

ANÁLISES DE TRANSIENTE E DE REGIME PERMANENTE DE UM ALGORITMO DE DESCONVOLUÇÃO CEGA

Diego Barreto Haddad

PADS, COPPE/Poli/UFRJ, Rio de Janeiro, Brasil

Coord. Telecomunicações, CEFET-RJ UNED Nova Iguaçu, Rio de Janeiro, Brasil

diego@pads.ufrj.br

Mariane Rembold Petraglia

PADS, COPPE/Poli/UFRJ, Rio de Janeiro, Brasil

mariane@pads.ufrj.br

Paulo Bulkool Batalheiro

PROSAICO, PEL/ DETEL/UERJ, Rio de Janeiro, Brasil

bulkool@pads.ufrj.br

Resumo – O presente artigo empreende um estudo analítico acerca das propriedades de equivariância, transiente e de regime permanente do algoritmo de desconvolução cega DCGN [1], na sua versão monocanal. A tarefa é desafiadora, dada a intensa não-linearidade deste algoritmo. Estimativas teóricas de diferentes graus de sofisticação são obtidas para os valores finais para os quais os coeficientes adaptativos convergem, assim como para o acompanhamento da etapa de transiente. Adicionalmente, é fornecido um critério de seleção dos parâmetros finais do algoritmo retornados por algumas técnicas propostas, porque alguns desses parâmetros não traduzem um comportamento real de uma execução do algoritmo.

Palavras-chave – Desconvolução Cega, Equalização, Análise de Transiente, Interferência Interssmbólica.

Abstract – This paper presents an analytic study about equivariance properties, transient and steady-state of the blind deconvolution algorithm proposed in [1], in its monochannel version. The task is challenging because of the high nonlinearity of this algorithm. Theoretical estimates of distinct sophistication levels are obtained for the final adaptive parameters of the algorithm, as well as for the tracking of the transient phase. In addition, a selection criterion of the steady-state parameters found by the proposed analysis techniques is presented, since some of them do not reflect the behavior of an actual implementation of the algorithm.

Keywords – Blind Deconvolution, Equalization, Transient Analysis, Intersymbol Interference.

1. INTRODUÇÃO

Este artigo se propõe a analisar alguns aspectos de um algoritmo de desconvolução cega de fontes. Mais desafiadoras do que as instantâneas, as configurações convolutivas, dada a ubiquidade do fenômeno dos múltiplos percursos, apresentam um repertório mais amplo de aplicações, justificando o intenso interesse da comunidade acadêmica neste tópico.

Enquanto técnicas de separação cega de fontes são consideradas bem-sucedidas quando as estimativas são versões filtradas das respectivas fontes, o objetivo das técnicas de desconvolução é mais restrito neste aspecto, pois a estas cabe também recuperar a estrutura temporal do sinal, uma restrição relevante em alguns contextos, como o de comunicações digitais¹. Isto justifica o fato de muitas técnicas de desconvolução cega apresentarem aplicações em comunicações de dados, como por exemplo as técnicas de Bussgang [2], que incluem como casos particulares os algoritmos de Sato [3] e de Godard (ou de módulo constante) [4], [5].

Uma outra distinção importante reside no fato de que técnicas de desconvolução cegas podem ser aplicadas a um contexto monocanal, no qual, inexistindo interferências advindas de outras fontes, apenas as interferências entre símbolos da mesma fonte devem ser removidos ou atenuados. Podemos, grosso modo, dividir as técnicas de equalização em dois grandes grupos: as que atuam no espaço de identificação da função de transferência dispersiva (modelo do canal) ou as que efetuam uma otimização no espaço de inversão desta função de transferência. Neste artigo estamos primordialmente interessados nesta segunda estratégia, a qual procura filtros (em geral, do tipo FIR) que, atuando sobre o sinal recebido, desfaçam ou atenuem o efeito do canal.

Em algumas aplicações, como em comunicações digitais, é possível inserir um trecho no sinal original que seja previamente conhecido. Este conhecimento *a priori* possibilita a inserção de técnicas de filtragem adaptativa supervisionada, que constituem um ramo maduro da área de processamento de sinais [6]. Quando não há (ou é indesejável) um acesso a trechos das fontes,

¹Nas misturas instantâneas, técnicas de separação cega também não corrompem a estrutura temporal das fontes.

cumprir recorrer a técnicas de processamento autodidata, as quais são poderosas por também não exigirem um conhecimento acerca da função de transferência do canal.

Embora haja na literatura muitos artigos dedicados à análise de desempenho de algoritmos de processamento supervisionado de sinais, relativamente poucas destas análises contemplam técnicas de processamento cego de separação/desconvolução de fontes. Como exemplos, abordagens analíticas de alguns desses algoritmos podem ser encontradas em [7], [8]; comparações de desempenho via simulações e médias do tipo Monte Carlo são efetuadas em [9], [10]. Uma pesquisa empreendida na base de artigos do IEEE revela poucos artigos de análise de algoritmos de desconvolução cega efetuados nos últimos anos, dos quais destacamos [11], [12], [13], [14]. Nenhum deles, pelo que investigamos, aborda o algoritmo DCGN. Dentre as abordagens analíticas de separação cega de fontes, a maioria trata do caso de misturas instantâneas. Este número reduzido de análises reflete a maior dificuldade de tratamento analítico das técnicas cegas. Duas são as características basicamente responsáveis por essa dificuldade: não-linearidade e impossibilidade de desacoplamento do problema multidimensional em diversos problemas unidimensionais independentes [15]. Estas razões, entre outras, justificam a dificuldade de obtenção de fórmulas analíticas que estimem a ISI (interferência interssimbólica, do inglês *intersymbol interference*) ou o limite superior do fator de aprendizagem, por exemplo. Mesmo na técnica de equalização CMA (*constant modulus algorithm*), mais simples do que o algoritmo analisado neste artigo, encontram-se problemas analíticos difíceis, que requerem por exemplo o recurso à topologia (no caso de sinais de entrada coloridos [16]), à geometria (volume das regiões de convergência [17]), à caracterização da região de dispersão [18], à determinação de classes específicas de canais [19] ou o recurso à análise via preservação de energia [20].

