

TÉCNICA DE ACELERAÇÃO PARA O AG APLICADO À SÍNTESE SE DISPOSITIVOS ÓPICOS MULTICAMADAS

Gilvan Borges

Marco José de Sousa

gilvan82@yahoo.com.br

marcojs@ufpa.br

Departamento de Engenharia Elétrica e Computação, Universidade Federal do Pará
Av. Augusto Corrêa, n.1, 66075-110, Belém - PA – Brazil

João Crisóstomo Weyl A Costa

jweyl@ufpa.br

Departamento de Engenharia Elétrica e Computação, Universidade Federal do Pará

Resumo. *Este trabalho apresenta uma técnica de aceleração para o algoritmo genético (AG) aplicado na síntese de filtros ópticos multicamadas. A técnica baseia-se na redução da quantidade de pontos de análise (amostras) necessários para a função objetivo (FO). Para contrabalançar a perda de confiabilidade da FO, devido a redução do número de amostras, este artigo propõe um mecanismo similar a um controle de qualidade aplicado às populações a cada geração do AG. A redução de amostras e o controle de qualidade combinados aumentaram o desempenho computacional do AG em até 1000%. Embora aplicado à síntese de filtro ópticos, este artigo sustenta que a técnica de aceleração poderia ser aplicada em outras espécies de problemas de síntese ou otimização com poucas adaptações.*

Palavras-chave: *Filtros ópticos, Multicamadas, Algoritmo Genético, Aceleração computacional.*

1. INTRODUÇÃO

Filtros de multicamadas ou de filmes finos são um tipo de filtro óptico empregado normalmente para transmitir ou bloquear a luz, com mais ou menos intensidade, em função da frequência (ou do comprimento de onda). Possuem larga aplicação, como por exemplo, em redes de comunicação ópticas, espectrografia, instrumentos ópticos diversos e em sensores ópticos de várias espécies.

Os filtros multicamadas possuem dimensões reduzidas, são mecanicamente e termicamente robustos e, conseqüentemente, são facilmente manipuláveis e aplicáveis. Sua fabricação é realizada através da deposição controlada de diferentes materiais dielétricos sobre um substrato. Os filmes finos resultantes da deposição possuem espessuras da ordem do comprimento de onda da luz que, para aplicações no espectro visível, situa-se em torno de 500 nm. Tipicamente as camadas são constituídas por apenas dois materiais de índices de refração diferentes alternados. A escolha dos materiais, da quantidade de camadas e da espessura de cada uma delas determina o comportamento do filtro em função da frequência, do ângulo de incidência e da polarização da luz.

Dependendo do comportamento desejado para um filtro óptico, existem algumas metodologias para determinar a combinação adequada de materiais e espessuras. Estruturas de camadas de $\frac{1}{4}$ de comprimento de onda de espessura e cavidades Fabry-Perot podem originar filtros rejeita-faixa simples ou passa-faixa múltiplos. Porém, filtros arbitrários, como os normalmente empregados em equalização, exigem uma combinação de materiais e espessuras freqüentemente imprevisíveis.

Para projetar filtros de filmes finos complexos, frequentemente tem-se empregado algoritmos evolutivos como o algoritmo genético (AG) [1]. Dentre todas as meta-heurísticas mais conhecidas, o AG é certamente um algoritmo bem estabelecido e documentado [2], embora dele existam muitas espécies (cuja classificação ainda é tema de divergências). Por outro lado, o problema de síntese de filmes finos empregando AG pode ser consideravelmente complexo e computacionalmente ostensivo, mesmo empregando modelos de análise simples como os matriciais [3], [4]. É necessário desenvolver técnicas especializadas para permitir otimização e síntese de filmes finos de forma eficiente e desprovida de supervisão humana. Para este propósito, este artigo apresenta uma série de modificações sobre um AG baseado no modelo canônico de Schwefel [5].

O objetivo das modificações é uma substancial e direta melhoria do desempenho computacional do algoritmo. A principal característica dessas modificações é permitir ao AG manipular duas variantes da mesma função objetivo: uma computacionalmente ostensiva e ao mesmo tempo apurada (FOE - Função Objetivo Estendida); outra pouco ostensiva e pouco apurada (FOR - Função Objetivo Reduzida). O AG modificado utiliza normalmente a versão da função objetivo FOR enquanto utiliza com economia a FOE, obtendo, dessa forma, um ganho de processamento que pode chegar a 1000% em relação ao AG original.

