

## Uma Técnica para a Redução da Cardinalidade de Padrões de Entrada de uma Rede Neural Artificial

Roberto K. H. Galvão<sup>1</sup>, Teresa C. B. Saldanha<sup>2</sup>

Takashi Yoneyama<sup>3</sup>, Mário César U. de Araújo<sup>4</sup>

<sup>1,3</sup>CTA - ITA - Div. Eng. Eletrônica - 12228-900 - São José dos Campos/SP

<sup>2,4</sup>UFPB - Dep. Química, Cid. Universitária, Campus I - C.P. 5093 - 58051 - 970 - João Pessoa/PB

E-mails: <sup>1</sup>kawakami@ele.ita.cta.br, <sup>3</sup>takashi@ele.ita.cta.br, <sup>2,4</sup>laqa@quimica.ufpb.br

### Abstract

*This paper proposes a new technique for reducing the cardinality of input patterns of artificial neural networks. The algorithm, denominated "Successive Projections" (SPA), tries to extract, from a set of vectors, a subset whose elements are maximally independent. Beginning from a conveniently chosen point, the SPA picks, at each step, the vector which has the largest projection in the subspace orthogonal to the vectors already selected. The technique was applied to the problem of choosing the best wavelengths for Simultaneous Multicomponent Analysis via Molecular Absorption Spectrophotometry. The prediction capability of the resulting model is similar to that of models employing Principal Component Analysis. However, due to the Parsimony Principle, the model obtained via SPA may be indicated as more adequate.*

### 1. Introdução

Um tópico de grande importância na área de Redes Neurais Artificiais (RNA) consiste na redução da cardinalidade dos padrões a serem apresentados à rede. Tal procedimento, ao diminuir o número de pesos sinápticos na camada de entrada, permite obter estruturas mais parcimoniosas, o que minimiza problemas de "data overfitting" e melhora a capacidade de generalização da rede. Com efeito, o Princípio da Parcimônia, dita que, "dados dois modelos com igual capacidade de previsão, é preferível adotar o mais simples" [1].

Entre as ferramentas convencionais usadas para esse propósito, pode-se citar, por exemplo, a Análise de Componentes Principais, a Transformada de Karhunen-Loève, a Transformada Discreta de Cosseno e, mais recentemente, a Transformada Wavelet [2]. Tais técnicas mapeiam os padrões para um espaço onde a informação de relevância pode estar contida em um número reduzido de coeficientes. Uma possível dificuldade de tal abordagem é que nem sempre os coeficientes obtidos guardam uma relação física direta com os dados originais. Desse modo, a interpretação do conhecimento embutido em uma RNA que opera sobre tais coeficientes torna-se mais complexa.

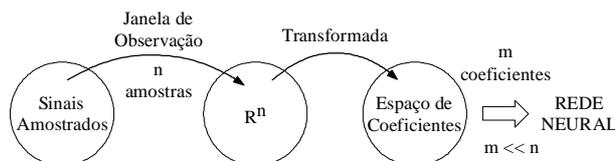


Fig. 1: Redução de Dados usando uma Transformação Intermediária

Neste artigo, é proposta uma metodologia alternativa que não emprega transformações intermediárias no processo de redução de dados. O objetivo consiste em se extrair, de uma matriz com  $m$  padrões de treinamento, cada um de comprimento  $n$ , um número pré-fixado de linhas que sejam maximamente independentes entre si. O procedimento utilizado, denominado "Projeções Sucessivas" (APS), é sub-ótimo, mas resultados numéricos sugerem que a seleção de variáveis realizada é próxima da ótima.

O algoritmo de "Projeções Sucessivas" é testado em um problema de análise química multicomponente simultânea via espectrofotometria de absorção molecular. Neste caso, as variáveis envolvidas são as absorbâncias medidas em diferentes comprimentos de onda.

### 2. O problema de redução de dados dentro do contexto de mínimos quadrados

O problema de redução de dados pode ser ilustrado mais facilmente com uma rede neural linear. Redes lineares efetuam um mapeamento entrada-saída da forma:

$$Y = WX \quad (1)$$

onde:  $Y$  = vetor de saídas ( $s \times 1$ )  
 $X$  = vetor de entradas ( $n \times 1$ )  
 $W$  = matriz de pesos sinápticos ( $s \times n$ )

Cada linha do vetor  $X$  corresponde a uma variável de entrada da rede.

