

UMA ABORDAGEM USANDO REDES NEURAIIS ARTIFICIAIS PARA RESOLUÇÃO DE PROBLEMAS DE OTIMIZAÇÃO RESTRITA

Ivan Nunes da Silva *

DEE / Faculdade de Engenharia (FE)
Universidade Estadual Paulista (UNESP)
Bauru – SP
ivan@feb.unesp.br

Wagner Caradori do Amaral

DCA / Faculdade de Eng. Elétrica e Computação (FEEC)
Universidade Estadual de Campinas (UNICAMP)
Campinas – SP

Lúcia V. R. de Arruda

Progr. Pós-Grad. Eng. Elétrica e Inform. Industrial (CPGEI)
Centro Federal de Educ. Tecnol. do Paraná (CEFET-PR)
Curitiba – PR

* *Corresponding author*/autor para quem as correspondências devem ser encaminhadas

Recebido em 08/2003; aceito em 06/2004 após 1 revisão
Received August 2003; accepted June 2004 after one revision

Resumo

Sistemas baseados em redes neurais artificiais fornecem altas taxas de computação devido ao uso de um número massivo de elementos processadores simples. Redes neurais com conexões realimentadas fornecem um modelo computacional capaz de resolver uma rica classe de problemas de otimização. Este artigo apresenta uma nova abordagem para resolver problemas de otimização restrita utilizando redes neurais artificiais. Mais especificamente, uma rede de Hopfield modificada é desenvolvida cujos parâmetros internos são calculados usando a técnica de subespaço válido de soluções. A partir da obtenção destes parâmetros a rede tende a convergir aos pontos de equilíbrio que representam as possíveis soluções para o problema. Exemplos de simulação são apresentados para justificar a validade da abordagem proposta.

Palavras-chave: otimização restrita; redes neurais artificiais; redes de Hopfield.

Abstract

Systems based on artificial neural networks have high computational rates due to the use of a massive number of simple processing elements and the high degree of connectivity between these elements. Neural networks with feedback connections provide a computing model capable of solving a large class of optimization problems. This paper presents a novel approach for solving constrained optimization problems using artificial neural networks. More specifically, a modified Hopfield network is developed and its internal parameters are computed using the valid-subspace technique. These parameters guarantee the convergence of the network to the equilibrium points, which represent the feasible solutions to problem. Simulated examples are presented to demonstrate the validity of the proposed method.

Keywords: constrained optimization; artificial neural networks; Hopfield networks.

1. Introdução

Problemas de otimização não-linear restrita referem-se, geralmente, ao problema de minimizar ou maximizar uma função objetivo não-linear sujeita a um conjunto de restrições (lineares e/ou não-lineares) de igualdade e/ou desigualdade (Bazaraa & Shetty, 1979).

Embora existam na literatura diversos métodos de programação matemática que possam ser aplicados na resolução de problemas de otimização não-linear restrita, há a necessidade crescente de investigar métodos alternativos que explorem arquiteturas de processamento inerentemente paralelas e adaptativas. Assim sendo, as redes neurais artificiais torna-se uma abordagem promissora que pode ser aplicada eficientemente nestes tipos de problemas. Entre as principais vantagens em se utilizar a abordagem de redes neurais artificiais em otimização não-linear restrita destacam-se as seguintes: i) a capacidade intrínseca de operação em paralelo; ii) a simplicidade de implementação em hardware; e iii) o alcance de altas taxas de computação por intermédio de elementos simples de processamento.

Basicamente, as redes neurais utilizadas para resolver problemas de otimização não-linear restrita são desenvolvidas utilizando parâmetros de penalidade (Kennedy & Chua, 1988; Vazquez *et al.*, 1990; Zhang *et al.*, 1992). Os pontos de equilíbrio destas redes, correspondentes às soluções do problema, são obtidos através de uma escolha adequada dos parâmetros de penalidade que devem ser suficientemente grandes para garantir a convergência da rede (Kennedy & Chua, 1988). Entretanto, a escolha destes parâmetros é uma tarefa árdua, sendo normalmente realizada através de técnicas de tentativas e erros, as quais podem demandar um esforço computacional bastante excessivo. Além disso, a qualidade da solução final também depende do ajuste destes parâmetros. Algumas outras dificuldades relacionadas ao processo de convergência para os pontos de equilíbrio da rede, que representam as soluções do problema de otimização, são também evidenciadas. Uma análise minuciosa dos resultados apresentados nestes artigos demonstra que muitas vezes resultados infactíveis são apontados como soluções do problema.

Com a finalidade de contornar estes problemas de convergência, de aperfeiçoar a eficiência de simulações por computador, e de fornecer uma nova metodologia para mapeamento de problemas de otimização não-linear restrita, uma rede de Hopfield modificada foi desenvolvida. A referida rede foi implementada com o propósito de que seus pontos de equilíbrio correspondam à solução do problema de otimização não-linear restrita. As principais características desta rede são: i) inexistência de constantes de ponderação; ii) todas as restrições estruturais envolvidas ao problema de otimização não-linear restrita são agrupadas em um único termo de restrição; iii) não há interferência entre termo de otimização e termo de restrição; e iv) nenhum parâmetro de iniciação é requerido para a execução da simulação. Uma análise do comportamento dinâmico da rede será também realizada para ilustrar o processo de convergência da rede em direção às soluções ótimas (pontos de equilíbrio).

Portanto, o problema a ser estudado neste artigo consiste de aplicar uma rede de Hopfield modificada a fim de minimizar uma função de energia E^{ot} , a qual representa a função objetivo do problema, estando sujeita a diversas restrições de igualdade e desigualdade. Essas restrições são todas agrupadas em um único termo de energia denominado E^{conf} , o qual possui o objetivo de confinar todas as restrições envolvidas com o problema de otimização.

Para este propósito a organização deste artigo está como segue. Na Seção 2, a rede de Hopfield modificada é apresentada, e a técnica do subespaço-válido de soluções utilizada para projetar os parâmetros internos da rede é descrita. A Seção 3 contém uma análise do comportamento

dinâmico da rede no contexto do subespaço-válido. Na Seção 4, o mapeamento de problemas de otimização não-linear restrita é formulado utilizando a rede de Hopfield modificada. Resultados de simulação são apresentados na Seção 5 para validar a abordagem proposta. Na Seção 6, as conclusões e os pontos chave deste artigo são apresentados.

2. A Rede de Hopfield Modificada

Além de fornecer um novo método para resolver problemas de otimização não-linear restrita, redes neurais artificiais oferecem um método que explora arquiteturas de processamento intrinsecamente paralelas e adaptativas. Neste artigo, uma rede de Hopfield modificada, cujos pontos de equilíbrio representam uma solução do problema de otimização não-linear restrita, é desenvolvida.

A rede de Hopfield é provavelmente o melhor exemplo conhecido de uma rede recorrente. Como definida em Hopfield (1984), esta rede apresenta geralmente uma única camada com conexões realimentadas entre os nós. Na maioria dos casos, os nós (neurônios) são completamente interconectados, ou seja, todos os neurônios da rede estão conectados aos outros e a si próprio. A equação nodal para a rede de Hopfield contínua no tempo é dada por:

$$\dot{u}_i(t) = -\eta u_i(t) + \sum_{j=1}^N T_{ij}(t) v_j(t) + i_i^b \quad (1)$$

$$v_i(t) = g(u_i(t)) \quad (2)$$

onde:

$u_i(t)$ é o estado corrente do i -ésimo neurônio.

$v_i(t)$ é a saída do i -ésimo neurônio.