Para a diferenciação entre as duas versões da função objetivo FOR e FOE, este artigo propõe a utilização de diferentes quantidades de amostras no processo de levantamento da curva de refletividade do filtro. A FOR costuma utilizar aproximadamente de 1 a 10% da quantidade de amostras utilizada pela FOE, sendo esta percentagem fundamental para a estimação do ganho computacional possível. Evidentemente a utilização de muito poucas amostras por parte da FOR pode acarretar uma série de efeitos indesejáveis sobre o AG. Este artigo propõe contornar estes problemas através do uso de *amostras móveis* por parte da FOR e do uso de um *procedimento de correção* elitista ao qual a FOE está subordinada.

2. O ALGORITMO GENÉTICO

2.1. Codificação

A formulação adotada neste artigo para a análise de filtros multicamadas foi a matricial frequentemente referenciada como método das matrizes características, cujo desenvolvimento é mostrado em [3] e [4]. Considera-se que a operação fundamental deste método de análise seja o cálculo, para a incidência da luz normal, do coeficiente de reflexão $r(\lambda, \mathbf{X})$ em função do comprimento de onda λ e de um filtro representado por \mathbf{X} :

$$\mathbf{X} = \{ \mathbf{X}_1; \mathbf{X}_2 \} = \{ a_1, a_2, \dots, a_k, \dots, a_N; b_1, b_2, \dots, b_k, \dots, b_N \}. \quad (1)$$

Onde os parâmetros a_k e b_k representam, respectivamente, o índice de refração e a espessura da camada k do filtro. O subscrito k pode variar de 1 (primeira camada) ao número total de camadas N (última seção). Se a cobertura (onde a luz é injetada) e o substrato fossem incluídos em (1), seus subscritos seriam respectivamente 0 e $N + 1$.

Empregando a codificação real, o vetor \mathbf{X} poderia ser mapeado, elemento a elemento de \mathbf{X}_1 e \mathbf{X}_2 , diretamente como um vetor cromossomo \mathbf{Y} . Entretanto, este artigo considera que as camadas possuem apenas dois valores de índice de refração alternados, L ou H , respectivamente o valor mínimo (*Low*) e máximo (*High*). Isso possibilita uma severa simplificação do cromossomo, pois torna-se praticamente desnecessário o mapeamento de \mathbf{X}_1 em \mathbf{Y} . Apenas um elemento (gene) de \mathbf{Y} é necessário para estabelecer qual dos materiais, L ou H , iniciará a seqüência. Adicionalmente, mais um gene pode ser incluído em \mathbf{Y} para representar o comprimento do filtro, permitindo que mais essa característica possa ser determinada pelo AG. Assim, o cromossomo \mathbf{Y} pode ser definido como se segue:

$$\mathbf{Y} = \{ A; B_1, B_2, B_3, \dots, B_k, \dots, B_N; C \} \quad (2)$$

onde os genes A , B_k , e C são valores reais pertencentes ao intervalo $[0; 1]$. O filtro \mathbf{X} pode ser obtido de \mathbf{Y} através das seguintes relações:

$$a_k = \begin{cases} L + (H - L) \cdot \text{round}(A), & \text{para } k = 1; \\ (L \cdot H) / a_{k-1}, & \text{para } k > 1 \end{cases} \quad (3)$$

$$b_k = b_{\min} + (b_{\max} - b_{\min}) \cdot B_k \quad (4)$$

Onde $\text{round}(\cdot)$ representa uma função de arredondamento para o valor inteiro mais próximo, retornando 0 para um argumento menor que 0.5 e 1 para um argumento maior ou igual 0.5. Os parâmetros b_{\min} e b_{\max} são respectivamente a espessura mínima e máxima permitida para uma camada.