Dadas matrizes de treinamento ("calibração")  $X_{cal}$  e  $Y_{cal}$  com  $m$  colunas (padrões) cada uma, a solução para  $W$  que minimiza o erro de saída no sentido de mínimos-quadrados é dada por [3]:

$$W = Y_{cal} X_{cal}^T (X_{cal} X_{cal}^T)^{-1} \quad (2)$$

onde  $Xcal^T(XcalXcal^T)^{-1}$  é chamada pseudo-inversa de Moore-Penrose da matriz  $Xcal$ .

Note-se que, se existirem elevadas correlações entre as linhas de  $Xcal$  (colinearidade), o cálculo da pseudo-inversa pode ser numericamente mal-condicionado. Desse modo, é desejável realizar uma etapa de redução de dados, de modo a usar apenas as variáveis (linhas de  $Xcal$ ) que sejam realmente “informativas”.

### 3. O algoritmo das projeções sucessivas

O algoritmo a ser descrito nesta seção procura selecionar, de um conjunto de vetores, um subconjunto cujos elementos sejam maximamente independentes entre si. Na presente aplicação, os vetores em questão são as  $n$  linhas da matriz  $Xcal$ , cada uma associada a uma certa variável. A  $i$ -ésima linha de  $Xcal$  será designada por  $x^i$ , ou seja:

$$Xcal = [x^1 \ x^2 \ \dots \ x^i \ \dots \ x^n]^T \quad (3)$$

O algoritmo segue um procedimento recursivo em que, partindo de um vetor inicial, é selecionado um novo vetor a cada iteração. Os passos do algoritmo estão descritos na fig.2. Supõe-se que se deseja reduzir o número de variáveis de  $n$  para  $N$ , incluindo aquela associada ao vetor de partida.

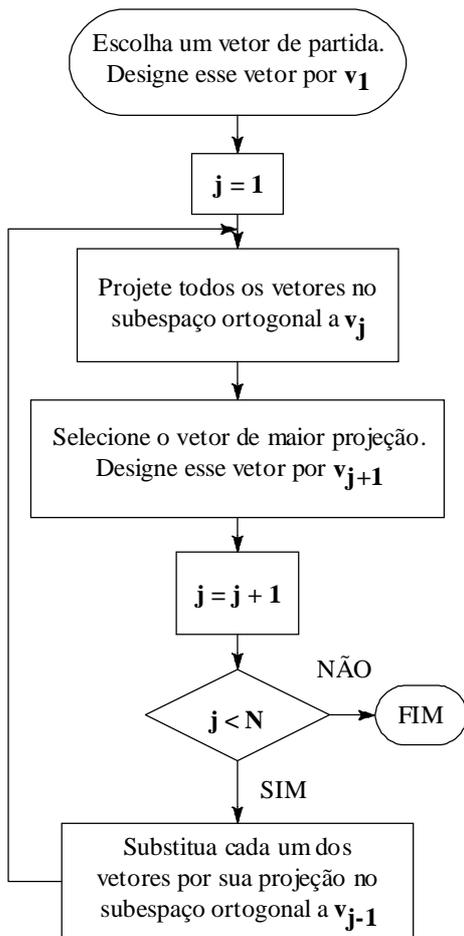


Fig.2: Algoritmo das Projeções Sucessivas

Ao final do processo, as variáveis escolhidas são aquelas associadas aos vetores que foram selecionados.

### 3.1. Exemplo

Suponha que, dada uma matriz  $Xcal_{(5 \times 3)}$  (isto é,  $m = 3$  padrões de treinamento com  $n = 5$  variáveis cada um) deseje-se selecionar 3 variáveis. A figura abaixo apresenta esquematicamente essa situação, em que cada linha de  $Xcal$  é representada por um vetor em um espaço de 3 dimensões. Sem perda de generalidade, toma-se a terceira linha ( $x^3$ ) como ponto de partida, fazendo  $v_1 = x^3$ .

