbf{w}_k = 0$ para $i \neq k$ (restrição 3).

A solução consiste, portanto, em um problema de otimização, com a variância como função objetivo a ser maximizada. Uma das técnicas de solução consiste no uso de multiplicadores de Lagrange para incorporar as restrições à função objetivo. Para o caso de um componente qualquer, com vetor de coeficientes \mathbf{w} , seja F a função a ser maximizada, definida por:

$$F = \mathbf{w}^T \Sigma \mathbf{w} - \lambda (\mathbf{w}^T \mathbf{w} - 1) \quad (3.10)$$

em que a restrição de vetor de pesos normalizado está incorporada em F no segundo termo, que é multiplicado pelo multiplicador de Lagrange, λ .

O vetor das derivadas parciais de F em relação ao vetor de pesos é:

$$\frac{\partial F}{\partial \mathbf{w}} = 2\Sigma \mathbf{w} - 2\lambda \mathbf{w} \quad (\text{vetor } p \times 1) \quad (3.11)$$

Fazendo-se esse vetor igual a zero, obtém-se:

$$\Sigma \mathbf{w} = \lambda \mathbf{w} \quad (3.12)$$

que é um problema característico de autovalores e autovetores. Assim, a solução não trivial (para $\mathbf{w} \neq \mathbf{0}$) da Equação 3.12 envolve a obtenção dos p autovalores λ_k , $k = 1, 2, \dots, p$, da matriz de covariância e dos correspondentes autovetores, os quais correspondem aos p vetores de pesos \mathbf{w}_k , para $k = 1, 2, \dots, p$. Por exemplo, seja $(\lambda, \mathbf{w})_k$ um par de autovalor – autovetor da matriz de covariâncias. Então, da Eq. 3.12:

$$(\mathbf{\Sigma} - \lambda \mathbf{I})\mathbf{w} = 0 \quad (3.13)$$

Como foi imposta a restrição de que os autovetores são normalizados, então:

$$\mathbf{w}^T \cdot \mathbf{w} = 1 \quad (3.14)$$

Pré-multiplicando-se a Eq. 3.13 por \mathbf{w}^T obtém-se:

$$\mathbf{w}^T (\mathbf{\Sigma} - \lambda \mathbf{I})\mathbf{w} = 0 \quad (3.15)$$

ou:

$$\mathbf{w}^T \mathbf{\Sigma} \mathbf{w} = \lambda \cdot \mathbf{w}^T \mathbf{w} . \quad (3.16)$$

Combinando-se as Eqs. 3.9, 3.14 e 3.16, obtém-se:

$$s_e^2 = E(e \cdot e^T) = E(\mathbf{w}^T \mathbf{X} \cdot \mathbf{X}^T \mathbf{w}) = \mathbf{w}^T \mathbf{\Sigma} \mathbf{w} = \lambda . \quad (3.17)$$

Ou seja, a variância do componente é igual ao autovalor correspondente da matriz de covariâncias. A matriz de covariâncias é positiva definida, o que implica que seus autovalores são positivos. Assim, para a obtenção dos componentes principais em ordem decrescente de importância quanto à variabilidade dos dados, ordenam-se os autovalores da matriz $\mathbf{\Sigma}$ em ordem decrescente, de modo que:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0 \quad (3.18)$$

e calculam-se os correspondentes autovetores, que são os pesos dos componentes principais, em ordem decrescente de importância do ponto de vista da variância abrangida. Assim, o vetor de pesos do primeiro componente, \mathbf{w}_1 , é o autovetor correspondente ao maior autovalor, λ_1 , da matriz de covariâncias, $\mathbf{\Sigma}$. Se todos os autovalores da matriz de

covariâncias forem distintos, então os autovetores são ortogonais, o que satisfaz à restrição 3. Assim, obtêm-se os pares $(\lambda, \mathbf{w})_k$, $k = 1, 2, \dots, p$, que são as variâncias e os respectivos coeficientes dos p componentes principais e_k , $k = 1, 2, \dots, p$. A soma dos p autovalores corresponde à variância total do sistema.

Pode ocorrer, em situações raras, de haver igualdade entre dois ou mais dos autovalores. Nesses casos, embora os autovetores de Σ possam ser ortogonais, eles não são únicos, exigindo a definição de seus valores por algum critério arbitrário. No entanto, essa situação normalmente não ocorre em situações reais, com grandes bases de dados.

