



SPOLM 2008

ISSN 2175-6295

Rio de Janeiro- Brasil, 05 e 06 de agosto de 2008.

## UM ALGORITMO DE SUAVIZAÇÃO HIPERBÓLICA PARA RESOLUÇÃO DE PROBLEMAS DE CLASSIFICAÇÃO

**Alberto de Oliveira Moreno**

Universidade do Estado do Rio de Janeiro  
Instituto de Matemática e Estatística  
Rua São Francisco Xavier, 524 – 6º andar  
Rio de Janeiro, RJ 20.550-900, Brasil  
[aomoreno@ime.uerj.br](mailto:aomoreno@ime.uerj.br)

**Adilson Elias Xavier**

Universidade Federal do Rio de Janeiro  
Programa de Engenharia de Sistemas e Computação  
P.O. Box 68511  
Rio de Janeiro, RJ 21.941-972, Brasil  
[adilson@cos.ufrj.br](mailto:adilson@cos.ufrj.br)

**Michael Souza**

Universidade Federal do Rio de Janeiro  
Programa de Engenharia de Sistemas e Computação  
[michael@cos.ufrj.br](mailto:michael@cos.ufrj.br)

### RESUMO

A formulação matemática clássica para o problema de classificação associado à análise de agrupamentos é baseada em modelos não-diferenciáveis de otimização, que são resolvidos por métodos que carecem de precisão e eficiência. No presente trabalho, propomos a resolução do problema de classificação por uma formulação suavizadora hiperbólica que conduz a um problema diferenciável, cuja solução pode ser obtida com maiores robustez e eficiência por métodos clássicos e mais poderosos de otimização. A eficiência e a precisão do método são comprovadas através de experimentos numéricos com problemas apresentados na literatura.

**Palavras-chave:** classificação, agrupamento, suavização hiperbólica.

### ABSTRACT

The classical mathematical formulation for the classification problem associated to cluster analysis is based on non differentiable optimization problems that are solved by methods that have few precision and efficiency. In this work, it is proposed to solve the classification problem through the hyperbolic smoothing technique, where the problem is reformulated in a completely differentiable form and whose solution by classic and more powerful methods can be obtained with great robustness and efficiency. We hope the proposed methodology prove an improvement both reliability and the efficiency in comparison with other alternative procedures.

**Keywords:** classification, clustering, hyperbolic smoothing.

## 1. INTRODUÇÃO

Diariamente, são produzidos e manipulados em todo o mundo um número explosivamente crescente de arquivos contendo dados sobre universos das mais diversas naturezas. No entanto, apenas a disponibilidade desses dados não é suficiente para dar-se um uso eficiente e eficaz aos mesmos. É necessário tratá-los, categorizá-los e classificá-los para promover o seu uso de forma eficiente, para afinal explorar ao máximo suas potencialidades. Assim, sobretudo nos últimos 30 anos, foram publicados inumeráveis artigos dedicados ao estudo do tratamento de informações visando o aprimoramento da sua utilização ([3], [10], [12]).

Usualmente, as técnicas de classificação são aplicadas em bases de dados cujos elementos estão divididos em classes. O objetivo dessas técnicas é caracterizar de forma sintética cada uma das classes existentes no banco de dados e, assim, simplificar o processo de classificação dos novos elementos que serão inseridos na base de dados. As técnicas de classificação envolvem a atribuição dita supervisionada, pois parte-se de um conjunto treinamento (base de dados) com elementos já classificados e podem ser aplicadas em uma grande variedade de problemas advindos dos mais diferentes contextos como, por exemplo, categorização de textos [8], reconhecimento de dígitos [7] e expressão genética [5].

O objetivo principal deste trabalho é propor um novo e eficiente método para o problema de classificação. Esse novo método é fortemente inspirado pelos trabalhos [4].

## 2. A METODOLOGIA DE CLASSIFICAÇÃO VIA SUAVIZAÇÃO HIPERBÓLICA (MCSH)

A formulação matemática clássica para o problema de classificação é baseada em modelos não-diferenciáveis de otimização, que são resolvidos por métodos que carecem de precisão e eficiência. Em nossa abordagem, o modelo não-diferenciável e não-convexo apresentado Bagirov em [4] é aproximado por uma sequência de modelos suavizados e completamente diferenciáveis, cujas soluções podem ser obtidas com maiores robustez e eficiência por métodos clássicos de otimização. Observa-se que essa estratégia tem como vantagem, não só a diferenciabilidade dos problemas considerados, que possibilita o uso de técnicas consagradas e eficientes de otimização, mas também a redução da quantidade de mínimos locais. Em consequência, nosso algoritmo pode conduzir a soluções que são mínimos locais mais profundos e, portanto, de boa qualidade.