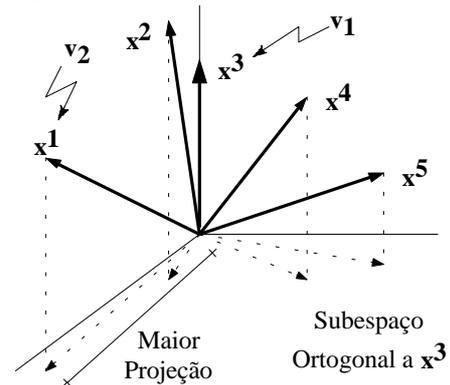


Fig.3: O vetor  $x^3$  foi escolhido como ponto de partida

Como se vê na figura, o vetor  $x^1$  tem a maior projeção no subespaço ortogonal a  $v_1$ . Portanto, faz-se  $v_2 = x^1$ .

Substituindo os vetores por suas projeções, tem-se a figura abaixo (note que a projeção de  $x^3$  tem norma zero).

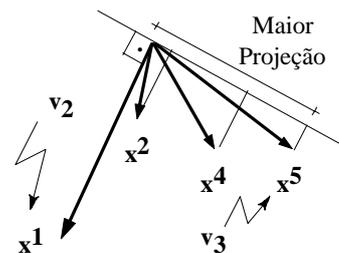


Fig.4: Cada vetor foi substituído por sua projeção no subespaço ortogonal a  $x^3$

Para selecionar o 3º vetor, repetem-se as projeções, dessa vez sobre o subespaço ortogonal a  $v_2$ . Como resultado, o vetor  $x^5$  é escolhido, fazendo-se  $v_3 = x^5$ .

Tendo sido selecionados 3 vetores, o algoritmo é encerrado, e são escolhidas as variáveis 1, 3 e 5. Note-se que não seria possível escolher uma quarta variável por esse método, pois todos os vetores resultantes de uma nova projeção teriam norma zero. Assim, o algoritmo só permite selecionar tantas variáveis quantos forem os padrões de treinamento.

Cumpra-se notar que o APS guarda uma semelhança com o conhecido processo de ortogonalização de Gram-Schmidt [4]. Contudo, este último gera **um novo conjunto de vetores** a partir do conjunto inicial, o que não é o objetivo do APS.

Ressalte-se ainda que o APS é uma técnica do tipo “blind”, uma vez que **as saídas desejadas  $Y_{cal}$  não são empregadas**.

### 3.2. Escolha do vetor de partida

Para sistematizar o procedimento como um todo, resta estabelecer um critério para a escolha do vetor de partida, caso o mesmo não possa ser selecionado por inspeção. Tal procedimento está descrito no fluxograma da fig.5, que emprega o conceito de número de condição como uma medida do grau de dependência das linhas de uma matriz (vide apêndice B)

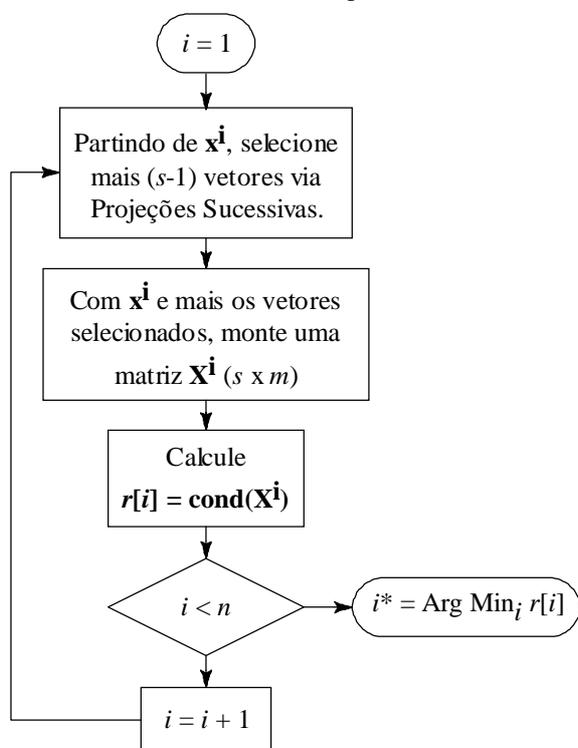


Fig.5: Procedimento para a determinação do vetor de partida para o algoritmo das Projeções Sucessivas