Este artigo contempla a técnica de desconvolução cega proposta em [1]. Embora esta técnica pretenda recuperar exatamente a estrutura temporal do sinal, mesmo quando não é bem-sucedida no tocante à equalização pode, em contextos multicanaís, efetuar uma separação de sinais de voz. Serão investigadas neste artigo questões relacionadas à equivariância, a estimativas dos coeficientes para os quais o algoritmo converge, ao comportamento do mesmo no transiente e a critérios de seleção das soluções que, pertencendo a um conjunto de possibilidades fornecido pela análise teórica, descrevem um comportamento do algoritmo que se verifica na prática.

Cabe ressaltar que algumas das análises deste artigo já foram publicadas em [21], [22], [23]. Porém há modificações na apresentação que efetuamos aqui neste artigo, como estudo de configurações mais complexas na Seção 3.2.2 e na explicitação de fórmulas analíticas num novo caso (Seção 3.5.1). Cabe ressaltar que este artigo apresenta seções inéditas, como as referentes ao método EDO (Seção 3.4) e ao critério de seleção das soluções obtidas pela análise teórica (Seção 3.6), permitindo uma visão unificada do repertório de técnicas de análise por nós empreendidas para este algoritmo.

2. ALGORITMO DCGN

Um dos algoritmos mais eficazes para desconvolução cega de fontes² é apresentado em [1], sendo tratado neste trabalho pela sigla DCGN (de “Desconvolução Cega por meio de Gradiente Natural”). O algoritmo DCGN é uma evolução do algoritmo original apresentado em [24], cujo gradiente natural apresentava um viés que reduzia o desempenho final, porque a iteração de atualização dos parâmetros incluía termos oriundos de amostras de sinais de entrada as quais não influenciavam a função custo. Estes termos adicionais foram introduzidos ao se calcular o “gradiente natural”; uma forma diferente de truncamento da seqüência de entrada elimina estes termos prejudiciais (vide [1] para maiores detalhes).

Como as análises deste artigo se concentram na configuração monocanal deste algoritmo, por motivos de clareza apenas a apresentação da versão monocanal do algoritmo será efetuada, ainda que possa ser compreendida como um caso particular da configuração multicanal.

2.1 VERSÃO MONOCANAL DO ALGORITMO DCGN

Seja \mathbf{h} o filtro de comprimento M correspondente ao canal, definido como:

$$\mathbf{h} = [h(0) \quad h(1) \quad \dots \quad h(M-1)], \quad (1)$$

e $s(k)$ o sinal (ou fonte) a recuperar. Logo, no receptor, o sinal recebido $x(k)$ é uma versão de $s(k)$ corrompida por interferência interssimbólica. Seja \mathbf{w}_k o filtro adaptativo de comprimento L do equalizador na k -ésima iteração, cuja saída $y(k)$, idealmente, deve ser uma estimativa da fonte desprovida do efeito dispersivo do canal. A estrutura do algoritmo de desconvolução é apresentada na Fig. 1. Se definimos o vetor-linha:

$$\mathbf{w}_k = [w_k(0) \quad w_k(1) \quad \dots \quad w_k(L-1)], \quad (2)$$

então a saída $y(k)$ pode ser escrita como:

$$y(k) = \mathbf{w}_k \mathbf{x}_k, \quad (3)$$

com \mathbf{x}_k definido por:

$$\mathbf{x}_k = [x(k) \quad x(k-1) \quad \dots \quad x(k-L+1)]^T. \quad (4)$$

²No contexto de técnicas de filtragem adaptativa não-supervisionada, este trabalho se concentra em aplicações de sinais de áudio e de voz.

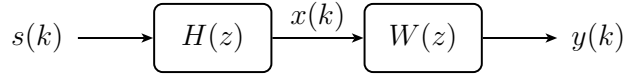


Figura 1: Estrutura de desconvolução que será empregada neste artigo. O filtro $H(z)$ (desconhecido) modela a transformação empreendida pelo canal e $w(z)$ é o filtro equalizador.

A função custo do algoritmo DCGN é descrita por:

$$\mathcal{F}(w(z)) = -\frac{1}{2\pi i} \oint \log|w(z)|z^{-1} dz - E[\log f(y(k))], \quad (5)$$

onde $w(z) = \sum_{l=0}^{L-1} w_k(l)z^{-l}$ e $f(\cdot)$ é a pdf (função densidade de probabilidade) da fonte. Caso tal pdf nos seja desconhecida, empregamos uma estimativa da mesma. Pode-se demonstrar que esta função-custo é, a menos de uma constante independente do equalizador, proporcional à informação mútua da sequência de sinais de saída $\{y(k)\}$, quando $f(\cdot)$ é a pdf da fonte [1]. Esta função-custo pode ser alternativamente colocada como [25]:

$$\mathcal{F}(\mathbf{w}) = -\frac{1}{2\pi} \int_0^{2\pi} \log \left| \sum_{l=0}^{L-1} w(l)e^{-i\theta l} \right| d\theta - E[\log f(y(k))]. \quad (6)$$

A equação de atualização do algoritmo DCGN emprega a otimização via gradiente natural, podendo ser explicitamente escrita por [1]:

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \beta [\mathbf{w}_k - \phi(y(k))\mathbf{z}_k^T], \quad (7)$$

onde $\phi(y(k))$ é a função *score* definida por $\phi(y(k)) = -\frac{d \log f(y(k))}{dy(k)}$, $\mathbf{z}_k = \mathbf{R}_k \mathbf{x}_k$, com \mathbf{R}_k sendo uma matriz simétrica expressa por:

$$\mathbf{R}_k = \begin{bmatrix} r_k(0) & r_k(1) & \cdots & r_k(L-1) \\ r_k(-1) & r_k(0) & \cdots & r_k(L-2) \\ \vdots & \ddots & \ddots & \vdots \\ r_k(-L+1) & r_k(-L+2) & \cdots & r_k(0) \end{bmatrix}, \quad (8)$$

onde $r_k(l) = \sum_{p=0}^{L-1-|l|} w_k(p)w_k(p+|l|)$. [1] sustenta ser difícil justificar teoricamente o desempenho do algoritmo, dada a enorme dificuldade na obtenção de uma análise estatística completa de seu transiente e do comportamento em regime permanente para configurações genéricas. Embora a tarefa seja complexa, neste artigo são empreendidas algumas derivações que elucidam analiticamente algumas características deste algoritmo em determinadas configurações.