T_{ij} é o peso conectando o i -ésimo neurônio ao j -ésimo neurônio.

i_i^b é a entrada do i -ésimo neurônio (input bias).

$\eta u_i(t)$ é um termo de decaimento passivo.

Na equação (2), $g(u_i(t))$ é uma função de ativação, monótona crescente, que limita a saída de cada neurônio para um intervalo pré-definido. Em Hopfield (1984) mostra-se que os pontos de equilíbrio da rede correspondem aos valores de $v(t)$ que minimizam a função de energia da rede. A função de energia associada à rede de Hopfield é definida por:

$$E(t) = -\frac{1}{2} v(t)^T \cdot T \cdot v(t) - v(t)^T \cdot i^b \quad (3)$$

De fato, a função de energia dada em (3) é uma função de Lyapunov (Vidyasagar, 1992) e, neste caso, a estabilidade assintótica da rede pode ser demonstrada utilizando o segundo método de Lyapunov, isto é, mostrando que as derivadas parciais de (3) são não crescentes. Esta condição é alcançada quando a matriz T é simétrica. Uma demonstração envolvendo o processo de convergência da rede de Hopfield é apresentada no Apêndice.

Assim, o mapeamento de problemas de otimização utilizando uma rede de Hopfield consiste em determinar, em cada tipo de problema, a matriz de pesos T e o vetor de entradas i^b associados à função de energia da rede (3).

A dificuldade em mapear problemas de otimização distintos através de uma rede de Hopfield convencional está em satisfazer as várias restrições que são impostas por cada tipo de problema. Uma técnica simples de mapeamento codifica as restrições como termos na função de energia que são minimizados quando as restrições são satisfeitas, ou seja:

$$E(t) = E^{ot}(t) + c_1.E_1^{rest}(t) + c_2.E_2^{rest} + \dots + c_k.E_k^{rest} \tag{4}$$

onde c_i são constantes positivas que ponderam cada uma das restrições. Assim, a rede atua com o propósito de minimizar simultaneamente uma função de energia (E^{ot}) correspondente à função objetivo associada ao problema, e as funções (E_k^{rest}) que representam as k -ésimas restrições envolvidas no problema. Portanto, se qualquer uma destas restrições for violada, a solução é considerada ‘inválida’ (Silva, 1997).

No processo de otimização com uma rede de Hopfield convencional, a multiplicidade de termos de restrições em (4) pode cancelar uma ou várias restrições simultaneamente. Como resultado, as soluções obtidas no final do processo de otimização podem ser inactíveis; e o desempenho da rede é sensível aos valores dos parâmetros c_i , que são geralmente especificados através de procedimentos não formais (Silva *et al.*, 1997).

Para contornar estes problemas, realizou-se em Aiyer *et al.* (1990) uma análise da convergência da rede que permite verificar como E^{ot} e os termos de restrições E_k^{rest} em (4) podem ser separados dentro de subespaços diferentes para evitar a inactibilidade do problema. Este subespaço que agrupa todas as restrições impostas pelo problema é denominado subespaço-válido de soluções, com equação definida por:

$$\mathbf{v}(t+\Delta t) = \mathbf{T}^{conf}.\mathbf{v}(t) + \mathbf{i}^{conf} \tag{5}$$

onde \mathbf{T}^{conf} é uma matriz projeção ($\mathbf{T}^{conf}.\mathbf{T}^{conf} = \mathbf{T}^{conf}$) que projeta o vetor \mathbf{v} para o subespaço-válido, e $\mathbf{T}^{conf}.\mathbf{i}^{conf} = \mathbf{0}$ (\mathbf{i}^{conf} é um vetor ortogonal ao subespaço-válido). Um estudo detalhado dos aspectos teóricos relacionados ao subespaço-válido é apresentado em Silva *et al.* (1997).

Assim, utilizando o subespaço-válido com o objetivo de agrupar todas as restrições envolvidas a um determinado problema, a função de energia dada em (4) passa a ser definida por:

$$E(t) = E^{ot}(t) + c_0.E^{conf}(t) \tag{6}$$

onde E^{conf} confina todas as restrições E_k^{rest} , pertencentes a equação (4), para o subespaço-válido. A função de energia E^{ot} e E^{conf} dadas na equação (6) são definidas por:

$$E^{ot}(t) = -\frac{1}{2}\mathbf{v}(t)^T.\mathbf{T}^{ot}.\mathbf{v}(t) - \mathbf{v}(t)^T.\mathbf{i}^{ot} \tag{7}$$

$$E^{conf}(t) = -\frac{1}{2}\mathbf{v}(t)^T.\mathbf{T}^{conf}.\mathbf{v}(t) - \mathbf{v}(t)^T.\mathbf{i}^{conf} \tag{8}$$

A necessidade de assegurar que E^{ot} seja otimizado quando todas as restrições contidas em E^{conf} estejam são satisfeitas, implica em atribuir um valor elevado para a constante c_0 em (6). Esta condição torna a simulação da rede ineficiente, desde que a maioria do esforço computacional é para forçar o confinamento das restrições.

Portanto, com o objetivo de evitar esses problemas de convergência, uma rede de Hopfield modificada que utiliza a técnica de subespaço-válido de soluções é então proposta. A arquitetura desenvolvida (Figura 1) diretamente garante a validade das restrições agrupadas

por E^{conf} , dispensando assim a utilização da constante de ponderação c_0 . Assim, a principal contribuição da arquitetura proposta, em constaste com a maioria das abordagens neurais propostas na literatura, está na inexistência de parâmetros de ponderação e/ou penalidade, implicando então na não necessidade de especificação dos mesmos. Uma análise da convergência da rede de Hopfield modificada proposta neste trabalho é realizada da Seção 3.

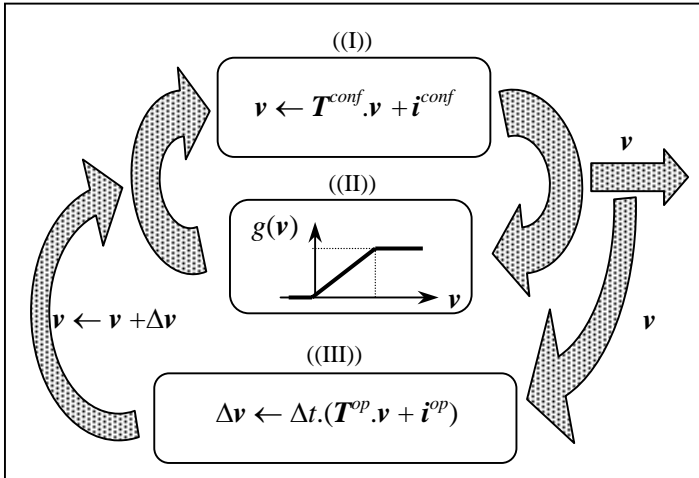


Figura 1 – A Rede de Hopfield Modificada.

A dinâmica da rede de Hopfield modificada, ilustrada na Figura 1, é explicitada através de três passos principais:

(I) Minimização de E^{conf} : corresponde à projeção de \mathbf{v} sobre o subespaço-válido que confina todas as restrições impostas pelo problema:

$$\mathbf{v}(t+\Delta t) = \mathbf{T}^{conf} \cdot \mathbf{v}(t) + \mathbf{i}^{conf}$$

Esta operação realiza uma minimização indireta de E^{conf} .

(II) Aplicação de uma função de ativação do tipo ‘rampa simétrica’, restringindo \mathbf{v} dentro de um hiper cubo pré-definido:

$$g_i(v_i) = \begin{cases} \lim_i^{inf} & , \text{ se } \lim_i^{inf} > v_i \\ v_i & , \text{ se } \lim_i^{inf} \leq v_i \leq \lim_i^{sup} \\ \lim_i^{sup} & , \text{ se } v_i > \lim_i^{sup} \end{cases} \quad (9)$$

onde $v_i(t) \in [\lim_i^{inf} , \lim_i^{sup}]$. Esta função de ativação é similar àquela utilizada em (1).