No algoritmo proposto em [4], assume-se que os elementos da base de dados (conjunto de treinamento) estão divididos em duas classes  $C_1$  e  $C_2$ , e as iterações do algoritmo seguem a seguinte lógica:

**(Passo 0)** Faça  $k := 0$ ,  $\varepsilon > 0$ ,  $X_{1k} := X_{2k} := \emptyset$  e defina as funções

$$d(Y, a) \equiv \min_{y \in Y} \|y - a\| \quad e \quad f(Y, A) \equiv \sum_{a \in A} d(Y, a);$$

**(Passo 1)** Para  $i = 1, 2$ , determine os centros  $c_i := \arg \min_{x \in \mathbb{R}^n} f(X_{ik} \cup \{x\}, C_i)$ ;

**(Passo 2)** Defina os subconjuntos

$$K_1 := \{a \in C_1 : d(X_{1k} \cup \{c_1\}, a) \leq d(X_{2k} \cup \{c_2\}, a)\}$$
$$K_2 := \{a \in C_2 : d(X_{2k} \cup \{c_2\}, a) \leq d(X_{1k} \cup \{c_1\}, a)\};$$

**(Passo 3)** Para  $i = 1, 2$ , faça  $x_{ik} := \arg \min_{x \in \mathbb{R}^n} f(\{x\}, K_i)$  e  $X_{i,k+1} := X_{ik} \cup \{x_{ik}\}$ ;

**(Passo 4)** Se  $k > 1$  e  $\frac{|f(X_{i,k+1}, C_i) - f(X_{ik}, C_i)|}{f(X_{i1}, C_i)} < \varepsilon \quad \forall i = 1, 2$ , então pare. Caso contrário, faça  $k := k + 1$  e vá para o Passo 1.

A cada iteração, no passo 1, é necessário obter as soluções de problemas não-diferenciáveis e não-convexos de otimização da forma

$$\min_{x \in \mathbb{R}^n} f(X_{ik} \cup \{x\}, C_i) = \min_{x \in \mathbb{R}^n} \min_{a \in C_i} \{\|x - a\|, d(X_{ik}, a)\} \quad (1)$$

onde  $X_{ik}$  e  $C_i$  são conjuntos conhecidos e, portanto, para cada  $a \in C_i$ ,  $d(X_{ik}, a)$  é uma constante. Sendo assim, o problema (1) pode ser escrito sob a forma

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \sum_{a \in C_i} Z_{ik} & \quad (2) \\ \text{s. a } Z_{ik}(x, a) &= \min\{\|x - a\|, d(X_{ik}, a)\}. \end{aligned}$$

Em função disso, propomos um método de solução alternativo que consiste na resolução de uma sequência de problemas diferenciáveis de otimização que se aproximam do problema (2). Em nosso algoritmo MCSH, a função objetivo do problema (2) é substituída por suavizações hiperbólicas ([13], [11]).

Considerando a formulação (2), cada  $Z_{ik}$  deve necessariamente satisfazer as seguintes restrições de desigualdade:

$$Z_{ik} \leq \|x - a\|, \quad \forall a \in C_i; \quad (3)$$

$$Z_{ik} \leq d(X_{ik}, a), \quad \forall a \in C_i. \quad (4)$$

Substituindo as restrições de igualdade do problema (2) pelas desigualdades acima, obtém-se o seguinte problema relaxado:

$$\begin{aligned} \min_{a \in C_i} \sum_{a \in C_i} Z_{ik} & \quad (5) \\ \text{s. a } Z_{ik} &\leq \|x - a\|, \\ &Z_{ik} \leq d(X_{ik}, a), \\ &a \in C_i. \end{aligned}$$

Uma vez que as variáveis  $Z_{ik}$  são inferiormente ilimitadas, então o problema relaxado (5) também será inferiormente ilimitado. Segue que, para se obter a equivalência desejada entre os problemas (2) e (5), é necessário modificar este último. Para tal, primeiro define-se  $\varphi(y) = \max\{y, 0\}$  e observa-se que para o conjunto de desigualdades do problema (5) vale

$$\varphi(Z_{ik} - \|x - a\|) + \varphi(Z_{ik} - d(X_{ik}, a)) = 0, \quad a \in C_i. \quad (6)$$

Define-se

$$\begin{aligned} d_1 &= \min\{\|x - a\|, d(X_{ik}, a)\}, \\ d_2 &= \max\{\|x - a\|, d(X_{ik}, a)\}. \end{aligned}$$

Pela definição, tem-se  $d_1 \leq d_2$ . A Figura 1 mostra as duas parcelas de (6) em função de  $Z_{ik}$ , considerando a situação em que prevalece a relação mais frequente em que se manifesta a desigualdade estrita  $d_1 < d_2$ .