3. ANÁLISES DA VERSÃO MONOCANAL DO ALGORITMO DCGN

Para que se possa empreender uma abordagem analítica, importa formular algumas hipóteses a respeito das fontes. Primeiramente, suas amostras são supostas iid (independentes e identicamente distribuídas). Embora, num contexto de separação de fontes, [26] sugira que a hipótese de amostras iid não inutiliza o algoritmo cego quando a fonte apresenta dependência estatística entre amostras (tratando-se apenas de uma opção por não empregar esta dependência), [27] mostra o efeito desta dependência no deslocamento dos mínimos da superfície da função-custo. Como nas simulações da análise as fontes respeitarão esta hipótese, não nos preocuparemos com esta possibilidade.

3.1 EQUIVARIÂNCIA

Um algoritmo de desconvolução ou separação cega de fontes é dito ser equivariante caso seu desempenho e convergência independam do canal. Para o caso instantâneo e (super)determinado, alguns algoritmos apresentam esta desejável propriedade (para um exemplo, vide [28]), excetuando-se os casos degenerados de uma matriz de mistura singular.

Já em configurações convolutivas, muitos algoritmos são considerados equivariantes, embora muitas destas afirmativas sejam menos fundamentadas em sólida base matemática do que em heurísticas extensões do caso instantâneo (por exemplo, vide [29]). O algoritmo DCGN, embora empregue uma otimização via gradiente natural, não é equivariante em sentido estrito, dada a dependência de seu desempenho final com relação ao filtro do canal.

Nesta seção mostraremos que, num caso simples (mas não trivial) e empregando um conceito mais amplo da propriedade de equivariância, podemos concluir que esta é respeitada pelo DCGN. Sendo \mathbf{c}_k o vetor que contém as respostas combinadas do canal e do filtro equalizador, sejam as seguintes hipóteses:

- Hipótese I: $M = L = 2$;
- Hipótese II: $h(0)h(1) \neq 0$, apresentando $h(0)$ e $h(1)$ mesma ordem de magnitude e invariância no tempo;
- Hipótese III: o algoritmo converge para uma solução sem atraso e de baixa ISI, ou seja, $\lim_{k \rightarrow \infty} |c_k(0)| \gg \lim_{k \rightarrow \infty} |c_k(d)|$ para $d \neq 0$.
- Hipótese IV: o fator de aprendizagem β é reduzido o suficiente para que os termos $c_k(i)$ sejam próximos o suficiente de seus valores esperados e permitindo aproximar os coeficientes adaptativos como independentes do sinal de entrada, após a convergência.

Seja $\mathbf{c}_k = [c_k(0) \ c_k(1) \ c_k(2)] = [h(0)w_k(0) \ h(0)w_k(1) + h(1)w_k(0) \ h(1)w_k(1)]$. Admitida a primeira hipótese, podemos desmembrar a equação de atualização do vetor \mathbf{w} em duas equações:

$$\begin{aligned} w_{k+1}(0) &= w_k(0) + \beta w_k(0) - \beta \phi(y(k)) w_k^2(0) x(k) - \beta \phi(y(k)) w_k^2(1) x(k) \\ &\quad - \beta \phi(y(k)) w_k(0) w_k(1) x(k-1), \\ w_{k+1}(1) &= w_k(1) + \beta w_k(1) - \beta \phi(y(k)) w_k(0) w_k(1) x(k) - \beta \phi(y(k)) w_k^2(0) x(k-1) \\ &\quad - \beta \phi(y(k)) w_k^2(1) x(k-1), \end{aligned} \quad (9)$$

A Hipótese III nos permite aplicar o limite $k \rightarrow \infty$ de modo a, após algumas manipulações matemáticas, obtermos:

$$\begin{aligned} \Delta c(0) &= \beta c(0) - \beta \frac{c^2(0)}{h(0)} \phi(y(k)) x(k) - \beta \frac{h(0)c^2(2)}{h^2(1)} \phi(y(k)) x(k) \\ &\quad - \beta \frac{c(0)c(1)}{h(1)} \phi(y(k)) x(k-1), \\ \Delta c(2) &= \beta c(2) - \beta \frac{c(0)c(2)}{h(0)} \phi(y(k)) x(k) - \beta \frac{h(1)c^2(0)}{h^2(0)} \phi(y(k)) x(k) \\ &\quad - \beta \frac{c^2(2)}{h(1)} \phi(y(k)) x(k-1), \end{aligned} \quad (10)$$

onde elidimos os índices k nos termos $c_k(i)$ e $\Delta c(i) = c_{k+1}(i) - c_k(i)$. A partir da Hipótese IV associamos as equações acima às equações determinísticas abaixo:

$$\begin{aligned} c(0) &= \text{E} \left[\phi(y(k)) \left\{ \left[\frac{c^2(0)}{h(0)} + \frac{h(0)}{h^2(1)} c^2(2) \right] x(k) + \frac{1}{h(1)} c(0)c(2)x(k-1) \right\} \right], \\ c(2) &= \text{E} \left[\phi(y(k)) \left\{ \frac{1}{h(0)} c(0)c(2)x(k) + \left[\frac{h(1)}{h^2(0)} c^2(0) + \frac{1}{h(1)} c^2(2) \right] x(k-1) \right\} \right]. \end{aligned} \quad (11)$$