Outra situação que pode ocorrer, mais comum, é haver um ou mais autovalores iguais a zero. Tal situação indica que há dependência linear entre elementos da matriz de dados, \mathbf{x} , causado por redundância de variáveis ou observações. Matematicamente, tal fato indica que o *rank* da matriz Σ é menor que p . Nesses casos, a base de dados deve ser modificada e reorganizada para aplicação da técnica.

3.4. Análise baseada na matriz de covariâncias ou de correlação.

As matrizes Σ e \mathbf{R} são relacionadas entre si e fornecem basicamente informações sobre o grau de associação linear entre variáveis. No entanto, uma vez que a matriz de correlação é obtida a partir das variáveis padronizadas, perde-se informação sobre a variância em troca da informação sobre a correlação entre as variáveis. Com isso, os elementos das matrizes Σ e \mathbf{R} são diferentes, fazendo com que os autovalores e autovetores sejam, também, diferentes, o que altera os resultados da análise. A escolha entre uma e outra matriz para realizar a PCA depende normalmente da natureza e valores numéricos dos dados. Nos casos em que se trabalha com variáveis cujos valores numéricos são muito diferentes entre si (por exemplo, vazões da ordem de milhares de litros por hora e frações molares), os resultados da análise são afetados pela amplitude dos valores. Nesses casos, é melhor utilizar a matriz de correlação dos dados, pois isso faz com que a análise considere as variáveis padronizadas, ou seja, levando em consideração a amplitude em múltiplos de desvio padrão de cada variável centrada na média.

3.5. Interpretação de resultados

Como visto, a técnica de PCA consiste na obtenção de um conjunto de p combinações lineares, ou componentes, das p variáveis originais, combinações essas ordenadas em ordem decrescente da variância representada por cada componente. A orientação dessas combinações lineares de modo a absorverem, cada uma, a máxima

variância disponível implica obter os p autovalores e correspondentes autovetores da matriz de covariância, ou da matriz de correlação. Uma vez que a técnica de PCA destina-se a reduzir a dimensão de sistemas com muitas variáveis e detectar correlações entre elas, o tratamento dos resultados da análise visa tornar evidentes as eventuais correlações e a importância relativa dos componentes obtidos e das variáveis originais. Na interpretação dos resultados, objetiva-se a seleção de um subconjunto de m componentes ou variáveis ($m < p$), o que envolve normalmente um ou mais critérios, que variam conforme a aplicação.

Crítérios para seleção de um subconjunto de componentes principais

Nos casos em que o objetivo é a seleção de m componentes principais, na maioria dos casos o objetivo é a utilização do conjunto de m componentes em modelos matemáticos dos sistemas em questão. É comum ocorrerem casos em que $m < p$, como, por exemplo, em análises de espectros, em que as variáveis podem ser a absorbância de radiação em um dado comprimento de onda e comprimentos de onda próximos são, em geral, correlacionados. Assim, os modelos são construídos de forma a correlacionar a concentração de espécies químicas de interesse com um grupo de m componentes principais, muito menor que as p variáveis originais, o que representa uma grande vantagem, pois sistemas de medição de espectros com alta resolução normalmente operam com centenas ou milhares de comprimentos de onda, ou seja, grandes valores de p . Assim, os cálculos de concentração baseados nesses modelos envolvem duas etapas: 1) a partir dos valores das p variáveis originais, calcular os valores dos m componentes ("component scores"); 2) calcular as variáveis de saída do modelo a partir dos m componentes. Os critérios mais comumente adotados para selecionar os m primeiros componentes a serem utilizados para descrever um dado sistema multivariado são:

a) valor de corte para o autovalor correspondente a cada componente, ou variância representada por cada componente.

Este critério é adequado quando a análise é baseada na matriz de correlação, \mathbf{R} . A justificativa para esta regra é que, se as variáveis incluídas no estudo são independentes, todas têm correlação zero e variância individual igual a 1, na matriz de correlação. Desse modo, um componente cuja variância seja menor que 1 representa menos do que uma variável original individual e não precisa ser considerado. O critério, portanto, é:

Selecionar e_k se $\lambda_k \geq 1$, para $k = 1, \dots, p$.