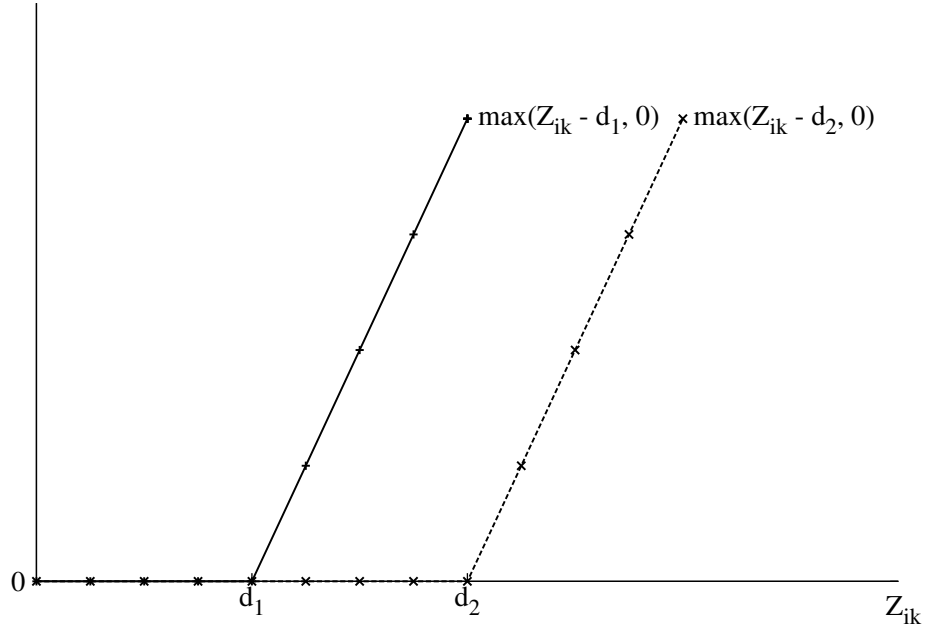


Figure 1: Os valores das parcelas do lado esquerdo da equação (6).

Usando as igualdades (6) no lugar das restrições de desigualdade do problema (5), seria obtido um novo problema onde  $Z_{ik}$  continuaria inferiormente ilimitado. Note que a função objetivo do problema (5) força a redução de cada  $Z_{ik}$ , portanto pode se pensar em limitar estas variáveis inferiormente substituindo o sinal "=" em (6) pelo sinal ">". Obtém-se assim o problema não canônico, i.e., com restrições de desigualdade estrita

$$\begin{aligned} \min \sum_{a \in C_i} Z_{ik} \\ \text{s. a } \varphi(Z_{ik} - \|x - a\|) + \varphi(Z_{ik} - d(X_{ik}, a)) > 0, \quad a \in C_i. \end{aligned} \quad (7)$$

A formulação canônica pode ser recuperada a partir de (7) pela adição de um termo perturbação  $\varepsilon$ . Considera-se assim o problema modificado

$$\begin{aligned} \min \sum_{a \in C_i} Z_{ik} \\ \text{s. a } \varphi(Z_{ik} - \|x - a\|) + \varphi(Z_{ik} - d(X_{ik}, a)) \geq \varepsilon, \quad a \in C_i. \end{aligned} \quad (8)$$

para  $\varepsilon > 0$ .

Uma vez que a região viável do problema (7) é o limite da região viável de (8) quando  $\varepsilon \rightarrow 0_+$ , pode-se então solucionar o problema (7) resolvendo uma sequência de problemas da forma (8) gerada por uma sequência de valores de  $\varepsilon$  que tende a zero.

### 3. SUAVIZAÇÃO DO PROBLEMA

Analisando o problema (8), vê-se que a definição da função  $\varphi$  envolve uma estrutura rígida e não-diferenciável, o que torna a sua solução computacionalmente muito complexa. Em vista disso, o método numérico que será adotado para resolver as restrições do problema (8) faz uso de uma estratégia de suavização [13]. Dessa perspectiva, define-se a função

$$\phi(y, \tau) = (y + \sqrt{y^2 + \tau^2})/2; \quad (9)$$

para  $y \in R$  e  $\tau > 0$ .

A função  $\phi$  tem as seguintes propriedades:

- (a)  $\phi(y, \tau) > \varphi(y), \forall \tau > 0$ ;
- (b)  $\lim_{\tau \rightarrow 0} \phi(y, \tau) = \varphi(y)$ ;
- (c)  $\phi(\cdot, \tau)$  é uma função  $C^\infty$  crescente e convexa.

A função  $\phi$  é uma boa aproximação da função  $\varphi$ . O parâmetro  $\tau$  indica o nível da aproximação, pois à medida que  $\tau$  tende 0, a função suavizadora  $\phi$  se aproxima da função original  $\varphi$ . Adotando-se as mesmas hipóteses da Figura 1, as parcelas das restrições do problema (8) e suas correspondentes aproximações suavizadas, dadas por (9), são exibidas na Figura 2.