Cumpra agora recorrer a algumas simplificações. Em nossas simulações, caso inicializemos \mathbf{w} como $[1 \ 0]$ e $h(0)$ seja maior do que $h(1)$ (o que é uma situação comum, já que as versões atrasadas do sinal tendem a estar mais atenuadas), $c(0)$ costuma ser muito maior do que $c(1)$ e $c(2)$ (admitindo convergência do algoritmo para uma boa solução). Supondo que isso ocorra, $y(k)$ deve ser muito próximo (a menos da constante multiplicativa $c(0)$) de $s(k)$ ou, em outras palavras, o sistema canal + equalizador não introduz atrasos, o que justifica, de certo modo, as hipóteses admitidas. Cabe no entanto ressaltar a forte restrição implicada pela hipótese III, mesmo para canais de fase mínima. Na prática, costuma haver um atraso não nulo na equalização, de sorte que a análise efetuada nesta seção se limita apenas a uma classe de canais.

Sendo iid as amostras de $s(k)$, $y(k)$ deve ser, aproximadamente, independente de $s(k-1)$ e de $s(k-2)$, o que nos permite efetuar a seguinte aproximação:

$$\text{E} [\phi(y(k))s(k-i)] \approx 0, \text{ para } i = 1, 2. \quad (12)$$

Como $x(k-i) = h(0)s(k-i) + h(1)s(k-i-1)$, chegamos a:

$$\begin{aligned} c(0) &\approx c^2(0)\text{E} [\phi(y(k))s(k)] + \frac{h^2(0)}{h^2(1)} c^2(2)\text{E} [\phi(y(k))s(k)] \\ &= \left[c^2(0) + \frac{h^2(0)}{h^2(1)} c^2(2) \right] \text{E} [\phi(y(k))s(k)]. \end{aligned} \quad (13)$$

Pelas Hipóteses II e III, podemos desprezar o termo $\frac{h^2(0)}{h^2(1)} c^2(2)$. Assim, chegamos a:

$$c(0) \approx c^2(0)\text{E} [\phi(y(k))s(k)], \quad (14)$$

de onde obtemos:

$$c(0) \approx \pm \frac{1}{\text{E} [\phi(y(k))s(k)]}, \quad (15)$$

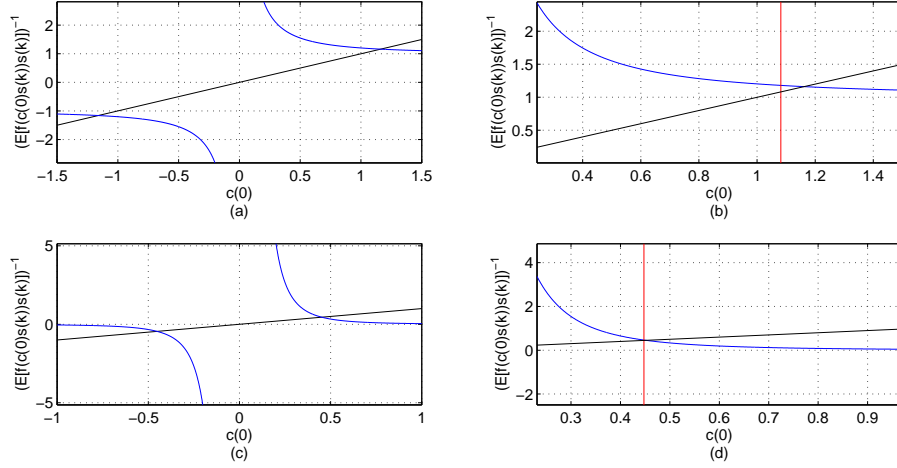


Figura 2: Avaliação da estimativa da Eq. (16), com inicialização $\mathbf{w}_0 = [1 \ 0]$, $\beta = 10^{-5}$ e 5×10^5 iterações. O sinal $s(k)$ foi gerado a partir da distribuição $f(s) = \frac{1}{2}e^{-|s|}$. A reta identidade está em preto e o termo $\frac{1}{E[f(c(0)s(k))s(k)]}$ (calculado via médias de Monte Carlo) em azul. Em vermelho, assinala-se o valor real para o qual o algoritmo converge (calculado por meio da média das últimas 3×10^5 iterações). (a) função $\phi(y(k)) = \tanh(y(k))$ e filtro do canal $\mathbf{h} = [1 \ 0 \ 3]$; (b) ampliação de parte do gráfico em (a); (c) função $\phi(y(k)) = y^3(k)$ e filtro do canal $\mathbf{h} = [-1, 5 \ 0, 4]$ (d) ampliação de parte do gráfico de (c).

onde a ambiguidade de escalamento responde pelo sinal \pm . A fórmula acima não é suficiente para a determinação de $c(0)$, já que $y(k)$ também depende de $c(1)$ e $c(2)$. Porém, a Hipótese III permite-nos supor que o equalizador é bem-sucedido, de modo que podemos empregar a aproximação $y(k) \approx c(0)s(k)$ e então obtemos o seguinte resultado:

$$c(0) \approx \frac{1}{E[\phi(c(0)s(k))s(k)]}. \quad (16)$$

Supondo conhecidas as estatísticas das fontes (o que é possível quando, por exemplo, é conhecida a modulação de um sinal transmitido) e a função ϕ , o termo $E[\phi(c(0)s(k))s(k)]$ pode ser calculado para qualquer valor de $c(0)$, seja por meio de resultados analíticos, seja, na impossibilidade destes serem efetuados, via médias de Monte Carlo.