(III) Minimização de E^{ot} : alteração de \mathbf{v} em direção a uma solução de custo ótimo (definida por \mathbf{T}^{ot} e \mathbf{i}^{ot}), correspondente aos pontos de equilíbrio da rede, que são as soluções para o problema de otimização considerado. Utilizando a função de ativação ‘rampa-simétrica’ e dado que $\eta=0$, pois as alterações nos neurônios são realizadas de forma síncrona, ou seja, todos os neurônios serão alterados simultaneamente, então a equação (1) torna-se:

$$\frac{d\mathbf{v}(t)}{dt} = \dot{\mathbf{v}} = -\frac{\partial E^{ot}(t)}{\partial \mathbf{v}}$$

$$\Delta \mathbf{v} = -\Delta t \cdot \nabla E^{ot}(\mathbf{v}) = \Delta t \cdot (\mathbf{T}^{ot} \cdot \mathbf{v} + \mathbf{i}^{ot}) \tag{10}$$

Assim, a minimização de E^{ot} consiste na alteração de \mathbf{v} na direção oposta ao gradiente da função E^{ot} em relação a \mathbf{v} . Estes resultados são também válidos quando uma função de ativação do tipo tangente hiperbólica é utilizada.

Deve-se ressaltar que a aplicação dos passos (I) a (III) são realizados de forma discreta. Neste caso, após as considerações realizadas no passo (III), a resolução da equação diferencial representada em (10) será similar ao método de Euler.

3. Análise da Dinâmica da Rede

Nesta seção, realiza-se uma análise linearizada da convergência da rede de Hopfield modificada cuja dinâmica de operação é implementada através dos passos (I)-(III) apresentados anteriormente. Em particular, considera-se que a região de operação na qual o vetor \mathbf{v} está contido é limitada pelo hiper-cubo definido pela função de ativação ‘rampa-simétrica’ (9). A equação nodal descrevendo o comportamento dinâmico desta rede é obtida a partir de (1) para $\eta = 0$ e $\mathbf{v}(t) = \mathbf{u}(t)$, ou seja:

$$\dot{v}_i(t) = \sum_{j=1}^N T_{ij} \cdot v_j(t) + i_i^b$$

A análise dinâmica da rede é portanto inicializada com a obtenção de uma equação para $\dot{\mathbf{v}}^{conf}$, componente de $\dot{\mathbf{v}}$ que pertence ao subespaço-válido.

Como na rede de Hopfield modificada a saída \mathbf{v} é constantemente confinada ao subespaço-válido pela aplicação do passo (I) descrito na Seção 2, isto é $\{\mathbf{v} \leftarrow \mathbf{T}^{conf} \cdot \mathbf{v} + \mathbf{i}^{conf}\}$, então a aplicação deste processo iterativo implica que qualquer componente de $\dot{\mathbf{v}}$ ortogonal a $\dot{\mathbf{v}}^{conf}$ é continuamente suprimida. Logo, a componente $\dot{\mathbf{v}}^{conf}$ (não $\dot{\mathbf{v}}$) caracteriza melhor a dinâmica global da rede. Assim, tem-se:

$$\begin{aligned} \dot{\mathbf{v}}^{conf} &= \mathbf{T}^{conf} \cdot \dot{\mathbf{v}} = \mathbf{T}^{conf}(\mathbf{T}^{ot} \cdot \mathbf{v} + \mathbf{i}^{ot}) \\ &= \mathbf{T}^{conf}(\mathbf{T}^{ot}(\mathbf{T}^{conf} \cdot \mathbf{v} + \mathbf{i}^{conf}) + \mathbf{i}^{ot}) \\ &= \mathbf{T}^{conf} \mathbf{T}^{ot} \mathbf{T}^{conf} \cdot \mathbf{v} + \mathbf{T}^{conf}(\mathbf{T}^{ot} \cdot \mathbf{i}^{conf} + \mathbf{i}^{ot}) \end{aligned} \tag{11}$$

A partir da equação (11), verifica-se que a componente $\dot{\mathbf{v}}^{conf}$ é constituída de duas partes, um termo constante $\{\mathbf{T}^{conf}(\mathbf{T}^{ot} \cdot \mathbf{i}^{conf} + \mathbf{i}^{ot})\}$; e um termo dependente de \mathbf{v} , $\{\mathbf{T}^{conf} \mathbf{T}^{ot} \mathbf{T}^{conf} \cdot \mathbf{v}\}$. Estas expressões podem ser simplificadas por:

$$\mathbf{T}^{conf} \mathbf{T}^{ot} \mathbf{T}^{conf} = \mathbf{A} \tag{12}$$

$$\mathbf{T}^{conf}(\mathbf{T}^{ot} \cdot \mathbf{i}^{conf} + \mathbf{i}^{ot}) = \mathbf{b} \tag{13}$$

Substituindo (12) e (13) em (11), obtém-se:

$$\dot{\mathbf{v}}^{conf} = \mathbf{A} \cdot \mathbf{v} + \mathbf{b} = \mathbf{A} \cdot \mathbf{v}^{conf} + \mathbf{b} \tag{14}$$

onde: $\mathbf{v}^{conf} = \mathbf{T}^{conf} \cdot \mathbf{v}$ e $\mathbf{T}^{conf} \cdot \mathbf{T}^{conf} = \mathbf{T}^{conf}$.

Com \mathbf{v} confinado ao subespaço-válido (isto é: $\mathbf{v} = \mathbf{T}^{conf} \cdot \mathbf{v} + \mathbf{i}^{conf}$ e $\mathbf{T}^{conf} \cdot \mathbf{i}^{conf} = \mathbf{0}$), E^{ot} pode ser expressado como:

$$E^{ot} = -\frac{1}{2} \mathbf{v}^T \mathbf{A} \mathbf{v} - \mathbf{b}^T \mathbf{v} \tag{15}$$

Na equação (15) observa-se que a dinâmica de $\dot{\mathbf{v}} = \dot{\mathbf{v}}^{conf} = \mathbf{A} \cdot \mathbf{v} + \mathbf{b}$ simplesmente resulta num processo de otimização de E^{ot} em relação ao subespaço-válido. Logo, o objetivo consiste em obter uma solução válida que minimiza E^{ot} .

Corolário: A solução geral de um sistema linear invariante no tempo (autônomo) do tipo descrito por $\dot{\mathbf{v}}^{conf} = \mathbf{A} \cdot \mathbf{v}^{conf} + \mathbf{b}$, com $\mathbf{v}^{conf}(\mathbf{0}) = \mathbf{v}_0^{conf}$ é dada por:

$$\mathbf{v}^{conf}(t) = e^{\mathbf{A}t} \mathbf{v}_0^{conf} + \int_0^t e^{\mathbf{A}(t-\tau)} \mathbf{b} \cdot d\tau \tag{16}$$

onde \mathbf{v}_0^{conf} é um vetor de elementos aleatórios correspondendo ao valor de \mathbf{v}^{conf} no instante de tempo $t = 0$. Deve-se observar que quando estamos começando a execução da rede, o Passo (I) é realizado em primeira instância seguido do passo (II). Neste caso, deve-se atribuir valores aleatórios para o vetor \mathbf{v} , representado aqui inicialmente por \mathbf{v}_0^{conf} , a fim de iniciar o processo de confinamento do mesmo para o subespaço-válido. Somente após a convergência ao subespaço-válido é que o passo (III) será aplicado pela primeira vez. Assim, daqui por diante, o valor atribuído \mathbf{v}^{conf} será igual ao valor de \mathbf{v} obtido no passo (III).