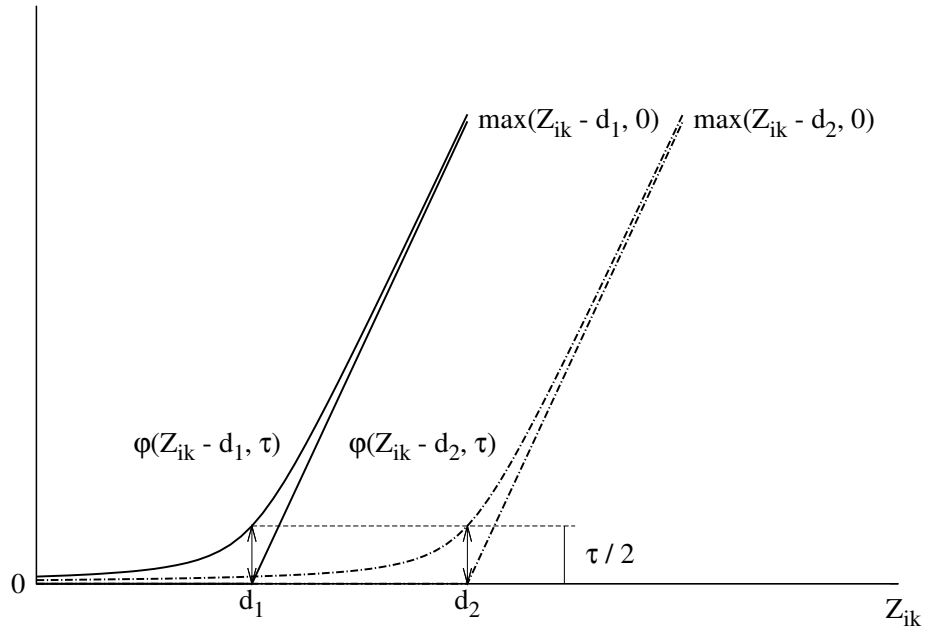


Figure 2: O valor das parcelas das restrições do problema (8) e suas respectivas aproximações suavizadas dadas pela função  $\phi$ .

Substituindo a função  $\varphi$  por  $\phi$  no problema (8) obtém-se

$$\min \sum_{a \in C_i} Z_{ik} \tag{10}$$

$$s. a \quad \phi(Z_{ik} - \|x - a\|, \tau) + \phi(Z_{ik} - d(X_{ik}, a), \tau) \geq \varepsilon, \quad a \in C_i.$$

para  $\varepsilon > 0$ .

Para se obter um problema completamente diferenciável é necessário, ademais, suavizar a função distância euclidiana  $\|x - a\|$ . Com este propósito, define-se a função

$$\theta(x, a, \gamma) = \sqrt{\|x - a\|_2^2 + \gamma^2}, \quad \text{para } \gamma > 0.$$

A função  $\theta$  tem as seguintes propriedades:

- (a)  $\lim_{\gamma \rightarrow 0} \theta(x^j, a^i, \gamma) = \|x^j - a^i\|$ ;
- (b)  $\theta$  é uma função  $C^\infty$ .

A função  $\theta(x, a, \gamma)$  é uma boa aproximação para a distância euclidiana  $\|x - a\|$ . O

parâmetro  $\gamma$  indica a qualidade da aproximação. Na medida em que  $\gamma$  se aproxima de 0, a função  $\theta$  tende para o valor exato da distância, conforme visualizado na figura 3.

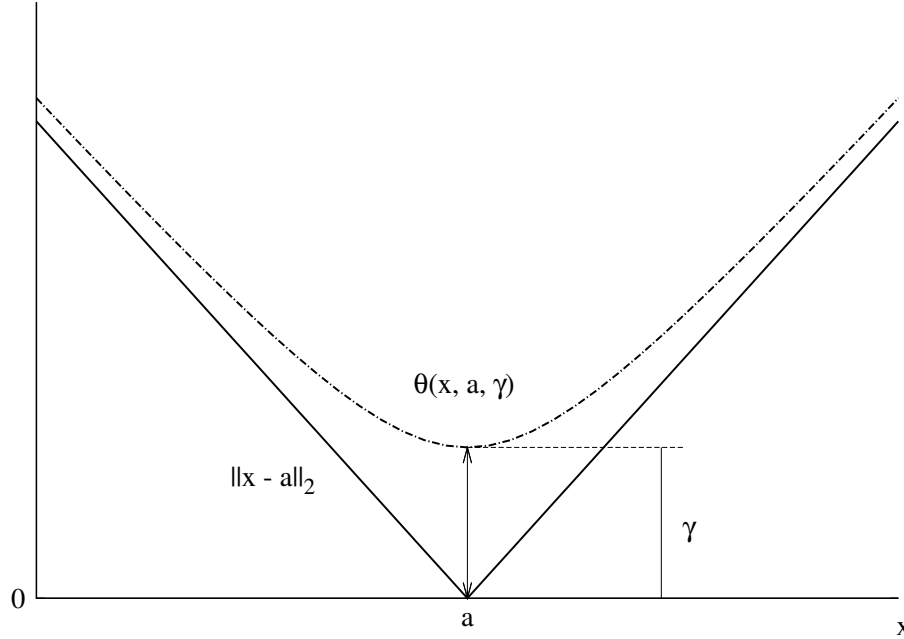


Figure 3: Distância euclidiana entre os pontos  $x$  e  $a$  e sua suavização dada pela função  $\theta(x, a, \gamma)$ .

Substituindo a distância  $\|x - a\|$  pela função  $\theta$  no problema (10), obtém-se o problema diferenciável

$$\begin{aligned} \min \sum_{a \in C_i} Z_{ik} & \quad (11) \\ s. a \quad \phi(Z_{ik} - \theta(x, a, \gamma), \tau) + \phi(Z_{ik} - d(X_{ik}, a), \tau) & \geq \varepsilon, \quad a \in C_i. \end{aligned}$$

As propriedades das funções  $\phi$  e  $\theta$  permitem encontrar a solução do problema (7) resolvendo uma sequência de subproblemas da forma (11) produzida pelo decréscimo dos parâmetros  $\gamma, \tau, \varepsilon$ .