A idéia se baseia na seguinte heurística: sendo  $s$  o número de saídas da rede, a matriz  $X_{cal}$  deve ter no mínimo  $s$  linhas para que se possa determinar a matriz de pesos  $W$ . O APS permite que se associe univocamente, a cada uma das variáveis, um número de linhas de  $X_{cal}$  menor ou igual ao número de padrões de treinamento ( $m$ ). Caso  $m > s$ , pode-se então associar, a cada uma das variáveis, uma matriz formada pela respectiva linha de  $X_{cal}$  e mais  $(s-1)$  outras linhas selecionadas via APS. Toma-se, então aquela variável que conduziu à matriz  $(s \times m)$  com o menor número de condição (isto é, com o menor grau de colinearidade entre suas linhas). Note-se que o ideal seria gerar

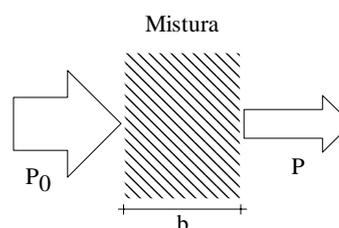
matrizes de  $N$  linhas ( $N =$  número de variáveis a serem selecionadas. Note-se que  $N > s$ ) nesse procedimento. Contudo isso seria computacionalmente custoso. A alternativa adotada é uma heurística que permite realizar de maneira judiciosa a seleção das  $s$  primeiras variáveis.

Note-se que, para determinar a melhor seleção de variáveis por busca exaustiva, seria necessário calcular o  $n^0$  de condição de  $\{n! [(n-N)! N!]^{-1}\}$  matrizes.

### 4. Um exemplo de aplicação: Calibração de análises multicomponentes simultâneas

A área de Química Analítica estuda métodos para a determinação, qualitativa e quantitativa, da composição química de um material [5]. Os *métodos clássicos* empregam reações químicas para extrair os componentes a serem identificados (*analitos*), por exemplo na forma de um precipitado ou de um gás. Alternativamente, existem os chamados *métodos instrumentais*, que empregam medidas de propriedades físicas, tais como condutividade, absorvância ou fluorescência, para realizar a análise de modo indireto. Por dispensarem etapas intermediárias, tais como secagem, destilação e pesagem, tais métodos são mais rápidos e precisos que as técnicas clássicas.

Um importante método de análise instrumental é a espectrofotometria de absorção molecular (EAM). Nesta técnica, a concentração de um analito é determinada como uma função da parcela de radiação luminosa que, passando através da mistura, é absorvida (fig.6)



$$A = \log(P_0/P)$$

Fig.6: Medida da Absorvância (A) de uma mistura.

$P_0$  = potência da radiação incidente;  $P$  = potência transmitida;  $b$  = percurso óptico da radiação.

De acordo com a *lei de Beer* [5], a relação entre a absorvância e a concentração  $C$  do analito é linear:

$$A = \epsilon b C \quad (4)$$

onde  $b$  é usualmente constante e depende das características morfológicas do aparelho utilizado. A constante  $\epsilon$  é denominada *absortividade molar*.

A análise multicomponente (mais de uma espécie a ser analisada) por EAM é tradicionalmente realizada utilizando-se um reagente específico para cada analito

ou separando o mesmo de seus interferentes. Todavia, nem sempre se dispõe de um reagente específico para cada analito, e a separação de elementos interferentes presentes na amostra é, em geral, trabalhosa. As pesquisas sobre análise multicomponente simultânea por espectrofotometria de absorção molecular (AMS-EAM) desenvolveram-se no sentido de evitar a exaustiva tarefa de separação de interferentes e de permitir a determinação simultânea de um número cada vez maior de analitos, favorecendo, conseqüentemente, a redução do tempo e do custo das análises.

A AMS-EAM emprega medidas de absorbância em um número de comprimentos de onda no mínimo igual ao número de espécies a serem analisadas. Neste contexto, o chamado procedimento de calibração multivariada consiste em levantar um modelo que relacione as absorbâncias medidas com as concentrações dos analitos. Para tal, é requerida a análise prévia de uma série de misturas de calibração (série de treinamento), cuja composição seja conhecida.