Definição: A exponencial de uma matriz quadrada \mathbf{A} é definida pela série de potências dada por:

$$e^{\mathbf{A}\zeta} = \sum_{k=0}^{\infty} \frac{(\mathbf{A}\zeta)^k}{k!}$$

Então, rescrevendo a equação (16) como séries de potências, tem-se:

$$\begin{aligned} \mathbf{v}^{conf}(t) &= \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathbf{A}^k \mathbf{v}_0^{conf} + \int_0^t \sum_{k=0}^{\infty} \frac{(t-\tau)^k}{k!} \mathbf{A}^k \mathbf{b} \cdot d\tau \\ &= \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathbf{A}^k \mathbf{v}_0^{conf} + \sum_{k=0}^{\infty} \frac{\mathbf{A}^k \mathbf{b}}{k!} \int_0^t (t-\tau)^k \cdot d\tau \\ &= \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathbf{A}^k \mathbf{v}_0^{conf} + \sum_{k=0}^{\infty} \frac{\mathbf{A}^k \mathbf{b}}{k!} \left(\frac{t^{k+1}}{k+1} \right) \\ &= \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathbf{A}^k \mathbf{v}_0^{conf} + \sum_{k=0}^{\infty} \frac{t^{k+1}}{(k+1)!} \mathbf{A}^k \mathbf{b} \end{aligned} \tag{17}$$

Para analisar o comportamento de \mathbf{v}^{conf} durante o processo de convergência da rede, considera-se os vetores \mathbf{v}^{conf} , \mathbf{v}_0^{conf} e \mathbf{b} escritos em termos de suas componentes que serão expressas no espaço coordenado gerado a partir dos autovetores normalizados de \mathbf{A} . Assim, considera-se que \mathbf{A} tem autovalores $\lambda_1, \lambda_2, \dots, \lambda_n$, cujos autovetores normalizados são $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$. Para distinguir os autovalores nulos e não-nulos de \mathbf{A} , define-se o conjunto \mathbf{Z} tal que $\lambda_i = 0$ para $i \in \mathbf{Z}$ e $\lambda_i \neq 0$ para $i \notin \mathbf{Z}$. Decompondo-se \mathbf{v}^{conf} , \mathbf{v}_0^{conf} e \mathbf{b} ao longo dos autovetores de \mathbf{A} , tem-se:

$$\mathbf{v}^{conf} = \sum_{i=1}^n v_i \mathbf{u}_i, \quad \mathbf{v}_0^{conf} = \sum_{i=1}^n o_i \mathbf{u}_i, \quad \mathbf{b} = \sum_{i=1}^n b_i \mathbf{u}_i \tag{18}$$

onde v_i, o_i e b_i referem-se respectivamente aos valores da i -ésima componente de \mathbf{v}^{conf} , \mathbf{v}_0^{conf} e \mathbf{b} , representado no espaço coordenado gerado a partir dos autovetores de \mathbf{A} . Portanto, a partir de (18), obtém-se:

$$\mathbf{A}^k \mathbf{v}_0^{conf} = \sum_{i=1}^n o_i \lambda_i^k \mathbf{u}_i, \quad \mathbf{A}^k \mathbf{b} = \sum_{i=1}^n b_i \lambda_i^k \mathbf{u}_i \tag{19}$$

Substituindo a equação (19) em (17) resulta-se em:

$$\begin{aligned} \mathbf{v}^{conf}(t) &= \sum_{k=0}^{\infty} \frac{t^k}{k!} \sum_{i=1}^n o_i \lambda_i^k \mathbf{u}_i + \sum_{k=0}^{\infty} \frac{t^{k+1}}{(k+1)!} \sum_{i=1}^n b_i \lambda_i^k \mathbf{u}_i \\ &= \sum_{i=1}^n o_i \mathbf{u}_i \sum_{k=0}^{\infty} \frac{t^k \lambda_i^k}{k!} + \sum_{i \notin \mathbf{Z}} \frac{b_i \mathbf{u}_i}{\lambda_i} \sum_{k=0}^{\infty} \frac{t^{k+1} \lambda_i^{k+1}}{(k+1)!} + \sum_{i \in \mathbf{Z}} b_i \mathbf{u}_i \sum_{k=0}^{\infty} \frac{t^{k+1} 0^k}{(k+1)!} \\ &= \sum_{i=1}^n e^{\lambda_i t} o_i \mathbf{u}_i + \sum_{i \notin \mathbf{Z}} \frac{b_i \mathbf{u}_i}{\lambda_i} \left(\sum_{k=0}^{\infty} \frac{t^k \lambda_i^k}{k!} - 1 \right) + \sum_{i \in \mathbf{Z}} b_i \mathbf{u}_i t \\ &= \sum_{i=1}^n e^{\lambda_i t} o_i \mathbf{u}_i + \sum_{i \notin \mathbf{Z}} \frac{b_i \mathbf{u}_i}{\lambda_i} (e^{\lambda_i t} - 1) + \sum_{i \in \mathbf{Z}} b_i \mathbf{u}_i t \end{aligned} \tag{20}$$

A equação (20) é completamente geral para quaisquer valores arbitrários de \mathbf{A} , \mathbf{b} e \mathbf{v}_0^{conf} . Entretanto, esta equação pode ser simplificada se \mathbf{A} e \mathbf{b} forem definidas como nas equações (12) e (13). Neste caso, os autovetores de \mathbf{A} com autovalores zero serão confinados ao transpor o subespaço ortogonal ao subespaço-válido, enquanto \mathbf{b} sempre permanece no subespaço-válido $\{b_i = 0 \text{ para } i \in \mathbf{Z}\}$. Assim, a equação (20) torna-se:

$$\mathbf{v}^{conf}(t) = \sum_{i=1}^n e^{\lambda_i t} o_i \mathbf{u}_i + \sum_{i \notin \mathbf{Z}} \frac{b_i \mathbf{u}_i}{\lambda_i} (e^{\lambda_i t} - 1) \tag{21}$$

Agora torna-se também importante examinar a equação (21) para valores pequeno e grande de t . Para t suficientemente pequeno, tem-se a seguinte aproximação:

$$e^{\lambda_i t} \approx 1 + \lambda_i t \tag{22}$$

Substituindo esta aproximação em (21), tem-se:

$$\mathbf{v}^{conf}(t) \approx \sum_{i=1}^n [o_i(1 + \lambda_i t) + b_i t] \cdot \mathbf{u}_i \tag{23}$$

Observa-se que para um \mathbf{v}_0^{conf} aleatoriamente pequeno, os termos o_i são frequentemente pequenos em comparação com os b_i . Logo, a equação (23) transforma-se em:

$$\mathbf{v}^{conf}(t) \approx t \sum_{i=1}^n b_i \mathbf{u}_i = t \cdot \mathbf{b} \tag{24}$$

Analisando a equação (24), nota-se que \mathbf{v}^{conf} inicialmente parte na direção do vetor \mathbf{b} . No limite, quando t é muito grande, a equação (21) indica que \mathbf{v}^{conf} tenderá em direção aos autovetores de \mathbf{A} correspondente ao maior autovalor positivo. Neste caso, a partir dos resultados advindos de Aiyer *et al.* (1990) e Vidyasagar (1992), os pontos de equilíbrio da rede podem ser iterativamente computados visto que o estado da rede partindo de uma posição inicial arbitrária sempre convergirá para um ponto de equilíbrio estável. Este ponto de equilíbrio é sempre limitado pela função de ativação ‘rampa-simétrica’ definida em (9).