O resultado expresso na Eq. (16) significa que o valor para o qual o coeficiente adaptativo de maior magnitude $c_k(0)$ converge (condição $k \rightarrow \infty$) praticamente independe da resposta do canal. Esta robustez é uma característica desejável e associada à equi-variância. Além deste resultado não ser facilmente generalizável para outras configurações, para o caso em questão verificamos que esta robustez não ocorre com $c_k(1)$ e $c_k(2)$, o que significa que estamos longe de elucidar exatamente em que sentido o algoritmo pode ser associado à propriedade de equi-variância. De todo modo, a Eq. (16) revela que, de alguma forma, podemos interpretar a técnica DCGN como atendendo à equi-variância na configuração sem atraso. É claro que

A Fig. 2 permite uma comparação da estimativa oriunda da Eq. (16) com a obtida pelo algoritmo DCGN. Sendo perfeita a estimativa, a reta vertical vermelha deve coincidir com o ponto de interseção da curva em azul e da reta. Para os dois casos apresentados nesta figura, a Eq. (16) revelou boa concordância com os resultados experimentais. A evolução da ISI destes casos é calculada segundo a fórmula [1]:

$$ISI = \frac{\sum_{l=0}^{M+L-1} c_k^2(l)}{\max_l c_k^2(l)} - 1, \quad (17)$$

e apresentada na Fig. 3.

3.2 ESTIMATIVA DE PRIMEIRA ORDEM DOS PARÂMETROS FINAIS DO ALGORITMO DCGN

Estimar o comportamento em regime permanente é uma tarefa complexa para o algoritmo DCGN, dada a não linearidade de sua equação de atualização. Nesta subseção apresentamos uma técnica que emprega a linearização da função $\phi(\cdot)$. A análise a ser efetuada associa uma determinada configuração a um sistema não linear de equações, cuja solução resulta em estimativas dos parâmetros $w_k(i)$ para os quais o algoritmo converge. Sejam as hipóteses:

- Hipótese I: O algoritmo converge para uma solução de bom desempenho (ou, equivalentemente, baixa ISI);
- Hipótese II: A fonte apresenta amostras i.i.d., de média zero, variância σ_s^2 e distribuição simétrica;
- Hipótese III: o fator de aprendizagem é suficientemente reduzido de modo a podermos substituir, em regime permanente, os parâmetros $w_k(i)$ por seus valores esperados, supondo-os também independentes do sinal de entrada;

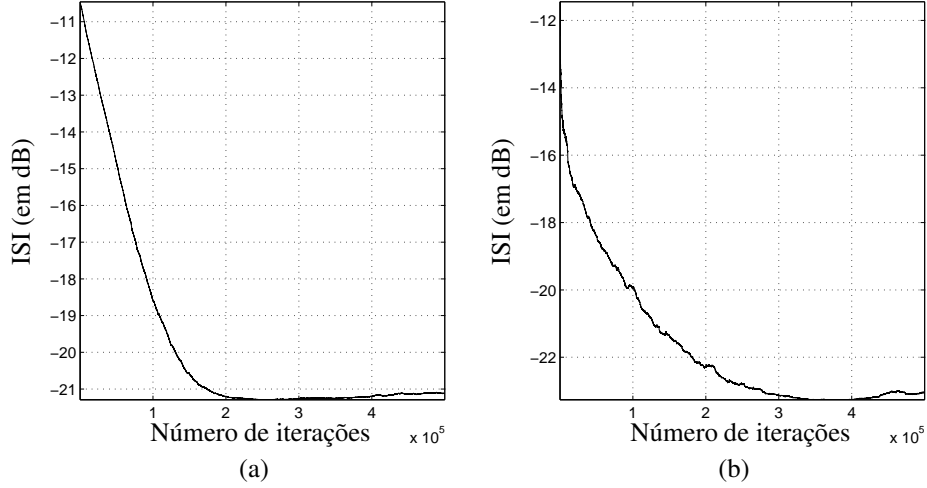


Figura 3: Evolução da ISI nos testes efetuados na Fig. 2, com inicialização $\mathbf{w}_0 = [1 \ 0]$ e $\beta = 10^{-5}$. (a) função $\phi(y(k)) = \tanh(y(k))$ e filtro do canal $\mathbf{h} = [1 \ 0, 3]$; (b) função $\phi(y(k)) = y^3(k)$ e filtro do canal $\mathbf{h} = [-1, 5 \ 0, 4]$.

Por simplicidade, efetuaremos a estimativa de primeira ordem (EPO) para o caso $M = L = 2$ e atraso final unitário (ou seja, $y(k) \approx c(1)s(k-1)$), para não tornar as equações demasiado grandes; não obstante, ao contrário da análise de equivariância, a EPO não se restringe a este caso, como poderá ser observado nas simulações. Uma desvantagem da EPO é a exigência de se arbitrar um atraso final do algoritmo, em geral desconhecido *a priori*.

Importa enfatizar o relaxamento das hipóteses com relação à análise de equivariância efetuada anteriormente. Impondo a condição $k \rightarrow \infty$ e empregando as Hipóteses I e III, temos:

$$\begin{aligned} 0 &\approx w(0) - (w^2(0) + w^2(1))E[\phi(y(k))x(k)] - w(0)w(1)E[\phi(y(k))x(k-1)], \\ 0 &\approx w(1) - (w^2(0) + w^2(1))E[\phi(y(k))x(k-1)] - w(0)w(1)E[\phi(y(k))x(k)], \end{aligned} \quad (18)$$

onde a elisão do índice k revela a condição de regime permanente. A principal dificuldade na resolução das equações acima reside nos termos não lineares $E[\phi(y(k))x(k)]$ e $E[\phi(y(k))x(k-1)]$, os quais serão, na EPO proposta, linearizados.

Sendo unitário o atraso final e empregando a Hipótese I, temos $y(k) = c(0)s(k) + c(1)s(k-1) + c(2)s(k-2) \approx c(1)s(k-1)$, o que nos permite utilizar uma linearização de Taylor do termo $\phi(y(k))$:

$$\phi(y(k)) \approx \phi(c(1)s(k-1)) + c(0)\phi'(c(1)s(k-1))s(k) + c(2)\phi'(c(1)s(k-1))s(k-2), \quad (19)$$

onde $\phi'(\cdot)$ é a derivada de $\phi(\cdot)$.