Na prática, é interessante considerar uma tolerância no valor de corte, selecionando componente cujas variâncias sejam um pouco menores que 1, para levar em consideração

variações e erros em medidas. Por exemplo, Jolliffe (1986), baseado em simulações, sugere o valor de corte igual a 0,7.

b) fração acumulada da variância total do sistema, representada pelos primeiros componentes, de 1 a m .

Este critério consiste em se estabelecer um valor mínimo aceitável da fração da variância total representada pelos m primeiros componentes. Normalmente adotam-se valores de cerca de 80 a 90%. Assim, o critério é:

$$Frac.\lambda_m = \frac{\sum_{k=1}^m \lambda_k}{\sum_{k=1}^p \lambda_k}, \quad \text{para } m = 1, \dots, p \quad (3.19)$$

o processo é interrompido quando $Frac.\lambda_m \geq$ fração mínima adotada.

c) visualização por meio de gráficos de distribuição.

Gráficos do tipo “scree plot” são úteis para visualização da importância dos componentes e são usados para decisão quanto ao número de componentes a serem utilizados para descrever um sistema multivariado. Na Figura 3.3 é mostrado um exemplo de gráfico do autovalor de cada componente em função de sua ordem, ou seja, a variância explicada por cada componente, na ordem de importância. A mesma informação pode ser representada na forma da fração da variância total explicada pelos componentes ordenados (também mostrada na Fig. 3.3). A seleção do número m de componentes principais a serem considerados em um dado estudo baseia-se na inclinação do gráfico: a partir de um certo número, a curva passa a ter inclinação praticamente constante e pequena. O número m dos primeiros componentes a serem selecionados no estudo corresponde à região de maior inclinação da curva (à esquerda), antes do ponto em que há alteração visível na inclinação.

Gráficos como o mostrado na Fig. 3.3 são importantes por possibilitarem visualizar a distribuição da variância entre os componentes. Nos casos em que o gráfico tem forma similar ao mostrado na figura, é possível visualizar que os primeiros 3 a 4 componentes são responsáveis pela maior parte da variância e que o restante da variância está distribuído entre os demais componentes. No entanto, há casos em que não é possível concentrar a maior parte da variância nos primeiros componentes, ocorrendo uma queda gradual na curva do gráfico, indicando que a variância do sistema está distribuída entre os componentes. Tais situações ocorrem nos casos em que não há correlação clara entre

variáveis, ou grupos de variáveis. Nesses casos, não há ganhos significativos em substituir as variáveis originais por componentes principais.

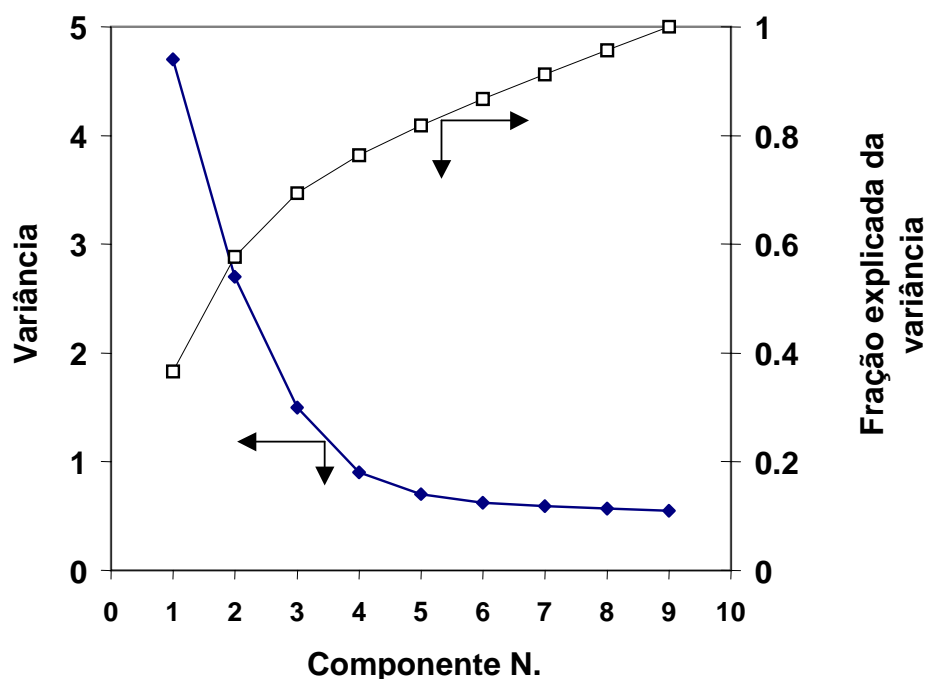


Figura 3. Exemplo de um “scree plot”.