No procedimento de calibração, a escolha das variáveis espectrais (comprimentos de onda) que resultem na máxima exatidão ainda é um ponto crítico da AMS-EAM, principalmente quando ocorre alta sobreposição espectral. Sabe-se que a precisão dos resultados normalmente aumenta com o número de variáveis empregadas. Todavia, isso também pode causar uma diminuição da exatidão, em virtude da possível introdução de ruído. Por esse motivo, a utilização da faixa espectral completa não é recomendada para se obter as melhores estimativas das concentrações. Assim sendo, a maior parte dos trabalhos sobre AMS-EAM enfatiza a necessidade de selecionar os comprimentos de onda analíticos. A literatura registra diversos critérios e metodologias para escolha das variáveis espectrais mais informativas tais como "generalized simulating annealing" [5], "stepwise elimination" [6], algoritmos genéticos [6] e outros.

Note-se que, se é adotado um modelo linear da forma (a linearidade aqui decorre da lei de Beer):

$$Y = WX \quad (5)$$

onde

Y = Concentrações dos analitos ( $s \times 1$ )

X = Absorbâncias medidas ( $n \times 1$ )

n = número de comprimentos de onda utilizados

s = número de analitos

então o problema de calibração pode ser resolvido via mínimos-quadrados, conforme discutido na seção 2. Nesse caso, o APS pode ser aplicado à matriz de absorbâncias das misturas de treinamento (Xcal) para determinação das variáveis espectrais mais informativas.

Neste exemplo, serão usados os dados da AMS-EAM de misturas de complexos de Co, Cu, Mn, Ni e Zn com 4-(2-piridilazo)resorcinol (PAR) na região ultravioleta (255-320 nm) [7].

## 4.1. Material e Métodos

Foram utilizadas  $m = 23$  misturas (17 misturas distintas, mais uma repetição da 17ª mistura e cinco repetições da 9ª mistura) para calibração e 12 misturas para teste. As matrizes de absorbâncias de calibração (Xcal) e de teste (Xteste) foram obtidas usando um espectrofotômetro UV-VIS HP-8453 com arranjo de fotodiodos (resolução de 1nm)

A matriz de concentração Ycal foi construída a partir de um planejamento fatorial fracionário [9], dentro dos seguintes limites [mg/L]: Co (0,05 - 0,53); Cu (0,15 - 1,60); Mn (0,10 - 1,06); Ni (0,15 - 1,59); Zn (0,15 - 1,61). A matriz Yteste foi construída tomando-se aleatoriamente concentrações dentro dos limites definidos para Ycal.

O APS foi implementado no ambiente MATLAB v. 4.2 em um microcomputador IBM-PC com processador Pentium 233Mhz, 32Mb RAM, 512kb de cache e sistema operacional Windows 95.

## 5. Resultados e discussão

Uma vez que são 5 as espécies a serem analisadas, o número mínimo de comprimentos de onda que devem ser empregados é  $N = 5$ . Para ilustrar o APS, serão selecionados 10 comprimentos de onda.

Neste problema em particular, melhores resultados foram obtidos normalizando-se cada coluna de Xcal pelo seu máximo antes de aplicar o APS. Contudo, uma vez realizado o processo de seleção, a matriz W foi determinada usando os dados de calibração originais (i.é., não normalizados).

### 5.1. Escolha do comprimento de onda inicial

A fig.7 apresenta o número de condição das matrizes ( $5 \times 23$ ) formadas a partir de cada um dos comprimentos de onda.

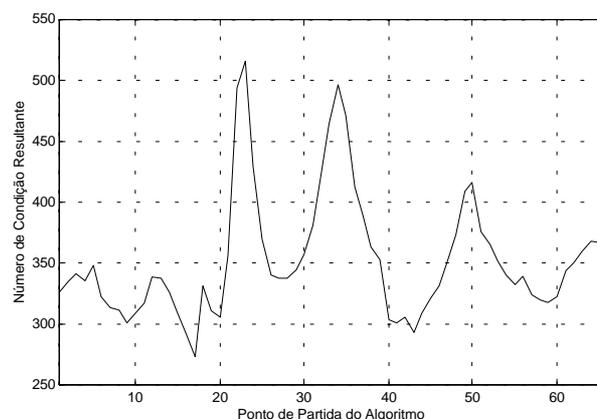


Fig.7: Escolha do comprimento de onda para iniciar o APS, usando como critério o número de condição da matriz de absorbâncias

O mínimo da curva indica o melhor comprimento de onda para iniciar o algoritmo como sendo o 17 (271nm). O tempo de processamento despendido na determinação desse comprimento de onda foi de 5.6s.