Embora o processo de convergência em direção aos pontos de equilíbrio seja assegurado, deve-se observar que isto não implica na obtenção de um ponto de mínimo global em relação à função de energia $E^m(t)$ associada à rede de Hopfield modificada.

4. Formulação do Problema de Otimização Não-Linear Restrita

Considera-se o seguinte problema de otimização não-linear restrita contendo p -restrições de desigualdade, com variáveis limitadas, ou seja:

$$\text{Minimizar } E^{ot}(\mathbf{v}) = f(\mathbf{v}) \tag{25}$$

$$\text{sujeito a } E^{conf}(\mathbf{v}): h_i(\mathbf{v}) \leq 0, \quad i \in \{1..p\} \tag{26}$$

$$\mathbf{z}^{\min} \leq \mathbf{v} \leq \mathbf{z}^{\max} \tag{27}$$

onde \mathbf{v} , \mathbf{z}^{\min} , $\mathbf{z}^{\max} \in \mathfrak{R}^N$; $f(\mathbf{v})$ e $h_i(\mathbf{v})$ são funções contínuas e diferenciáveis. As condições (26) e (27) definem um conjunto fechado em \mathfrak{R}^N no qual o vetor \mathbf{v} deve permanecer para representar uma solução válida para o problema de otimização (25). Uma solução para o problema pode ser obtida por uma rede de Hopfield modificada cujo subespaço-válido garante a satisfação das condições definidas em (26). Além disso, o hipercubo inicial representado pelas restrições de desigualdade (variáveis canalizadas) em (27) é definido pela função de ativação ‘rampa-simétrica’.

Os parâmetros \mathbf{T}^{conf} e \mathbf{t}^{conf} , pertencentes ao subespaço-válido, são calculados transformando-se as restrições de desigualdade em (26) num conjunto de restrições de igualdade através da utilização de uma variável auxiliar para cada restrição de desigualdade:

$$h_i(\mathbf{v}) + \sum_{j=1}^p \delta_{ij} \cdot w_j = 0 \tag{28}$$

onde $w_j \geq 0$ são as variáveis auxiliares que podem ser tratadas como variáveis v_i , e δ_{ij} é definida pela função impulso de Kronecker dada por:

$$\delta_{ij} = \begin{cases} 1, & \text{se } i = j \\ 0, & \text{se } i \neq j \end{cases} \tag{29}$$

Após esta transformação, o problema definido pelas equações (25), (26) e (27) podem ser reescritas como:

$$\text{Minimizar } E^{ot}(\mathbf{v}^+) = f(\mathbf{v}^+) \tag{30}$$

$$\text{sujeito a } E^{conf}(\mathbf{v}^+): h_i(\mathbf{v}^+) = 0 \tag{31}$$

$$\mathbf{z}^{\min} \leq \mathbf{v}^+ \leq \mathbf{z}^{\max}, \quad i \in \{1..N\} \tag{32}$$

$$0 \leq \mathbf{v}^+ \leq \mathbf{z}^{\max}, \quad i \in \{N+1..N^+\} \tag{33}$$

onde $N^+ = N + p$, e $\mathbf{v}^{+T} = [\mathbf{v}^T \ \mathbf{w}^T] \in \mathfrak{R}^{N^+}$ é um vetor de variáveis estendidas. Deve-se notar que E^{ot} não dependerá das variáveis auxiliares \mathbf{w} .

A matriz projeção \mathbf{T}^{conf} da equação do subespaço-válido é obtida através da projeção de \mathbf{v}^+ , obtido a partir da minimização de $E^{ot}(\mathbf{v}^+) = f(\mathbf{v}^+)$, para o subespaço tangente à superfície delimitada pelas restrições dadas por (31). Assim, uma equação para \mathbf{T}^{conf} pode ser definida por (Luenberger, 1984):

$$\mathbf{T}^{conf} = \mathbf{I} - \nabla \mathbf{h}(\mathbf{v}^+)^T \cdot (\nabla \mathbf{h}(\mathbf{v}^+) \cdot \nabla \mathbf{h}(\mathbf{v}^+)^T)^{-1} \cdot \nabla \mathbf{h}(\mathbf{v}^+) \tag{34}$$

onde:

$$\nabla \mathbf{h}(\mathbf{x}) = \begin{bmatrix} \frac{\partial h_1(\mathbf{x})}{\partial x_1} & \frac{\partial h_1(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial h_1(\mathbf{x})}{\partial x_N} \\ \frac{\partial h_2(\mathbf{x})}{\partial x_1} & \frac{\partial h_2(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial h_2(\mathbf{x})}{\partial x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_p(\mathbf{x})}{\partial x_1} & \frac{\partial h_p(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial h_p(\mathbf{x})}{\partial x_N} \end{bmatrix} = \begin{bmatrix} \nabla h_1(\mathbf{x})^T \\ \nabla h_2(\mathbf{x})^T \\ \vdots \\ \nabla h_p(\mathbf{x})^T \end{bmatrix} \tag{35}$$

Substituindo-se a matriz \mathbf{T}^{conf} dada por (34) na equação do subespaço-válido (5), obtém-se:

$$\mathbf{v}^+ \leftarrow [\mathbf{I} - \nabla \mathbf{h}(\mathbf{v}^+)^T \cdot (\nabla \mathbf{h}(\mathbf{v}^+) \cdot \nabla \mathbf{h}(\mathbf{v}^+)^T)^{-1} \cdot \nabla \mathbf{h}(\mathbf{v}^+)] \cdot \mathbf{v}^+ + \mathbf{i}^{conf} \tag{36}$$

Por fim, resultados da teoria de estabilidade de Lyapunov (Vidyasagar, 1992) devem ser introduzidas na equação (36) para garantir a estabilidade do sistema não-linear, e conseqüentemente, forçar a convergência da rede para os pontos de equilíbrio que representam uma solução para o sistema. Pela definição de Jacobiano, quando \mathbf{v} tende ao ponto de equilíbrio implica que $\mathbf{v}^e = \mathbf{0}$. Neste caso, o valor de \mathbf{i}^{conf} deve ser também nulo, pois caso contrário a igualdade não é satisfeita para a condição de equilíbrio, ou seja, $\mathbf{v}^e = \mathbf{v}(t) = \mathbf{v}(t + \Delta t) = \mathbf{0}$. Assim, a equação (31) pode ser aproximada por:

$$\mathbf{h}(\mathbf{v}^+) \approx \mathbf{h}(\mathbf{v}^e) + \mathbf{J} \cdot (\mathbf{v}^+ - \mathbf{v}^e) \tag{37}$$

onde: $\mathbf{J} = \nabla \mathbf{h}(\mathbf{v}^+)$ e $\mathbf{h}(\mathbf{v}^+) = [h_1(\mathbf{v}^+) \ h_2(\mathbf{v}^+) \ \dots \ h_p(\mathbf{v}^+)]^T$

Na vizinhança do ponto de equilíbrio $\mathbf{v}^e = \mathbf{0}$, obtém-se a seguinte equação:

$$\lim_{\mathbf{v}^+ \rightarrow \mathbf{v}^e} \frac{\|\mathbf{h}(\mathbf{v}^+)\|}{\|\mathbf{v}^+\|} = 0 \tag{38}$$

Introduzindo-se os resultados das equações (37) e (38) na equação (36), obtém-se:

$$\mathbf{v}^+ \leftarrow \mathbf{v}^+ - \nabla \mathbf{h}(\mathbf{v}^+)^T \cdot (\nabla \mathbf{h}(\mathbf{v}^+) \cdot \nabla(\mathbf{v}^+)^T)^{-1} \cdot \mathbf{h}(\mathbf{v}^+) \tag{39}$$

Portanto, a equação (39) sintetiza a equação do subespaço-válido para sistemas de equações não-lineares. Neste caso, a equação original do subespaço-válido em (5), o qual está representado pelo passo (I) na Figura 1, deve ser substituída pela equação (39). A aplicação sucessiva do passo (I) seguido do passo (II), ilustrado na Figura 1, faz com que \mathbf{v} seja uma solução que satisfaça todas as restrições impostas pelo problema de otimização não-linear.