Empregando as Hipóteses II e III, o termo $E[\phi(y(k))x(k)]$ pode ser expresso por:

$$E[\phi(y(k))x(k)] = h(1)\overbrace{E[\phi(c(1)s(k-1))s(k-1)]}^{\chi_1(c(1))} + h(0)c(0)\overbrace{E[\phi'(c(1)s(k-1))]}^{\chi_2(c(1))}\overbrace{E[s^2(k)]}^{\sigma_s^2}. \quad (20)$$

De forma semelhante, chegamos a:

$$E[\phi(y(k))x(k-1)] = h(0)\chi_1(c(1)) + h(1)c(2)\chi_2(c(1))\sigma_s^2. \quad (21)$$

Importa ressaltar o fato de que $\chi_1(c(1))$ e $\chi_2(c(1))$ são dependentes das estatísticas da fonte $s(k)$. Estas aproximações nos permitem reescrever as Eqs. (18) da forma a seguir:

$$\begin{aligned} 0 &\approx w(0) - (h(1)w^2(0) + h(1)w^2(1) + h(0)w(0)w(1))\chi_1(c(1)) \\ &\quad - (h^2(0)w^3(0) + h^2(0)w(0)w^2(1) + h^2(1)w(0)w^2(1))\chi_2(c(1))\sigma_s^2, \\ 0 &\approx w(1) - (h(0)w^2(0) + h(0)w^2(1) + h(1)w(0)w(1))\chi_1(c(1)) \\ &\quad - (h^2(1)w^2(0)w(1) + h^2(1)w^3(1) + h^2(0)w^2(0)w(1))\chi_2(c(1))\sigma_s^2, \end{aligned} \quad (22)$$

onde utilizamos o fato de que $c(0) = h(0)w(0)$ e $c(2) = h(1)w(1)$ (por simplicidade, optamos por não expressar $c(1)$ como $h(0)w(1) + h(1)w(0)$).

Encontrar as soluções do sistema de equações acima implica estimar os parâmetros $w(0)$ e $w(1)$, o que nos permitiria, sem recursos a simulações ou a médias do tipo Monte Carlo, obter teoricamente o desempenho do algoritmo numa determinada configuração (ou seja, para determinados valores de $h(0)$ e $h(1)$). O sistema de equações é não linear, o que dificulta uma análise teórica genérica.

3.2.1 CASO PARTICULAR: DISTRIBUIÇÕES EXPONENCIAIS

Em geral, o sistema de equações (22) não apresenta soluções analíticas. Por isso, investigaremos um caso particular de $\phi(\cdot)$ e da pdf da fonte $s(k)$, que permite a determinação destas soluções. Esta análise também servirá para explicitar a metodologia do cálculo de $\chi_1(c(1))$ e de $\chi_2(c(1))$.

Suponhamos que a distribuição de $s(k)$ seja exponencial, de média zero e de parâmetro $\kappa > 0$:

$$f(s) = \frac{1}{2\kappa} e^{-\frac{|s|}{\kappa}}. \quad (23)$$

Escolhendo a função *score* ϕ de forma ótima, encontramos:

$$\phi(s) = -\frac{d \log\left(\frac{1}{2\kappa}\right)}{ds} + \frac{d \frac{|s|}{\kappa}}{ds} = \frac{\text{sign}(s)}{\kappa}. \quad (24)$$

Neste caso particular, é possível obter analiticamente o termo $\chi_1(c(1))$:

$$\begin{aligned} \chi_1(c(1)) &= E[s(k-1)\phi(c(1)s(k-1))] = \int_{-\infty}^{\infty} s \cdot \frac{\text{sign}(c(1)s)}{\kappa} \frac{1}{2\kappa} e^{-\frac{|s|}{\kappa}} \\ &= \frac{1}{2\kappa^2} \int_{-\infty}^{\infty} s \cdot \text{sign}(c(1)s) e^{-\frac{|s|}{\kappa}} ds, \end{aligned} \quad (25)$$

onde a modelagem das amostras de $s(k)$ como iid permite a elisão do índice k . Para $c(1) > 0$ temos:

$$\chi_1(c(1)) = \frac{1}{2\kappa^2} \left[-\int_{-\infty}^0 s e^{\frac{s}{\kappa}} ds + \int_0^{\infty} s e^{-\frac{s}{\kappa}} ds \right] = 1. \quad (26)$$

Para o caso $c(1) < 0$, temos que $\chi_1(c(1)) = -1$; daí:

$$\chi_1(c(1)) = \text{sign}(c(1)). \quad (27)$$

Sendo $\phi(s) = \frac{\text{sign}(s)}{\kappa}$, temos que $\phi'(s) = 0$. Isso implica $\chi_2(c(1)) = 0$ (exceto na descontinuidade presente em $s = 0$; porém esta condição possui probabilidade nula de ocorrência, o que nos permite desconsiderá-la). Daí, o sistema de equações (22) pode ser simplificado³:

$$\begin{aligned} h(1)w^2(0) + (h(0)w(1) - 1)w(0) + h(1)w^2(1) &\approx 0, \\ h(0)w^2(1) + (h(1)w(0) - 1)w(1) + h(0)w^2(0) &\approx 0, \end{aligned} \quad (28)$$

A partir das equações abaixo:

$$w(0) = \frac{1 - h(0)w(1) \pm \sqrt{1 + h^2(0)w^2(1) - 2h(0)w(1) - 4h^2(1)w^2(1)}}{2h(1)}, \quad (29)$$

$$w(1) = \frac{1 - h(1)w(0) \pm \sqrt{1 + h^2(1)w^2(0) - 2h(1)w(0) - 4h^2(0)w^2(0)}}{2h(0)}, \quad (30)$$

é possível explicitar de forma analítica a solução deste sistema de equações, por meio de substituição. O resultado final foi omitido devido à sua grande extensão.

Obviamente, a solução trivial $w(0) = w(1) = 0$ deve ser excluída, bem como eventuais soluções complexas, que podem surgir para valores maiores de M e L .