Note-se que 271nm é um comprimento de onda em que os espectros individuais encontram-se bem discriminados, como pode ser observado na fig.8.

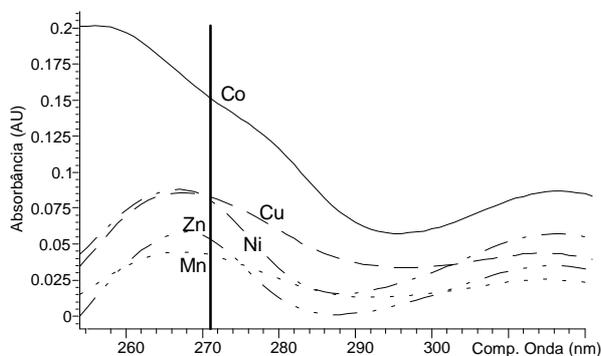


Fig.8: Espectros dos complexos individuais de Metal-PAR: região ultravioleta (255 a 320 nm)

## 5.2. Análise dos Resultados

Partindo de 271nm, foram selecionados os comprimentos de onda 255, 257, 258, 260, 267, 271, 275, 282, 295 e 320nm. De posse do comprimento de onda inicial, o APS levou 0,2 s para realizar essa seleção.

Para avaliar o quão próximo do ótimo está a seleção realizada, a fig.9 apresenta um histograma dos números de condição da matriz Xcal normalizada obtidos com 10000 seleções aleatórias de 10 comprimentos de onda. A solução encontrada pelo APS encontra-se marcada pela linha vertical (nº condição = **9111**).

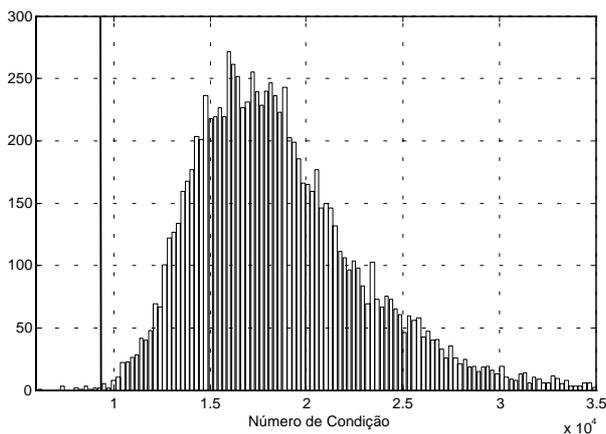


Fig.9 Histograma dos números de condição obtidos com diferentes seleções de comprimentos de onda

Das 10000 seleções realizadas:

- 15 (i.é., **0.15%** do total) resultaram em um nº de condição inferior a 9111.
- 8995 (**90.0%**) resultaram em um nº de condição superior a 1.5 vezes 9111.

O menor valor encontrado foi **6066**.

Para avaliar a adequabilidade dos comprimentos de onda selecionados, foram calculados, usando cada uma das 10000 seleções, os erros absolutos médios das previsões das concentrações das misturas de teste. O erro absoluto médio de previsão é definido como:

$$E = \frac{1}{5 \cdot 17} \sum_{i=1}^5 \sum_{j=1}^{12} |\hat{Y}(i, j) - Y_{\text{teste}}(i, j)| \quad (6)$$

onde:

$$\hat{Y} = WX_{\text{teste}} \quad (7)$$

A fig.10 apresenta o histograma dos erros absolutos médios cometidos. O erro cometido empregando os comprimentos de onda selecionados via APS encontra-se marcado com uma linha vertical (**0.057**).