Dado um centróide  $x$ , a propriedade (c) da função hiperbólica  $\phi$  implica que as restrições do problema (11) serão ativas. Portanto, o problema (11) é equivalente a

$$\begin{aligned} \min \sum_{a \in C_i} Z_{ik} & \quad (12) \\ s. a \quad h_{ik}(Z_{ik}, x) = \phi(Z_{ik} - \theta(x, a, \gamma), \tau) + \phi(Z_{ik} - d(X_{ik}, a), \tau) - \varepsilon & = 0, \quad a \in C_i. \end{aligned}$$

A dimensão do espaço das variáveis do problema (12) é  $|C_i| + n$ . Como em geral, o número de observações dado por  $|C_i|$  é grande, o problema (12) tem um grande número de variáveis. Todavia, ele possui uma estrutura separável, pois cada variável  $Z_{ik}$  aparece em uma única restrição de igualdade. Além disso, a derivada parcial de  $h_{ik}(Z_{ik}, x)$  com respeito a  $Z_{ik}$  é estritamente positiva, pois  $h_{ik}(Z_{ik}, x)$  é crescente com respeito a  $Z_{ik}$  (vide Figura 2). Portanto, é possível usar o Teorema da Função Implícita para calcular cada componente  $Z_{ik}$  como função

do centróide  $x$ .

Dessa forma, o problema irrestrito

$$\min f(x) = \sum_{a \in C_i} Z_{ik}(x). \quad (13)$$

é obtido, onde cada  $Z_{ik}(x)$  resulta do cálculo de uma raiz de cada equação

$$h_{ik}(Z_{ik}, x) = \phi(Z_{ik} - \theta(x, a, \gamma), \tau) + \phi(Z_{ik} - d(X_{ik}, a), \tau) - \varepsilon = 0, \quad a \in C_i. \quad (14)$$

Devido a propriedade (a) da função hiperbólica  $\phi$ , as duas primeiras parcelas da equação acima são estritamente crescentes com respeito a variável  $Z_{ik}$  e, portanto, a equação acima possui raiz única.

Novamente, devido ao Teorema da Função Implícita as funções  $Z_{ik}(x)$  têm todas as derivadas com respeito a variável  $x$ . Assim, é possível calcular o gradiente da função objetivo do problema (13) da seguinte forma:

$$\nabla f(x) = \sum_{a \in C_i} \nabla Z_{ik}(x), \quad (15)$$

onde

$$\nabla Z_{ik}(x) = -\nabla h_{ik}(Z_{ik}) / \frac{\partial h_{ik}(Z_{ik}, x)}{\partial Z_{ik}}. \quad (16)$$

A metodologia acima emprega as mesmas idéias apresentadas em [1] utilizadas no desenvolvimento do algoritmo do gradiente reduzido dedicado à resolução do problema geral de programação não-linear sujeito a restrições de igualdade. Dessa forma, é fácil resolver o problema (13) utilizando qualquer método baseado em informações da derivada de primeira ordem. É necessário enfatizar que o problema (12) é definido em um espaço  $n$ -dimensional, ou seja, trata-se de um problema de pequeno porte.

Em suma, pelo uso da suavização hiperbólica, conforme acima apresentado deve ser resolvida uma sequência de problemas da forma abaixo, obtida pela manipulação conveniente dos parâmetros  $\tau, \varepsilon, \gamma$ :

$$\min \sum_{a \in C_i} Z_{ik} \quad (17)$$

$$s. a \quad h_{ik}(Z_{ik}, x) = \phi(Z_{ik} - \theta(x, a, \gamma), \tau) + \phi(Z_{ik} - d(X_{ik}, a), \tau) - \varepsilon = 0, \quad a \in C_i,$$

onde

$$\phi(y, \tau) = (y + \sqrt{y^2 + \tau^2})/2, \quad (18)$$

$$\theta(x, a, \gamma) = \sqrt{\|x - a\|_2^2 + \gamma^2}. \quad (19)$$

Ademais, esse problema pode ser resolvido pela forma irrestrita (13).

Agora, será feita uma digressão sobre os efeitos adicionais engendrados pelo uso da suavização. Para efeito de uma ilustração bem singela, será considerado um problema com unicamente três pontos num espaço de uma única dimensão.

Na Figura 4, são representadas as parcelas  $Z_{i1}, Z_{i2}, Z_{i3}$  da função objetivo do problema original como função de  $x$ . A Figura 5 é a representação suavizada do mesmo caso representado pela Figura 4. As parcelas suavizadas representadas na Figura 5 foram geradas a partir das soluções  $Z_{i1}, Z_{i2}, Z_{i3}$  das restrições de igualdade do problema suavizado (11) em função de  $x$  e um conjunto de três valores pré-fixados dos parâmetros  $\tau, \gamma, \varepsilon$ .