Crítérios para seleção de um subconjunto de variáveis

A substituição das variáveis originais por suas combinações lineares, na forma de componentes principais, muitas vezes não é desejada. Isto ocorre nos casos em que as variáveis são de natureza diferente, como em processos químicos, em que variáveis representando vazão, temperatura, pressão, composição, nível e outras são combinadas nos componentes principais. Nesses casos, o uso de componentes principais significa a perda do significado físico das variáveis envolvidas no processo. Então, a redução da dimensão do sistema envolve a seleção de um subconjunto de m variáveis, as quais são as mais importantes, com base na sua contribuição para a variância dos dados. Em um determinado componente, algumas variáveis podem contribuir positiva ou negativamente, sendo essa contribuição de maior ou menor intensidade. A contribuição de cada variável pode ser verificada a partir dos coeficientes das variáveis nos componentes principais (“component loadings”). Assim, a seleção das m variáveis mais importantes no sistema envolve, primeiramente, a obtenção dos componentes principais. Uma vez obtidos os componentes principais, a importância de algumas variáveis em relação a outras pode ser

avaliada a partir dos valores absolutos de seus coeficientes em cada um dos componentes principais, levando-se em consideração a importância relativa destes.

O procedimento para seleção consiste em, para cada componente e_k selecionado entre os m primeiros, $k = 1, 2, \dots, m$, selecionar a variável j para a qual $|w_{kj}|$ seja máximo, desde que essa variável não tenha sido selecionada antes (para valores menores de k). Assim, começando no primeiro componente ($k = 1$), a variável cujo coeficiente tem maior valor absoluto é aquela mais importante do sistema, quanto à contribuição para a variância. Essa variável deve, então, ser selecionada para participar de modelos. Em geral há outras variáveis em um mesmo componente cujos coeficientes têm valores absolutos próximos. Isso indica que essas variáveis são correlacionadas entre si. Portanto, com a seleção daquela com maior coeficiente essas outras correlacionadas passam a ser representadas. O procedimento é repetido para $k = 2, \dots, m$. No final do processo, haverá um conjunto de m variáveis que representam, em cada componente, a maior contribuição para a variância do sistema.

Exemplo

Realizar a PCA de um sistema com 3 variáveis aleatórias, x_1 , x_2 e x_3 , com a seguinte matriz de covariâncias:

$$\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Os autovalores e autovetores (normalizados) dessa matriz são:

$$\lambda_1 = 5,83; \quad \mathbf{w}_1^T = [0,383; -0,924; 0]$$

$$\lambda_2 = 2,00; \quad \mathbf{w}_2^T = [0; -0; 1]$$

$$\lambda_3 = 0,17; \quad \mathbf{w}_3^T = [0,924; 0,383; 0]$$

Assim, a variância total, que é igual à soma dos elementos da diagonal principal da matriz de covariâncias, equivale também à soma dos 3 autovalores, ou seja, 8. Os componentes principais são:

$$e_1 = 0,383x_1 - 0,924x_2$$

$$e_2 = x_3$$

$$e_3 = 0,924x_1 + 0,383x_2$$

Esses três componentes são ortogonais, ou seja, constituem variáveis independentes no sistema. No primeiro componente, que abrange 72,9% da variância total ($= 5,83/8$), a variável que apresenta maior peso (em módulo) é x_2 . Essa é, portanto, a principal variável que afeta a variância total do sistema. O segundo componente, responsável por 25% da variância total ($= 2/8$), é composto pela própria variável x_3 , o que é coerente, pois na matriz de covariância observa-se que essa variável não é correlacionada com as demais. O terceiro componente, que representa apenas cerca de 2% da variância total ($= 0,17/8$), pode ser desprezado, pois a perda de informações é pequena. Assim, o sistema pode ser descrito com o uso dos 2 componentes principais, ou das variáveis x_2 e x_3 , em vez das três variáveis originais.

Pode-se usar o mesmo procedimento com a matriz de correlação, No entanto, os resultados normalmente são diferentes quando os valores numéricos das variáveis são diferentes.