2.2. O AG e o procedimento de correção elitista

O AG proposto é resultado da adaptação do modelo canônico de Schwefel [5]. As diferenças entre os algoritmos são: a inclusão de um parâmetro E para o controle do elitismo; inclusão de um procedimento para a melhoria do desempenho computacional. Um diagrama do AG modificado é mostrado na Fig. 1 que mostra como os procedimentos de seleção, recombinação-mutação e avaliação relacionam-se ao longo de uma geração i , que processa a população \mathbf{P}_i criando a partir desta a população da geração seguinte \mathbf{P}_{i+1} . Primeiramente o procedimento de seleção por torneio [6], bloco S , é repetido M vezes a fim de gerar uma

população intermediária $\mathbf{P}_{i+1/2}$. Em seguida a recombinação uniforme e a mutação gaussiana, o bloco R/Mut, são aplicadas tendo como progenitores dois indivíduos de $\mathbf{P}_{i+1/2}$ e gerando um indivíduo descendente que é avaliado através da função objetivo FOR e incorporado à população intermediária \mathbf{P}_{i+1} . A aplicação de R/Mut e de FOR são repetidas Q vezes. Portanto, o tamanho da população \mathbf{P}_{i+1} é de pelo menos Q . Porém E melhores indivíduos de \mathbf{P}_i podem ser incorporados em \mathbf{P}_{i+1} (elitismo) e K elementos da população intermediária também podem ser incorporados, chegando a um tamanho de população \mathbf{P}_{i+1} de $E + K + Q$.

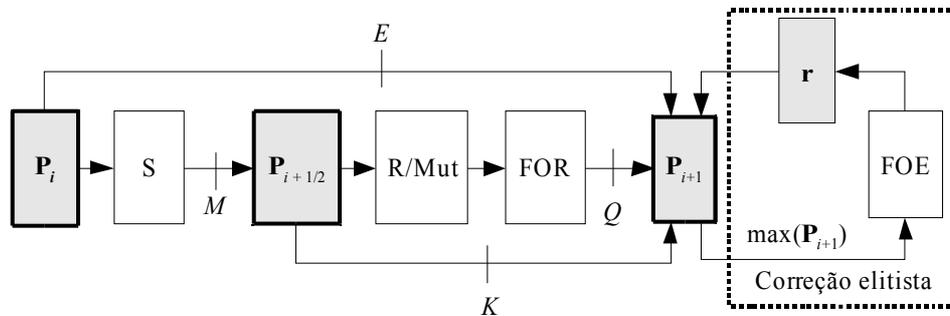


Figura 1 - Diagrama para o AG modificado.

O critério para a transferência de K indivíduos de $\mathbf{P}_{i+1/2}$ para \mathbf{P}_{i+1} é baseado no tempo de vida. Cada indivíduo, além do valor retornado pela função objetivo, tem associado a si mais um campo representando seu tempo restante de vida medido em números de gerações. Cada novo indivíduo tem esse campo iniciado com um valor κ que será decrementado a cada geração e então, quando o valor do campo atinge o valor mínimo zero, o indivíduo não poderá mais ser adicionado em \mathbf{P}_{i+1} a partir de $\mathbf{P}_{i+1/2}$. Quando todos os indivíduos selecionados possuem um tempo de vida restante maior que zero a população seguinte \mathbf{P}_{i+1} terá o tamanho máximo $Q+M+E$. O outro caso extremo ocorre quando todos os indivíduos de $\mathbf{P}_{i+1/2}$ possuem um tempo de vida zero, fazendo com que o tamanho de \mathbf{P}_{i+1} seja de apenas $Q+E$.

Além das operações relativamente comuns a todos algoritmos genéticos, a Fig. 1 exibe um bloco extra que interage apenas com a população \mathbf{P}_{i+1} . Este bloco representa o procedimento de correção elitista. Uma vez formada a população \mathbf{P}_{i+1} , o melhor indivíduo, filtrado através do operador $\max(\cdot)$, será submetido à FOE, que lhe atribuirá um outro valor objetivo certamente diferente do valor anterior conferido pela FOR. O novo valor objetivo é comparado ao do indivíduo de referência \mathbf{r} : caso seja melhor, o indivíduo extraído da população tomará o lugar de \mathbf{r} ; caso contrário, se o melhor indivíduo de \mathbf{P}_{i+1} (de acordo com a FOR) for inferior ao indivíduo \mathbf{r} , o indivíduo de referência será reincorporado à \mathbf{P}_{i+1} .

Uma vez que a FOR não pode ser tão confiável quanto a FOE, o processo de seleção baseado na FOR pode resultar teoricamente em uma população \mathbf{P}_{i+1} inferior a \mathbf{P}_i . O processo de correção proposto permite combater a possibilidade de retrocesso. Uma extensão natural do raciocínio seria a utilização não de apenas 1 indivíduo de referência, mas vários. A utilização de uma população referencial poderia ser útil no processo de otimização multiobjetivo, no qual usualmente procura-se por um frente de soluções (frente de Pareto) [7].