Na Figura 6, foram representadas a soma das três parcelas da função objetivo, bem como

das respectivas funções suavizadas. Observa-se na Figura 6, a redução do número de mínimos locais. Em particular, os mínimos locais existentes em  $a_1$  e  $a_2$  na função original são eliminados na função objetivo suavizada. Isso mostra o poder convexificador da proposta de suavização hiperbólica. Constata-se que a soma das parcelas suavizadas resultou numa função unimodal com mínimo global próximo ao da função original. Embora esse fato seja extremamente promissor, não se pode garantir que, para qualquer caso, o mesmo torne a ocorrer.

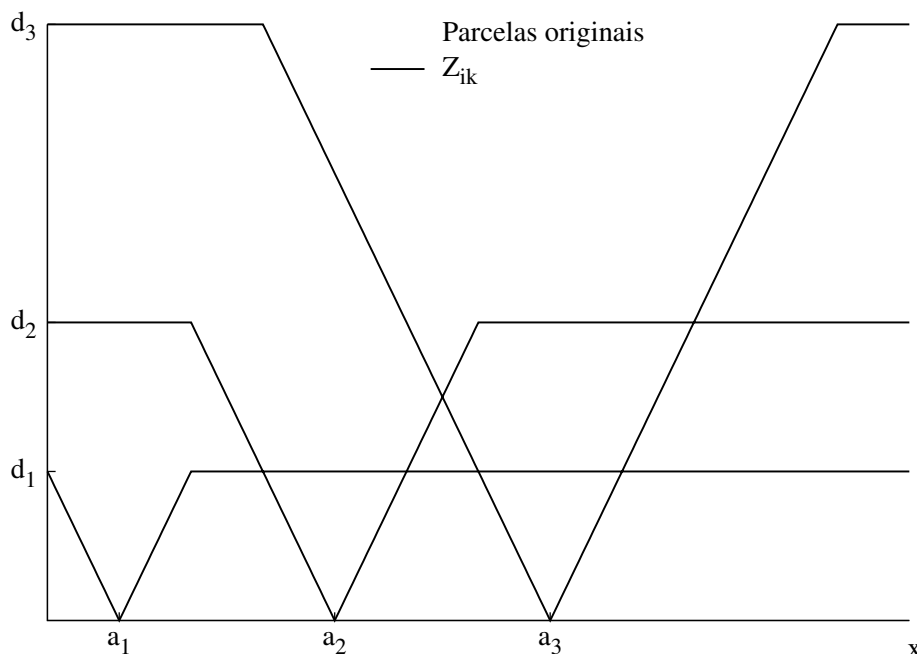


Figure 4: Parcelas originais  $Z_{ik}$  do problema (2).

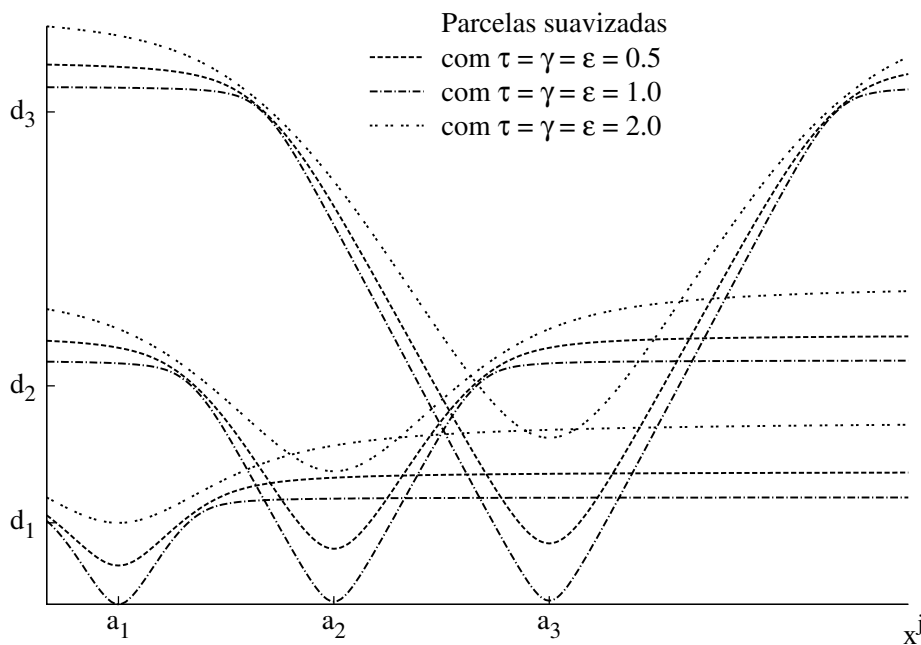


Figure 5: Parcelas do problema suavizado (11) com diferentes valores dos parâmetros  $\tau$ ,  $\gamma$ ,  $\epsilon$ .