O sistema de equações (29)-(30) apresenta uma solução para um caso particular que ilustra uma desejável propriedade do algoritmo. Seja o caso $h(0) = 0$ e $h(1) \neq 0$, na qual o canal apenas introduz um atraso unitário e um fator de escalamento $h(1)$ no sinal. Nesta configuração, manter o atraso unitário na saída do equalizador sem alterar a fonte significa termos $w(0) = 1/h(1)$ e $w(1) = 0$. Admitindo $w(1) = 0$, a Eq. (29) (descartando a solução trivial) fornece-nos justamente o valor $w(0) = 1/h(1)$, conforme esperado. O mesmo comportamento se observa caso façamos $h(0) \neq 0$ e $h(1) = 0$, quando o valor ideal do vetor \mathbf{w} seria $[0 \ 1/h(0)]$ (supondo atraso unitário). Como esta característica exige uma condição muito particular do canal (ou seja, que o mesmo apresente apenas um componente não nulo), não podemos associá-la à equivariância.

3.2.2 EXPERIMENTOS

Seja o caso $M = L = 3$. Como ilustração, apresentamos duas das três equações a ele associadas (com as configurações da Fig. 4):

$$h(2)w^2(0) + (h(1)w(1) + h(0)w(2) - 1)w(0) + h(2)w^2(1) + h(2)w^2(2) + h(1)w(1)w(2) = 0.$$

³Supomos aqui o caso $c(1) > 0$; para analisar a possibilidade $c(1) < 0$, basta inverter o sinal dos parâmetros $w(0)$ e $w(1)$.

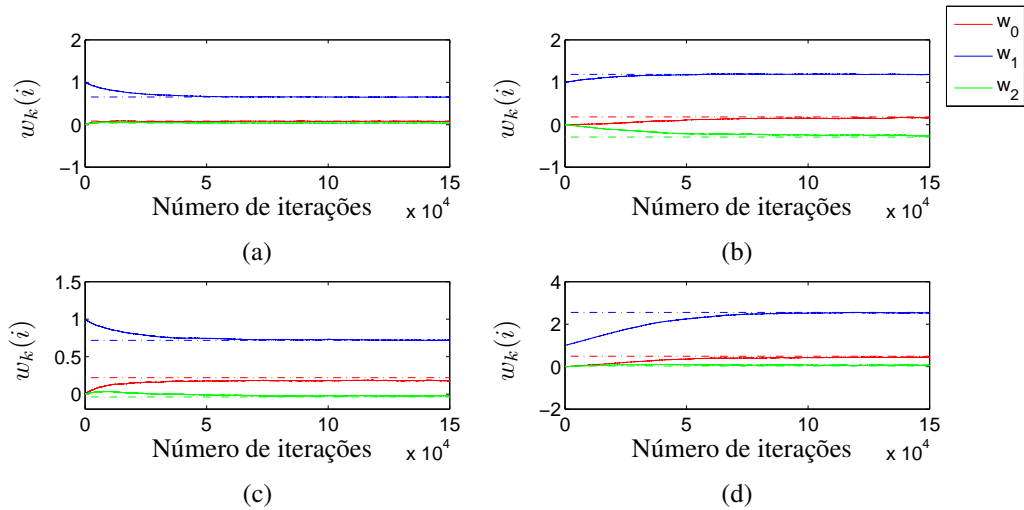


Figura 4: Evolução dos parâmetros $w_k(0)$, $w_k(1)$ e $w_k(2)$ para os quatro casos do experimento com $M = L = 3$. Os parâmetros utilizados foram $\beta = 5 \times 10^{-5}$, $f(s)$ exponencial com $\kappa = 0, 1$, $\phi(y(k)) = \frac{\text{sign}(y(k))}{\kappa}$ e vetor \mathbf{w} inicializado como $[0 \ 1 \ 0]$. As linhas horizontais refletem o valor teórico. (a) filtro $\mathbf{h}^{(1)}$; (b) filtro $\mathbf{h}^{(2)}$; (c) filtro $\mathbf{h}^{(3)}$ e (d) filtro $\mathbf{h}^{(4)}$.

$$h(0)w^2(2) + (h(2)w(0) + h(1)w(1) - 1)w(2) + h(1)w(0)w(1) + h(0)w^2(0) + h(0)w^2(1) = 0. \quad (31)$$

A resolução de um sistema de equações não lineares apresenta diversas dificuldades. Embora técnicas padrão de resolução destas equações (empregadas na instrução `solve` do MATLAB) gerem as soluções desejadas, existem também outras soluções que não refletem qualquer comportamento real do algoritmo (soluções espúrias). Esta dificuldade levanta a questão sobre como selecionar as soluções que fazem sentido. Mais à frente tais questões serão objeto de maior atenção, numa modelagem mais sofisticada (que não emprega a aproximação EPO). No momento, foram selecionadas as soluções do sistema que mais se aproximam do comportamento real do algoritmo.

Sejam as quatro possíveis escolhas para o filtro do canal com $M = 3$ abaixo apresentadas:

$$\begin{aligned} \mathbf{h}^{(1)} &= [0, 19687 \quad -1, 5556 \quad 0, 087869], \\ \mathbf{h}^{(2)} &= [0, 11839 \quad -0, 78388 \quad -0, 18287], \\ \mathbf{h}^{(3)} &= [-0, 37657 \quad 1, 3414 \quad 0, 073948], \\ \mathbf{h}^{(4)} &= [-0, 072526 \quad 0, 39283 \quad -0, 0047669], \end{aligned}$$

A Fig. 4 apresenta a evolução dos coeficientes e sua comparação com a solução obtida no sistema de equações (31), mostrando que a modelagem se revelou razoavelmente acurada (a análise supôs um atraso de duas amostras na saída do equalizador).