2.3. Diferenciação entre FOR e FOE

Neste artigo, o ponto de partida adotado tanto para a FOR quanto para a FOE é o inverso do erro médio quadrático:

$$F(\mathbf{X}) = \left\{ \sum_{n=1}^{n_{max}} \frac{[|r(\lambda_n, \mathbf{X})|^2 - R(\lambda_n)]}{n_{max}} \right\}^{-1} \quad (5)$$

Onde $|r(\lambda_n, \mathbf{X})|^2$ define a refletividade para o comprimento de onda λ_n e para a grade representada por \mathbf{X} . O parâmetro n_{max} representa o número de pontos (amostras) em que as duas curvas de refletividade, a calculada $|r(\lambda_n, \mathbf{X})|^2$ e a desejada $R(\lambda_n)$, são comparadas. Presume-se que a maior parte do esforço computacional deverá concentrar-se no cálculo da refletividade, de modo que todo o restante do processamento associado aos procedimentos convencionais do AG possam ser negligenciados. Portanto, o tempo de processamento por geração deverá ser proporcional a n_{max} .

Ambas as funções objetivo, FOR e FOE, podem ser versões de (5) utilizando respectivamente $n_{max} = n_R$ e $n_{max} = n_E$, onde $n_R < n_E$. O valor adequado para n_E é o mesmo utilizado em um AG normal desprovido de aceleração, enquanto n_R situa-se normalmente entre 1 e 10% de n_E . Isso permite com que a FOR seja entre 10 e 100 vezes mais rápida do que a FOE. Por outro lado, AG acelerado será mais rápido por um fator f inversamente proporcional a n_R , que pode ser estimado através da relação entre o número total de amostras processadas por geração para o AG normal e o número de amostras processadas por geração para o AG acelerado:

$$f = \frac{n_E \times Q}{n_E + n_R \times Q} \quad (6)$$

Onde Q é o mesmo parâmetro da Fig. 1.

Entretanto, a simples redução do número de amostras para a diferenciação da FOR pode prejudicar o funcionamento do AG. Utilizando sempre as mesmas posições das amostras para a comparação entre as curvas de refletividade calculada e desejada através de (5), o AG deve evoluir realizando o ajuste da refletividade ao alvo somente nos pontos coincidentes às amostras. Devido a isso, o AG provavelmente apresentará uma solução falsa ao final do processo, pois a curva da grandeza ajustada pelo AG provavelmente divergir da curva alvo quando ambas forem comparadas sob uma amostragem rigorosa.

Para evitar eventuais problemas relacionados ao excesso de espaçamento entre amostras, este trabalho propõe a utilização de *amostras móveis*, onde, a cada geração, as posições das amostras no espectro modificam-se aleatoriamente. Assim, para a FOR, utiliza-se n_R amostras uniformemente espaçadas de $\Delta\lambda = (\lambda_{max} - \lambda_{min})/n_R$, com a primeira amostra λ_1 distante de λ_{min} de d , sendo d uma variável aleatória, com distribuição uniforme entre 0 e $\Delta\lambda$. A Fig. 2 ilustra como as amostras podem ser posicionadas em função do deslocamento aleatório d modificado a cada geração.

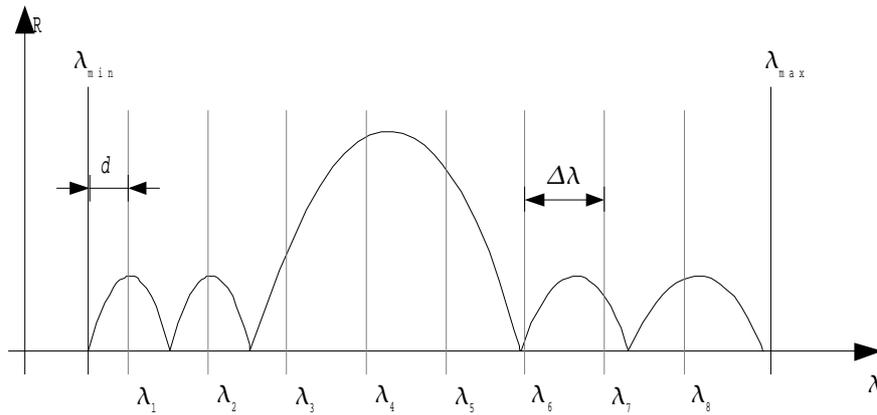


Figura 2 - Distribuição uniforme de amostras em função de uma distância d aleatória a partir de λ_{min} .