Os parâmetros \mathbf{T}^{ot} e \mathbf{i}^{ot} do termo de energia E^{ot} , dado por (7) e representado em (30), são definidos com o objetivo de que a solução ótima corresponda à minimização de E^{ot} . Este procedimento pode ser feito alterando-se o vetor \mathbf{v}^+ na direção oposta ao gradiente da função de energia E^{ot} . Assim, os pontos de equilíbrio da rede podem ser calculados assumindo-se os seguintes valores para \mathbf{T}^{ot} e \mathbf{i}^{ot} :

$$\mathbf{i}^{ot} = - \left[\begin{array}{cccc} \frac{\partial f(\mathbf{v}^+)}{\partial v_1} & \frac{\partial f(\mathbf{v}^+)}{\partial v_2} & \dots & \frac{\partial f(\mathbf{v}^+)}{\partial v_{N^+}} \end{array} \right]^T \tag{40}$$

$$\mathbf{T}^{ot} = \mathbf{0} \tag{41}$$

Vale aqui ressaltar que a obtenção das expressões de \mathbf{T}^{conf} e \mathbf{i}^{conf} , as quais caracterizam o subespaço válido de soluções, é realizada a partir do conhecimento a priori sobre o inter-relacionamento envolvendo as variáveis do problema. O processo envolvido com a especificação de \mathbf{T}^{conf} e \mathbf{i}^{conf} deve levar em consideração as propriedades do subespaço válido, ou seja, $\mathbf{T}^{conf} \cdot \mathbf{T}^{conf} = \mathbf{T}^{conf}$ e $\mathbf{T}^{conf} \cdot \mathbf{i}^{conf} = \mathbf{0}$.

Conforme mencionado anteriormente, visto que $\mathbf{v}^{+T} = [\mathbf{v}^T \ \mathbf{w}^T]$ então o vetor \mathbf{i}^{ot} dado em (40) seria representado por:

$$\mathbf{i}^{ot} = - \left[\begin{array}{cccccc} \frac{\partial f(\mathbf{v}^+)}{\partial v_1} & \frac{\partial f(\mathbf{v}^+)}{\partial v_2} & \dots & \frac{\partial f(\mathbf{v}^+)}{\partial v_N} & \frac{\partial f(\mathbf{v}^+)}{\partial w_1} & \frac{\partial f(\mathbf{v}^+)}{\partial w_2} & \dots & \frac{\partial f(\mathbf{v}^+)}{\partial w_p} \end{array} \right]^T \tag{42}$$

Como o processo de otimização de E^{ot} não depende dos valores das variáveis auxiliares \mathbf{w} , então a equação dada em (42) pode ser substituída pela seguinte expressão:

$$\mathbf{i}^{ot} = - \left[\begin{array}{cccc} \frac{\partial f(\mathbf{v}^+)}{\partial v_1} & \frac{\partial f(\mathbf{v}^+)}{\partial v_2} & \dots & \frac{\partial f(\mathbf{v}^+)}{\partial v_N} \\ \underbrace{\hspace{10em}}_{N\text{-componentes}} & \underbrace{\hspace{10em}}_{p\text{-componentes}} & \dots & \underbrace{\hspace{10em}}_{p\text{-componentes}} \end{array} \right]^T \tag{43}$$

A fim de sintetizar todos os passos envolvidos com o mapeamento de problemas de otimização restrita através da rede de Hopfield modificada, ilustra-se na Figura 2 os passos algorítmicos básicos envolvidos com a abordagem proposta.

1. **Iniciar** o vetor \mathbf{v} com valores aleatórios.
2. **Introduzir** as variáveis auxiliares no vetor \mathbf{v} .
3. **Converter** o vetor \mathbf{v} em \mathbf{v}^+ .
4. **Repita**
 - Enquanto** (\mathbf{v}^+ não convergir) **faça**
 - Obter** o valor de $\mathbf{h}(\mathbf{v}^+)$.
 - Obter** a matriz Jacobiana $\nabla \mathbf{h}(\mathbf{v}^+)$ a partir de (35).
 - Atualizar** o valor de \mathbf{v}^+ a partir de (39).
 - Aplicar** a função de ativação dada em (9).
 - Fim_enquanto**
 - Obter** o vetor \mathbf{i}^{ot} a partir de (43).
 - Atualizar** o valor de \mathbf{v}^+ utilizando (10).
 - Até que** (\mathbf{v}^+ fique estacionário).
5. **Fim.**

Figura 2 – Algoritmo básico da rede de Hopfield modificada.

Para ilustrar o desempenho da rede neural proposta, alguns resultados de simulação são apresentados na próxima seção.

5. Resultados de Simulação

A rede de Hopfield proposta nas seções anteriores foi utilizada para resolver três problemas de otimização não-linear restrita propostos na literatura.

O primeiro problema, conforme proposto em Bazaraa & Shetty (1979), a função objetivo está sujeita somente às restrições de desigualdade, ou seja:

$$\text{Min } f(\mathbf{v}) = e^{v_1} + v_1^2 + 4v_1 + 2v_2^2 - 6v_2 + 2v_3 \quad (44)$$

$$\text{sujeito a: } v_1^2 + e^{v_2} + 6v_3 \leq 15 \quad (C1)$$

$$v_1^4 - v_2 + 5v_3 \leq 25 \quad (C2)$$

$$v_1^3 + v_2^2 - v_3 \leq 10 \quad (C3)$$

$$0 \leq v_1 \leq 4 \quad (C4)$$

$$0 \leq v_2 \leq 2 \quad (C5)$$

$$v_3 \geq 0 \quad (C6)$$

Para este problema, com três restrições de desigualdade e com variáveis limitadas, o vetor solução (ponto de equilíbrio) obtido após a convergência da rede de Hopfield modificada é dado por $\mathbf{v} = [0,00001 \ 1,50002 \ 0,00000]^T$, com função objetivo $f(\mathbf{v}) = -3,49999$. Estes resultados são bem próximos ao valor da solução ótima fornecida por $\mathbf{v}^* = [0,00000 \ 1,50000 \ 0,00000]^T$ e $f(\mathbf{v}^*) = -3,50000$.

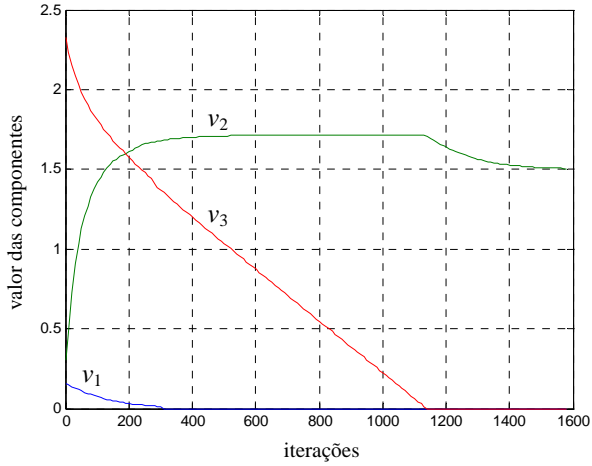


Figura 3 – Evolução do vetor de saída da rede (1º problema).