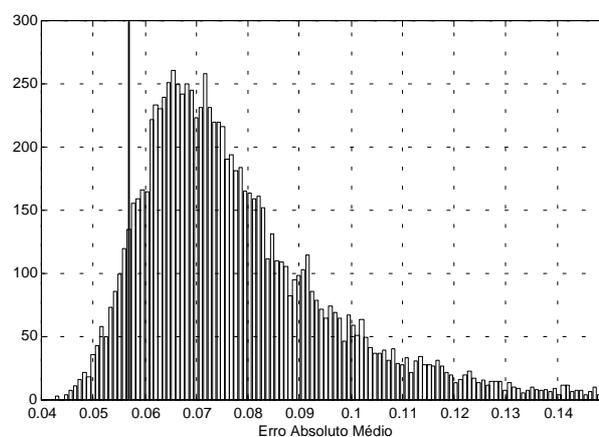


Fig.10: Histograma dos erros absolutos médios cometidos nas misturas de teste

Das 10000 seleções realizadas:

- 739 (**7.4%** do total) resultaram em um erro inferior a 0.057.
- 2 (**0.02%**) resultaram em um erro inferior a 0.75 vezes 0.057.
- 5658 (**56.6%**) resultaram em um erro superior a 1.25 vezes 0.057.

O menor erro obtido foi de **0.043**.

## 5.3. Observações

O procedimento para seleção de comprimentos de onda minimiza o erro de previsão apenas de forma indireta. Apesar disso, do ponto de vista do erro cometido na série de teste, a seleção via APS encontrou-se razoavelmente próxima do ótimo.

Para efeito de comparação, o erro médio absoluto cometido empregando PCR (“Regressão em Componentes Principais”) e PLS1 (“Mínimos Quadrados Parciais”) foi, respectivamente, **0.058** e **0.052**. Tais técnicas são padrões na área de Quimiometria [8], e usam análise de componentes principais para tratar o conjunto completo de dados. A capacidade de previsão do modelo obtido via APS mostra-se comparável à dos modelos obtidos via PCR e PLS1. Contudo, o APS

conduz a um modelo mais simples, que, por empregar um número reduzido de variáveis, dispensa a decomposição em componentes principais.

Em comparação com estratégias de seleção de variáveis que se baseiam em busca aleatória dirigida, tais como *simulated annealing* e algoritmos genéticos, o APS apresenta a vantagem de repetibilidade de cálculo e menor esforço computacional. Com efeito, sendo fixado o número de variáveis a serem selecionadas ( $N$ ) e o número de padrões de treinamento ( $m$ ), a complexidade computacional de algoritmos de busca aleatória cresce com o número de pontos do espaço de busca, que é proporcional ao **fatorial de  $n$** . No caso do APS, caso a primeira variável já tenha sido selecionada, a determinação das ( $N-1$ ) demais envolve um número de operações de projeção no espaço  $R^m$  igual a:

$$\sum_{i=1}^{N-1} (n-i) = (N-1)(n-N/2) \quad (8)$$

que cresce **linearmente** com  $n$ . Pode-se ver que, no exemplo da sub-seção 3.1, onde  $N = 3$  e  $n = 5$ , foram realizadas 7 projeções (quatro na fig.3 e mais três na fig.4).

Analisando o fluxograma da fig.5, pode-se ver que a escolha da primeira variável envolve  $n$  seleções de  $s$  variáveis. Portanto, o número de projeções realizadas é:

$$n(s-1)(n-s/2) \quad (9)$$

que cresce com  $n^2$ . Adicionalmente, são realizados mais  $n$  cálculos do número de condição de matrizes ( $s \times m$ ).

## 6. Conclusão

Neste artigo, foi proposta uma nova metodologia para redução de dados, denominada Algoritmo das Projeções Sucessivas (APS), que não emprega transformações intermediárias. Testes foram realizados em um problema de calibração de análises químicas multicomponentes simultâneas via espectrofotometria de absorção molecular (AMS-EAM).

O modelo empregando absorbâncias medidas nos comprimentos de onda selecionados via APS apresentou capacidade de previsão comparável a modelos obtidos via PCR e PLS1, que se baseiam na análise de componentes principais para tratar a informação espectral completa. Contudo, o modelo obtido via APS mostra-se mais parcimonioso, além de estabelecer uma relação direta entre as grandezas físicas (absorbâncias e concentrações).