Particularmente essa forma de diferenciação entre a FOR e a FOE pode fazer com que os $E+K$ indivíduos transferidos da população \mathbf{P}_i e $\mathbf{P}_{i+1/2}$ precisem ser reavaliados pela FOR, uma vez que o valor d se modificará desatualizando-os. Neste caso, (6) precisa ser reescrita respeitando essa observação:

$$f = \frac{n_E \times Q}{n_E + n_R \times (Q + E + K)} \quad (7)$$

3. RESULTADOS

Considere o projeto de um filtro para a máxima reflexão entre os comprimentos de onda 500 nm e 550 nm (espectro visível). O AG proposto foi aplicado tendo como parâmetros: $M = 50$, $Q = 50$, $E = 1$, $K = 0$ ($\kappa = 0$), $n_R = 10$, $n_E = 1000$, $N = 10$ (10 camadas), índices de refração (não dispersivos) $L = 1.25$ e $H = 1.75$, substrato e cobertura de índices de refração iguais a 1.5, b_{min} e b_{max} iguais a 25 nm e 300 nm respectivamente. As figuras 3 e 4 mostram respectivamente a curva de refletividade e o perfil de índices de refração obtidos para o filtro sintetizado.

Para este mesmo projeto, mas utilizando diferentes combinações de n_{max} , n_R e n_E , o AG foi executado 10 vezes para o cálculo de cada um dos valores médios da função objetivo (FO) mostrados em função da quantidade de amostras (n_{max} ou n_R) nas figuras 5 e 6. Para a Fig. 5, não foi empregada a técnica de amostras móveis, nem foi utilizado o procedimento de correção elitista. Utilizou-se n_{max} igual a 100, 300, 500, 750 e 1000. Para a Fig. 6, foi empregada a técnica de amostras móveis juntamente com o procedimento de correção elitista utilizando $n_E = 1000$ (para todos os pontos) e n_R igual a 1, 5, 10, 20, 50 e 100. Para permitir comparações, os filtros obtidos nas execuções para os propósitos das figuras 5 e 6 foram reavaliados por (8) utilizando-se 1000 amostras fixas.

Comparando a Fig. 5 com a Fig. 6, percebe-se que o desempenho do AG acelerado empregando apenas 10 amostras móveis é em média superior ao AG desprovido de aceleração utilizando 500 amostras fixas. Mesmo empregando apenas 1 amostra móvel, o AG acelerado consegue resultados ainda melhores do que o AG tradicional para $n_{max} = 100$.

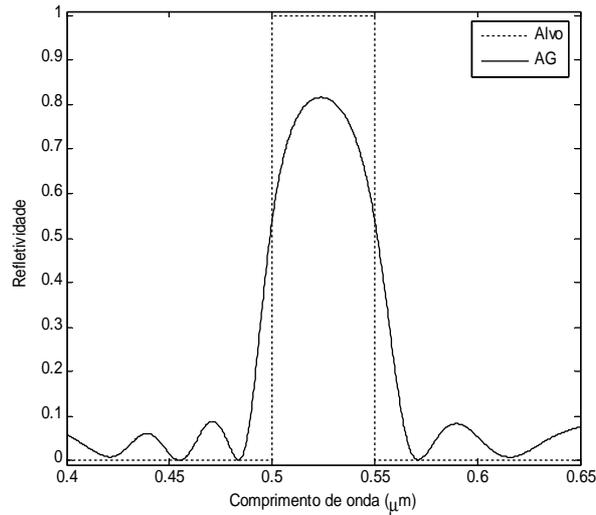


Figura 3 – Refletividade em função do comprimento de onda para o filtro sintetizado pelo AG proposto.

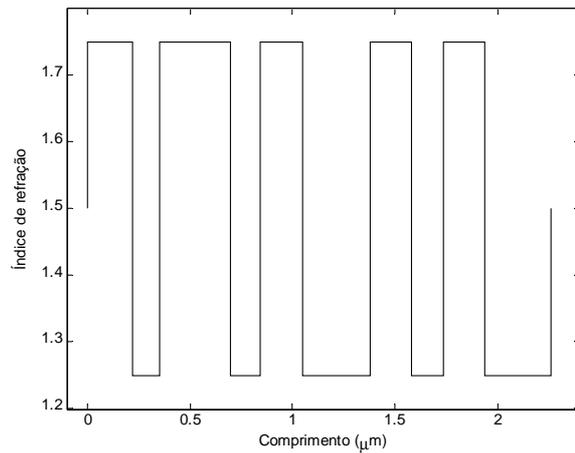


Figura 4 - Perfil de índices de refração para o filtro sintetizado pelo AG proposto.