A Figura 3 ilustra a evolução dos valores de v_1 , v_2 e v_3 obtidos pela rede de Hopfield modificada em relação ao número de iterações. O esboço do comportamento da função objetivo do problema é apresentado na Figura 4. As restrições dadas em (C4), (C5) e (C6) são diretamente tratadas pela função de ativação da rede que é definida em (9).

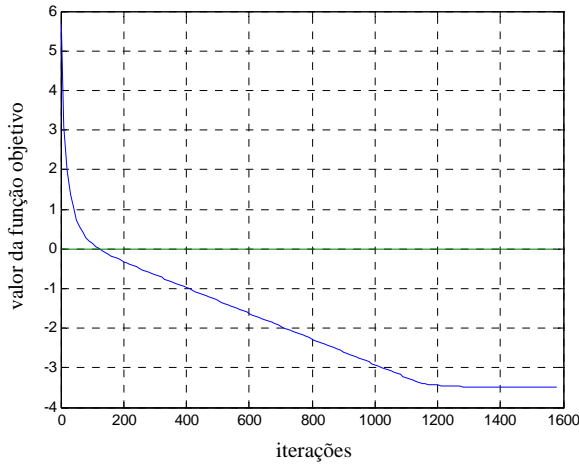


Figura 4 – Evolução da função objetivo (1º problema).

O segundo problema tratado pela rede de Hopfield modificada é constituído por restrições de igualdade e desigualdade, sendo o mesmo definido pelas seguintes expressões:

$$\text{Min } f(\mathbf{v}) = v_1^3 + 2v_2^2 \cdot v_3 + 2v_3$$

$$\text{sujeito a: } v_1^2 + v_2 + v_3^2 = 4$$

$$v_1^2 - v_2 + 2v_3 \leq 2$$

$$v_1, v_2, v_3 \geq 0$$

Em relação a este problema, a Figura 5 mostra a evolução dos valores de v_1 , v_2 e v_3 em função do número de iterações. O vetor solução obtido, após a convergência da rede, é dado por $\mathbf{v} = [0,00000 \ 4,00000 \ 0,00001]^T$, sendo o valor da função objetivo $f(\mathbf{v}) = E^{ot}(\mathbf{v}) = 0.00034$. A solução ótima para o problema é fornecida por $\mathbf{v}^* = [0,00000 \ 4.00000 \ 0.00000]^T$, com $E^{ot}(\mathbf{v}^*) = 0,00000$.

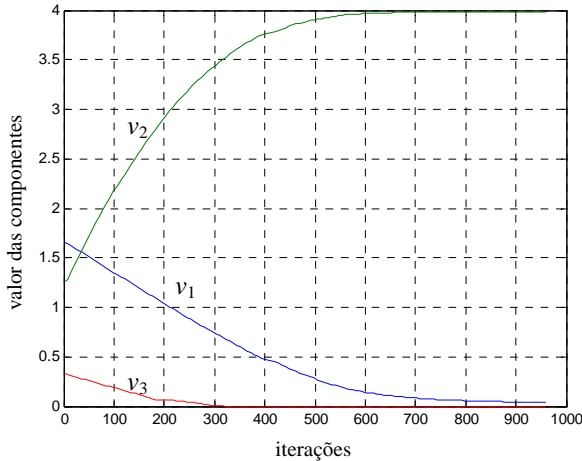


Figura 5 – Evolução do vetor de saída da rede (2º problema).

A Figura 6 ilustra o comportamento da função objetivo do problema em função do número de iterações. Os valores iniciais atribuídos ao vetor \mathbf{v} foram gerados aleatoriamente entre zero e um.

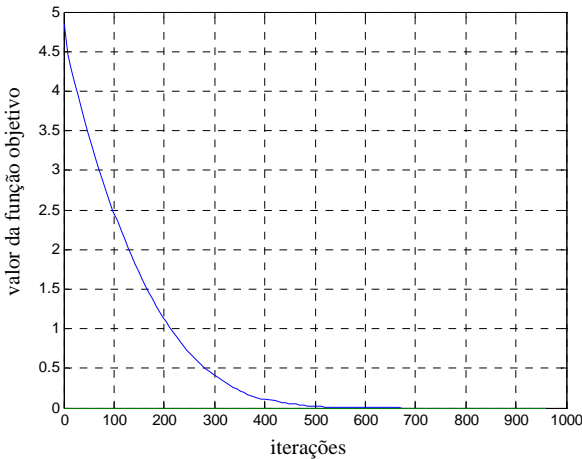


Figura 6 – Evolução da função objetivo (2º problema).

O terceiro problema, conforme proposto em Kennedy & Chua (1988), visa confirmar a aplicabilidade da rede proposta quando todas as restrições associadas ao mesmo são do tipo linear, ou seja:

$$\begin{aligned} \text{Min } f(\mathbf{v}) &= 0,4v_2 + v_1^2 + v_2^2 - v_1v_2 + \frac{1}{30}v_1^3 \\ \text{sujeito a: } v_1 + 0,5v_2 &\geq 0,4 \\ 0,5v_1 + v_2 &\geq 0,5 \\ v_1, v_2 &\geq 0 \end{aligned}$$

A Figura 7 apresenta a evolução dos valores dos elementos do vetor de saída da rede em relação ao número de iterações. Após a convergência da rede, o vetor solução obtido é dado por $\mathbf{v} = [0,3398 \ 0,3301]^T$, com $E^{ot}(\mathbf{v}) = 0,2456$. Estes resultados são bem próximos aos valores da solução exata fornecida por $\mathbf{v}^* = [0,3395 \ 0,3302]^T$, com função objetivo $E^{ot}(\mathbf{v}^*) = 0,2455$.

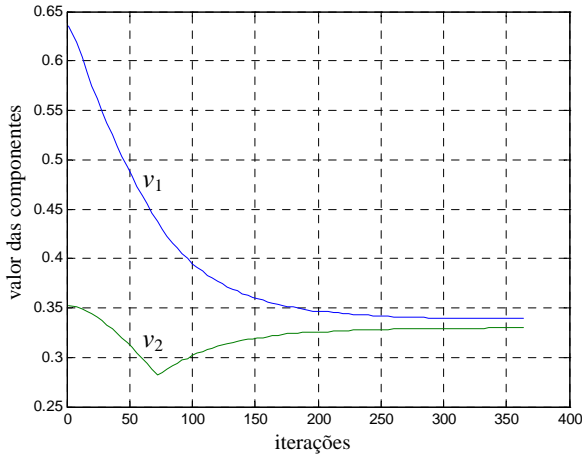


Figura 7 – Evolução do vetor de saída da rede (3º problema).

A Figura 8 apresenta o comportamento da função objetivo do problema em função do número de iterações. Neste caso, a rede foi iniciada com valores aleatórios, uniformemente distribuídos, entre zero e um.

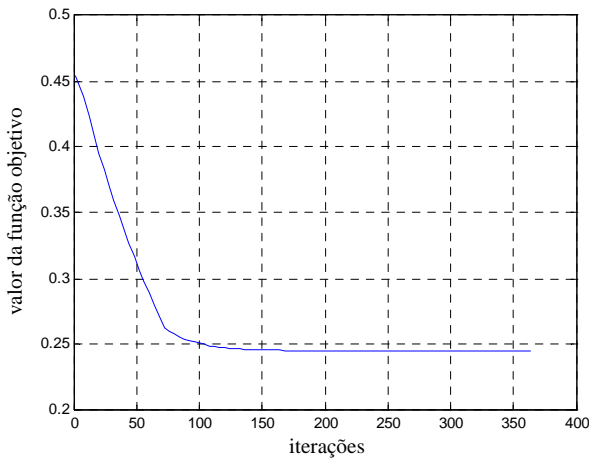


Figura 8 – Evolução da função objetivo (3º problema).