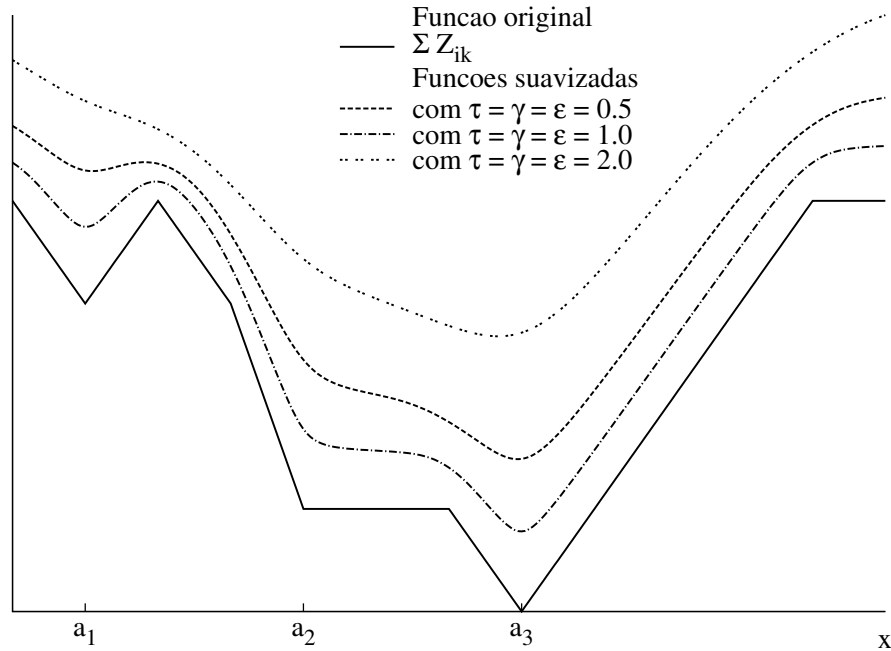


Figure 6: Função objetivo original  $Z_{ik}$  do problema (2) e funções objetivo do problema suavizado (11) com diferentes valores dos parâmetros  $\tau$ ,  $\gamma$ ,  $\varepsilon$ .

A transformação do problema original não-diferenciável e fortemente não-convexo em um problema diferenciável e menos não-convexo, i.e., com um menor número mínimos locais é uma das principais vantagens da suavização hiperbólica. As Figuras 5 e 6 ilustram essa afirmação no caso em que  $x \in R$  e  $i \in \{1,2,3\}$ .

Assim, de acordo com a apresentação acima o processo de solução do problema (17) será implementado através do seguinte algoritmo geral:

---

### Algoritmo SH (suavização hiperbólica)

---

#### Inicialização

Escolha ponto inicial  $x^0$  para parâmetros  $\gamma^1, \tau^1, \varepsilon^1$

Escolha valores de redução  $\rho_j$ ,  $j = 1,2,3$  no intervalo aberto  $(0,1)$ .

Faça  $k = 1$

#### Passo Principal

Enquanto algum critério de parada não for satisfeito, repita:

Resolva para  $j = 1,2$  o problema (17) com

$$\gamma = \gamma^k,$$

$$\tau = \tau^k,$$

$$\varepsilon = \varepsilon^k,$$

tomando como ponto inicial  $x^{k-1}$  e obtendo como  $x^k$  como solução.

Faça

$$\gamma^{k+1} = \rho_1 \gamma^k,$$

$$\tau^{k+1} = \rho_2 \tau^k,$$

$$\varepsilon^{k+1} = \rho_3 \varepsilon^k,$$

$$k = k + 1.$$


---

Como em outros métodos de suavização, a solução para o problema (17) será obtida através da resolução de uma seqüência de subproblemas de minimização irrestritos ( $k = 1, 2, \dots, K, m$ ) conforme indicado no Algoritmo SH.

Note que o Algoritmo SH faz  $\tau$  e  $\gamma$  se aproximarem de zero, logo as funções  $\phi$  e  $\theta$  presentes nas restrições do problema (17) aproximam-se das funções  $\varphi$  e  $\|\cdot\|$  respectivamente. Adicionalmente, faz  $\varepsilon$  se aproximar de zero, portanto, em um movimento simultâneo, o problema resolvido se aproxima gradativamente do problema original. É implicitamente assumido que o algoritmo encontra uma solução de mínimo global,  $x^k$ , do subproblema suavizado ou, pelo menos, um mínimo local profundo, ou seja, com valor muito baixo.

Sob essa hipótese e devido às propriedades de continuidade de todas as funções envolvidas, a seqüência  $x^1, x^2, \dots, x^K$  de valores ótimos deve tender igualmente a um valor ótimo profundo do problema original.

#### 4. EXPERIMENTOS COMPUTACIONAIS

O método proposto é recursivo e em cada iteração cria dois novos grupos, um para cada classe, e determina seus respectivos centros. Em seguida, cada uma das observações da base de dados é associada ao grupo de cujo centro está mais próxima. Após essa etapa, ocorre um refinamento na posição dos últimos centros gerados, levando-se em conta apenas as observações corretamente classificadas. Mesmo após o refinamento dos centros, podem ocorrer erros em que uma observação pertencente a uma dada classe é alocada, de forma equivocada, em um agrupamento pertencente à outra classe. Apesar dos esforços do algoritmo ao longo das iterações para redução desses erros, tendência empiricamente constatada, não há garantias de que eles sejam completamente eliminados até o final do processo.