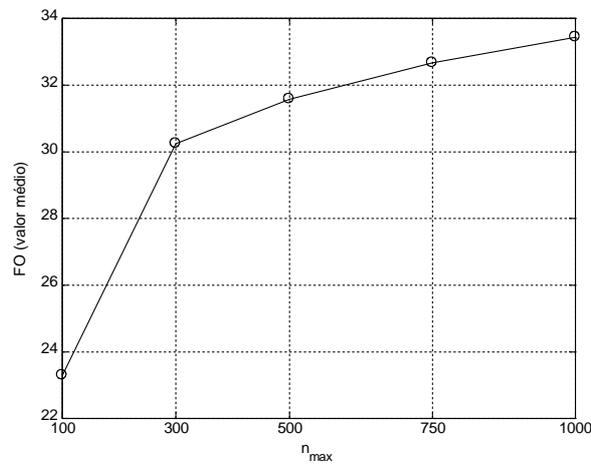


Figura 5 – Valores médios de FO em função do número de amostras obtidos para 10 execuções (por ponto) do AG sem aceleração.

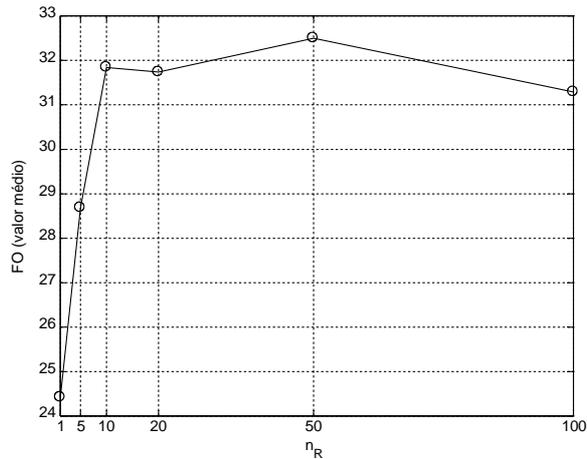


Figura 6 – Valores médios de FO em função do número de amostras obtidos para 10 execuções (por ponto) do AG com aceleração.

As figuras 7 e 8 mostram os tempos de processamento médios por geração relacionados respectivamente às execuções das figuras 5 e 6. O computador utilizado nas experiências foi um PC com processador AMD Athlon 2400+ com 512 MB de memória principal. O *software* do AG foi escrito em C++ para a plataforma GNU/Linux e compilado através do GNU gcc/g++.

A Fig. 9 compara o ganho f médio calculado para o AG acelerado com o ganho previsto através de (7) para $n_E = 1000$ e os mesmos valores de n_R utilizados na Fig. 6. Observa-se que quanto menor o número de amostras n_R , maior o erro entre a curva esperada e a obtida, chegando a quase 29% para o valor de abscissa igual a 1; quando o custo computacional dos procedimentos comuns AG (seleção, operadores genéticos, etc.) tornam-se consideráveis em relação ao custo computacional da função objetivo.

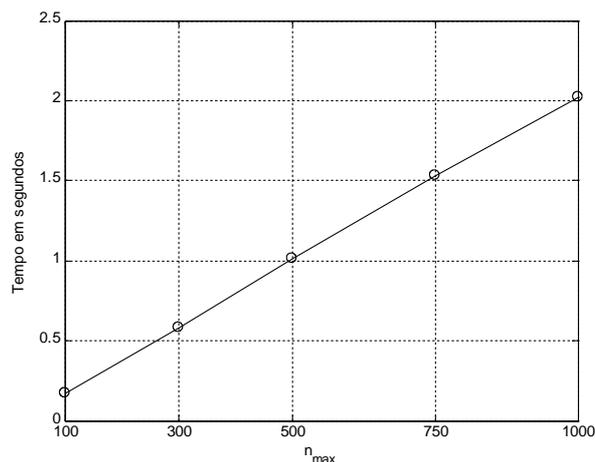


Figura 7 – Tempo de processamento médio por geração (segundos) em função de n_{max} para o AG desprovido de aceleração.