Estes resultados mostram a aplicação efetiva da rede de Hopfield modificada para resolver problemas de otimização não-linear. A rede foi também simulada considerando diversos valores iniciais atribuídos ao vetor de saída v . Para todas as simulações, a rede sempre convergiu para as mesmas soluções.

Embora o número de iterações seja elevado o tempo gasto para a obtenção das soluções tem sido bastante curto. Em relação a este aspecto vale ressaltar que embora tenhamos utilizado computadores seqüenciais em nossas simulações, o desempenho da rede alcançará eficiência máxima quando a mesma for executada em computadores com processadores operando em paralelo, pois neste caso cada unidade de processamento corresponderia a uma unidade neural.

Outro aspecto relevante diz respeito à qualidade das soluções obtidas pela rede. Embora os valores das soluções apresentem alguns desvios em relação aos valores ótimos, observa-se que os mesmos são menores quando comparados com outras abordagens neurais utilizadas nesses tipos de problemas.

É também motivo de registro que alguns testes realizados com problemas de dimensões maiores, tendo cerca de 800 restrições de desigualdade, mostraram que a rede pode tornar-se lenta em sua procura pelos pontos de equilíbrio que representam a solução do sistema. Este comportamento deve-se ao fato de estarmos realizando toda a simulação em computadores seqüenciais. A utilização de máquinas com processadores em paralelo, conforme mencionado anteriormente, pode contornar esta situação.

Finalmente, o mapeamento de problemas de otimização não-linear restrita através da rede de Hopfield modificada apresenta algumas particularidades em relação aos métodos primais (Bazaraa & Shetty, 1979) que são também utilizados na solução destes problemas, podendo se destacar as seguintes: i) a não necessidade do cálculo do conjunto ativo de restrições em cada iteração; ii) o vetor de inicialização da rede (solução inicial) não precisa pertencer ao conjunto factível definido pelas restrições; iii) não necessidade de obtenção, em cada iteração, de uma direção admissível de busca; e iv) mecanismo de busca independente da utilização de multiplicadores de Lagrange.

6. Conclusões

Neste artigo, uma rede de Hopfield modificada foi desenvolvida para resolver problemas de otimização não-linear restrita. Os resultados de simulação mostraram que a rede proposta fornece resultados expressivos, tornando assim um método alternativo para resolver estes problemas eficientemente. Os parâmetros internos da rede foram obtidos utilizando conceitos relativos às teorias de otimização não-linear e de estabilidade de Lyapunov. Todas as restrições associadas ao problema são mapeadas através do subespaço-válido de soluções.

O mapeamento de problemas de otimização não-linear restrita através da rede de Hopfield modificada além de fornecer um novo método para resolver estes problemas, não requer a definição (ajuste) de nenhum parâmetro de penalidade ou ponderação.

Referências Bibliográficas

- (1) Aiyer, S.V.B.; Niranja, M. & Fallside, F. (1990). A theoretical investigation into the performance of the Hopfield model. *IEEE Transactions on Neural Networks*, **1**, 53-60.
- (2) Bazaraa, M.S. & Shetty, C.M. (1979). *Nonlinear Programming*. John Wiley & Sons, New York.
- (3) Hopfield, J.J. (1984). Neurons with a graded response have collective computational properties like those of two-state neurons. *Proc. of the National Academy of Science*, **81**, 3088-3092.
- (4) Kennedy, M.P & Chua, L.O. (1988). Neural networks for nonlinear programming. *IEEE Transactions on Circuits and Systems*, **35**, 554-562.
- (5) Luenberger, D.G. (1984). *Linear and Nonlinear Programming*. Addison-Wesley, Reading, Massachusetts.
- (6) Silva, I.N. (1997). Uma Abordagem Neuro-nebulosa Para Otimização de Sistemas e Identificação Robusta. Tese de Doutorado, UNICAMP/FEEC/DCA.
- (7) Silva, I.N.; Arruda, L.V.R. & Amaral, W.C. (1997). Robust estimation of parametric membership regions using artificial neural networks. *International Journal of Systems Sciences*, **28**, 447-455.
- (8) Vázquez, A.R.; Castro, R.D.; Rueda, A.; Huertas, J.L. & Sinencio, S. (1990). Nonlinear switched-capacitor neural network for optimization problems. *IEEE Transactions on Circuits and Systems*, **37**, 384-398.
- (9) Vidyasagar, M. (1992). *Nonlinear Systems Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey.
- (10) Zhang, S.; Zhu, X. & Zou, L.H. (1992). Second-order neural nets for constrained optimization. *IEEE Transactions on Neural Networks*, **6**, 1021-1024.

Apêndice

A estabilidade assintótica da rede Hopfield pode ser mostrada utilizando o segundo método de Lyapunov (Vidyasagar, 1992), isto é, mostrando que o sistema dissipa energia com o passar do tempo. Para isso, dada uma função de Lyapunov (ou função de energia) associada à rede, mostra-se que suas derivadas temporais são não crescentes. Uma função de Lyapunov para a rede de Hopfield é definida por:

$$E(t) = -\frac{1}{2} \mathbf{v}(t)^T \cdot \mathbf{T} \cdot \mathbf{v}(t) - \mathbf{v}(t)^T \cdot \mathbf{i}^b + \eta \sum_{i=1}^N \int_0^{v_i(t)} g^{-1}(\alpha) \cdot d\alpha \quad (45)$$

cujas derivadas temporais são dadas por:

$$\dot{E}(t) = \frac{dE(t)}{dt} = (\nabla_{\mathbf{v}} E(t))^T \cdot \dot{\mathbf{v}}(t) \quad (46)$$

onde $\nabla_{\mathbf{v}}$ é o operador gradiente com relação a \mathbf{v} . Para a condição de \mathbf{T} simétrica, tem-se:

$$\nabla_v E(t) = -\mathbf{T} \cdot \mathbf{v}(t) - \mathbf{i}^b + \eta \cdot \mathbf{u}(t) \quad (47)$$

Comparando a equação (1) com (47), conclui-se que:

$$\nabla_v E(t) = -\dot{\mathbf{u}}(t) \quad (48)$$

Inserindo o resultado de (48) em (46), obtém-se:

$$\begin{aligned} E(t) &= -\dot{\mathbf{u}}(t)^T \cdot \dot{\mathbf{v}}(t) = \\ &= -\sum_{i=1}^N \dot{u}_i(t) \cdot \dot{v}_i(t) = \sum_{i=1}^N \dot{u}_i(t) \cdot \frac{\partial v_i}{\partial u_i} \cdot \frac{\partial u_i}{\partial t} = \\ &= -\sum_{i=1}^N (\dot{u}_i(t))^2 \cdot \frac{\partial v_i}{\partial u_i} \end{aligned} \quad (49)$$

Como $(\dot{u}_i(t))^2$ é sempre positivo, precisa-se verificar apenas que $\partial v_i / \partial u_i \geq 0$ para se mostrar que $\dot{E}(t) \leq 0$. Isto é trivialmente verdadeiro desde que a função de ativação da rede dada em (2) seja monótona crescente. A partir da equação (48), nota-se também que $\dot{E}(t) = 0$ apenas quando $\dot{\mathbf{u}}(t) = 0$. Isto completa a prova de estabilidade assintótica para a rede de Hopfield.