Observa-se ainda que, devido às características do algoritmo, pode ocorrer em certos casos um ligeiro acréscimo do valor da função objetivo entre duas iterações consecutivas, mas se mantém a tendência de decréscimo do valor da função objetivo ao longo das iterações na quase totalidade das iterações desses experimentos.

Nos experimentos computacionais realizados, utilizou-se como critério de parada uma tolerância na variação da função objetivo entre duas iterações consecutivas. A parada do algoritmo ocorre quando o módulo da diferença de valores da função objetivo, em duas iterações sucessivas, dividido pelo valor inicial da função objetivo for inferior a uma dada tolerância, conforme definido no passo 4 do algoritmo proposto na seção 2. Em todos os experimentos foi adotada a tolerância  $\varepsilon = 10^{-2}$ .

Para avaliação de resultados, aplicamos a metodologia MCSH às bases de dados Australian Credit Approval (ACA) e Breast Cancer Wisconsin - Diagnostic (BCW) disponibilizadas pela UCI-Machine Learning Repository [2].

Na tabela 1, para comparação dos resultados de nossos experimentos numéricos com a base (ACA), escolhemos o melhor resultado obtido pelos algoritmos de classificação com otimização local apresentados na seção 3.8 de [4] e o melhor resultado dos 23 algoritmos de classificação encontrados no capítulo 9 de [10]. Para a comparação dos resultados obtidos com a base (BCW), escolhemos os resultados obtidos por Hybrid misclassification minimization (HMM) [6] e Parametric misclassification minimization (PMM) [4]. A coluna *error* indica o percentual médio de elementos equivocadamente classificados no processo de validação ten-fold.

Base de dados (ACA)		Base de dados (BCW)	
Algoritmo	error(%)	Algoritmo	error(%)

Bagirov	14.4	HMM	2.4
Michie	13.1	PMM	2.2
MCSH	13.3	MCSH	3.7

Table 1: Resultados dos experimentos computacionais utilizando as bases de dados Australian Credit Approval (ACA) e Breast Cancer Wisconsin - Diagnostic (BCW).

## 5. CONCLUSÃO

No que tange a acurácia, o método proposto obteve desempenho comparável aos dos métodos apresentados. Com relação ao tempo de processamento, o método aqui proposto mostrou-se muito veloz, convergindo para soluções comparáveis a dos demais algoritmos considerados em tempo expressivamente baixo, 1.47 segundos na base (ACA) e 7.47 segundos na base (BCW).

Propõe-se em nosso trabalho um método alternativo para resolução do problema de classificação que se baseia no método proposto por Bagirov. Por esse método alternativo, os problemas não-diferenciáveis formulados por Bagirov são substituídos por problemas equivalentes suavizados, nos quais as restrições dos problemas originais são representadas de forma aproximada por funções hiperbólicas.

O algoritmo proposto foi implementado em linguagem FORTRAN 77. Os problemas de otimização considerados foram resolvidos utilizando o método Quase-Newton com atualização da aproximação da matriz hessiana dada pela fórmula Broyden-Fletcher-Goldfarb-Shanno (BFGS).

## REFERÊNCIAS

- [1] Abadie, J. e Carpenter, J. (1969). *Generalization of the wolfe reduced gradient method to the case of nonlinear constraints*. Academic Press.
- [2] Asuncion, A. e Newman, D.J. (2007). UCI Machine Learning Repository.
- [3] Bagirov, A. M. e Rubinov, A. M. e Yearwood, J. (2002). A global optimization approach to classification. *Optimization and Engineering*, 22:65-74.
- [4] Bagirov, A. M. e Yearwood, J. (2003) A new nonsmooth optimization algorithm for clustering problems. Technical report, University of Ballarat.
- [5] Brown, M. G. e Lin, D. C. W. e Sugest, C. M. e Furey, T. e Ares, M. e Haussler, D. (2000) Knowledge based analysis of microarray gene expression data mining using support vector machines. *Proceedings of National Academy of Sciences*, pages 262-267.
- [6] Chen , Chunhui e Mangasarian, O. L. (1995). Hybrid Misclassification Minimization. Technical report, MP-TR-1995-05, University of Wisconsin.
- [7] DeCoste, D. and Scholkopf, S. (2002) Training invariant support vector machines. *Machine Learning*, 46:161-190.
- [8] Joachins, T. (1969) A dynamic programming algorithm for cluster analysis. *Operations Research*, 17:1034-1057.
- [9] Mangasarian, O. L. (1994) Misclassification minimization. *Journal of Global*

*Optimization*, 5:309-323.

[10] Michie, D. e Spiegelhalter, D. J. e Taylor, C. C. (1994) *Machine Learning and Statistical Classification* of Artificial intelligence. Ellis Horwood.

[11] Moreno, A. O. (2008) *Um novo algoritmo para resolução de problemas de classificação*. PhD thesis, Universidade Federal do Rio de Janeiro.

[12] Vapnik, V. N. (1995) *The nature of statistical learning theory*. Springer Verlag, 1995.

[13] Xavier, A. E. (1993) Solução de problemas de programação não-diferenciáveis via suavização. Technical report, Universidade Federal do Rio de Janeiro.