Cumpramos ressaltar que o APS conduz sistematicamente aos mesmos resultados, ao contrário de técnicas de otimização envolvendo busca aleatória (tais como Algoritmos Genéticos e "Simulated Annealing", por exemplo). Além disso, a complexidade computacional do APS cresce apenas com o quadrado do número de variáveis do problema (ou linearmente, caso a primeira escolha seja feita por inspeção). Note-se, por fim, que a solução encontrada pode

eventualmente ser utilizada como um ponto de partida para algoritmos de busca aleatória.

## Referências

- [1] M. B. Seasholtz and B. R. Kowalski. The parcimony principle applied to multivariate calibration. *Anl. Chim. Acta* 277:165-177, 1993.
- [2] R. K. H. Galvão. Wavelet-Based Techniques for Adaptive Feature Extraction and Pattern Recognition. Tese de Doutorado, ITA, São José dos Campos, 1999.
- [3] C. L. Lawson and R. J. Hanson. Solving Least Squares Problems. Prentice-Hall, Englewood Cliffs, 1974.
- [4] E. Kreyszig. Introductory Functional Analysis with Applications. John Wiley, New York, 1978.
- [5] D. A. Skoog, F. J. Holler, and T. A. Nieman. Principles of Instrumental Analysis, 5<sup>th</sup> ed. Harcourt Brace & Co., Philadelphia, 1998.
- [6] J. H. Kalivas, N. Roberts, and J. M. Sutter. Global optimization by simulated annealing with wavelength selection for ultraviolet-visible spectrophotometry. *Anal. Chem.* 61:2024-2030, 1990.
- [7] C. B. Lucasius, M. L. M. Beckers, and G. Kateman. Genetic algorithms in wavelength selection: a comparative study. *Anal. Chim. Acta* 286: 135-153, 1994.
- [8] T. C. B. Saldanha. Uma Metodologia para Análise Espectrofotométrica Multicomponente Simultânea na Região Ultravioleta Usando o 4-(2 Piridilazo) Resorcinol (PAR) e Calibração Multivariada. Tese de Doutorado, UFPE - Depto. Química Fundamental, Recife, 1998.
- [9] B. B. Neto, I. S. Scarminio e R. E. Bruns. Planejamento e Otimização de Experimentos. Editora da UNICAMP, Campinas, 1995.

## Apêndice A: Projeções em um espaço Vetorial.

Dados dois vetores  $\mathbf{a}_{1 \times m}$  e  $\mathbf{b}_{1 \times m}$ , a projeção de  $\mathbf{b}$  na direção de  $\mathbf{a}$  é dada por:

$$\text{Proj}_{\mathbf{a}} \mathbf{b} = (\mathbf{b} \mathbf{a}^T) \mathbf{a} (\mathbf{a} \mathbf{a}^T)^{-1}$$

Portanto, a projeção de  $\mathbf{b}$  no subespaço ortogonal a  $\mathbf{a}$  é dada por:

$$\text{Proj}_{\mathbf{a}^\perp} \mathbf{b} = \mathbf{b} - \text{Proj}_{\mathbf{a}} \mathbf{b} = \mathbf{b} - (\mathbf{b} \mathbf{a}^T) \mathbf{a} (\mathbf{a} \mathbf{a}^T)^{-1}$$

## Apêndice B: Número de condição

O número de condição de uma matriz  $X_{p \times q}$  ( $p < q$ ) é definido como:

$$\text{cond}(X) = \sigma_{\max}(X) / \sigma_{\min}(X)$$

onde  $\sigma_{\max}(X)$  e  $\sigma_{\min}(X)$  são respectivamente o maior e o menor valor singular de  $A$ .

O número de condição mede o grau de singularidade de uma matriz. Por exemplo, uma matriz quadrada singular (i.é., não-inversível), por possuir um ou mais autovalores nulos, apresenta  $\sigma_{\min} = 0$  e, conseqüentemente, um número de condição infinito. Em outras palavras, quanto maior o número de condição da matriz  $X$ , maior a colinearidade entre suas linhas.