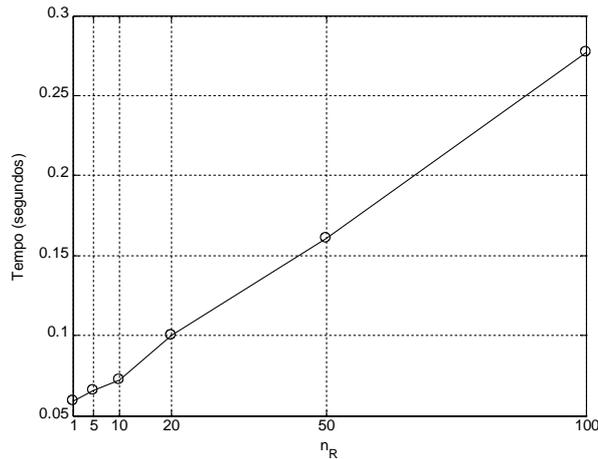


Figura 8 – Tempo de processamento médio por geração (segundos) em função de n_R para o AG acelerado.

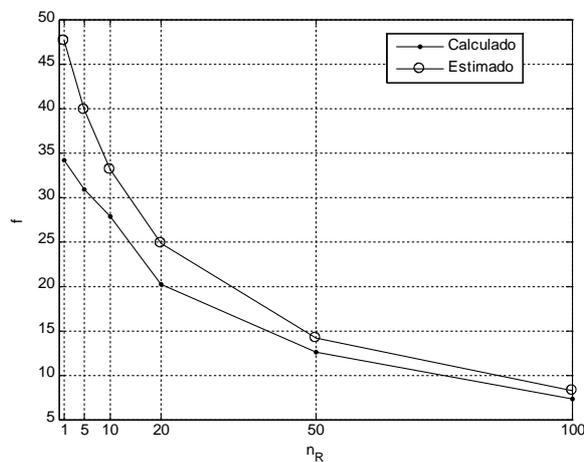


Figure 9 – Comparação entre de f médio calculado (obtido) e estimado para o AG proposto utilizando $n_E = 1000$.

4. CONCLUSÕES

As técnicas de aceleração mostraram-se bastante satisfatórias. Os ganhos computacionais observados chegaram próximo de 3000%, empregando $n_R = 10$ e $n_E = 1000$. Embora a aplicação em particular tenha sido a síntese de filtros ópticos, quaisquer problemas de síntese envolvendo algoritmos evolutivos e empregando funções objetivos que desenvolvam curvas, para serem comparadas ponto a ponto, podem se beneficiar das técnicas aqui mostradas. Em algumas classes de problemas seria também possível aplicar apenas o procedimento de correção elitista, desde que alguma forma de simplificação dos cálculos para a FO esteja disponível. É comum dispor de modelos menos realísticos que, embora possuam aplicação limitada, podem ser corridos com menos custo computacional. Estes modelos podem dar origem a FOR, enquanto os modelos apurados podem constituir a FOE.

REFERÊNCIAS

[1] E. Herbert Li and A. B. Djuri, 2001. Genetic algorithm for reflective filters design. *The Hong Kong Devices Meeting*.

[2] D. E. Goldberg, 1989. Genetic algorithms in search, optimization and machine learning. Addison – Wesley.

[3] Born, M. and Wolf, E., 1980. Principles of optics, 6nd Ed., Pergamon Press.

[4] S. T. Peng, Theodor Tamir, and Henry L. Bertoni 1975. Theory of periodic dielectric waveguides. *Transactions on Microwave Theory and Techniques*, Vol. MTT-23, No. 1.

[5] H.-P. Schwefel and G. Rudolph, 1995. Contemporary evolution strategies. *Advances in Artificial Life, Third ECAL Proceedings*, pages 893-907.

[6] Brad L. Muller, David E. Goldberg, 1995. Genetic algorithms, tournament selection, and the effects of noise. *Urbana, IL: University of Illinois at Urbana-Champaign, IlliGAL Report No. 95006*.

[7] Hsu-Chih Cheng and Yu-Lung Lo, 2005. The Synthesis of Multiple Parameters of Arbitrary FBGs Via a Genetic Algorithm and Two Thermally Modulated Intensity Spectra, *Journal of Lightwave Technology*, Vol. 23, No. 6.