A EPO também permite encontrar soluções de configurações de ordem mais elevada. Seja o caso $M = 7$ com as seguintes escolhas para o filtro do canal:

$$\begin{aligned} \mathbf{h}^{(1)} &= [0, 0235 \quad 0, 3727 \quad 0, 3812 \quad 2, 5901 \quad -0, 6070 \quad -0, 0815 \quad -0, 1358], \\ \mathbf{h}^{(2)} &= [-0, 2830 \quad -0, 1560 \quad 0, 2387 \quad 1, 4773 \quad -0, 0941 \quad -0, 1025 \quad -0, 0717], \\ \mathbf{h}^{(3)} &= [-0, 1841 \quad -0, 0863 \quad 0, 0870 \quad 1, 4511 \quad 0, 1498 \quad -0, 2256 \quad 0, 1428], \\ \mathbf{h}^{(4)} &= [0, 1092 \quad 0, 2614 \quad -0, 1463 \quad -1, 2872 \quad -0, 4135 \quad -0, 2158 \quad -0, 0624]. \end{aligned}$$

Modelando o atraso como de 6 amostras, as soluções da EPO são comparadas com os resultados finais do algoritmo nas Figuras 5-8, as quais demonstram a acurácia da estimativa gerada pela EPO. A evolução da ISI nestes casos também é apresentada, mostrando que a Hipótese I é satisfeita em todos eles.

3.2.3 CRÍTICA DA EPO

Ainda que permita uma estimativa dos valores para os quais o equalizador converge, a EPO proposta apresenta duas características indesejáveis: exige a convergência do algoritmo para uma baixa ISI (o que nem sempre ocorre) para efetuar a linearização e necessita de um conhecimento *a priori* do atraso final do sistema canal + equalizador, o qual depende tanto da inicialização do filtro quanto do canal (em geral, desconhecido).

A propriedade de gerar diversas soluções não é tão problemática, porque em geral, em nossas simulações, a escolha da solução que gera a menor ISI coincide com a solução desejada. E, além disso, mais à frente uma técnica de escolha cega da

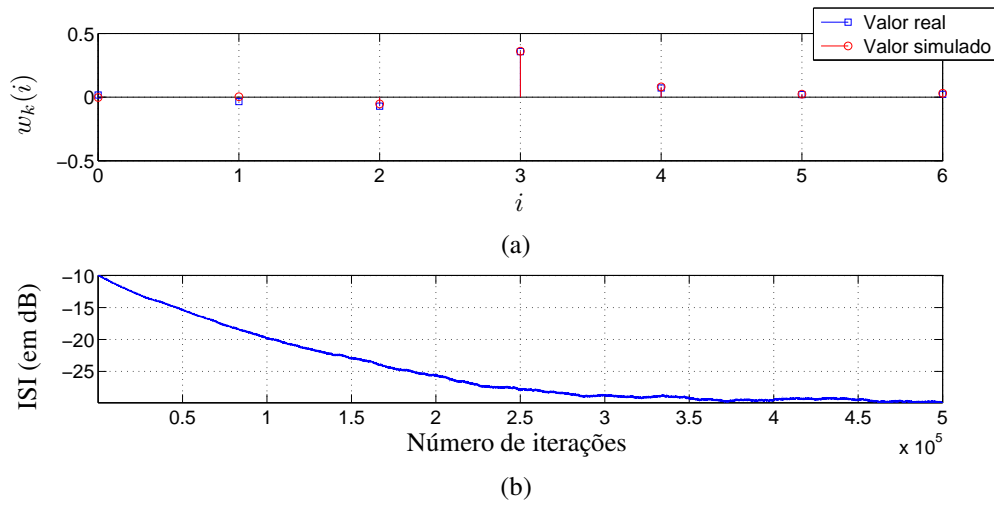


Figura 5: Análise da técnica EPO aplicada ao canal $\mathbf{h}^{(1)}$ do experimento com $M = L = 7$. Os parâmetros utilizados foram $\beta = 5 \times 10^{-5}$, $f(s)$ exponencial com $\kappa = 0,1$, $\phi(y(k)) = \frac{\text{sign}(y(k))}{\kappa}$ e vetor \mathbf{w} inicializado como $[0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0]$. (a) comparação dos valores reais com os estimados via EPO; (b) evolução da ISI (em dB).

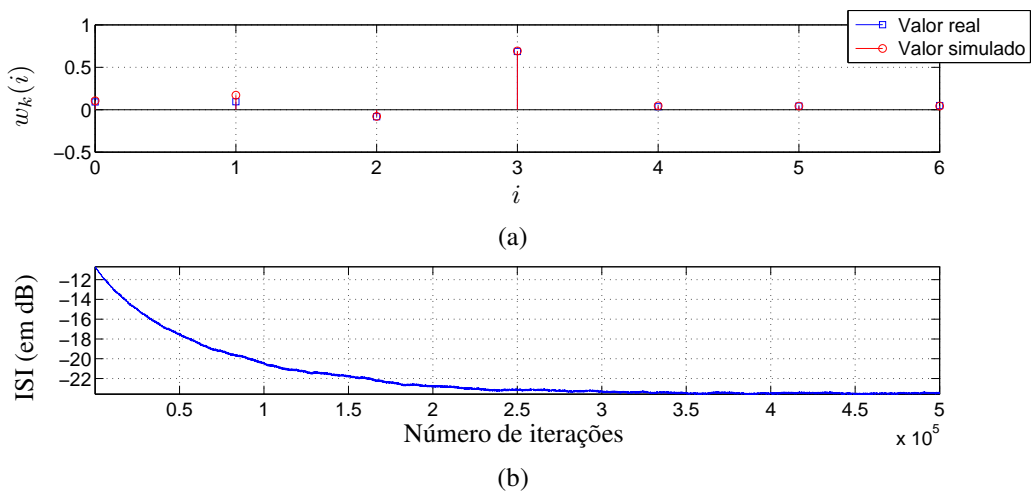


Figura 6: Análise da técnica EPO aplicada ao canal $\mathbf{h}^{(2)}$ do experimento com $M = L = 7$. Os parâmetros utilizados foram $\beta = 5 \times 10^{-5}$, $f(s)$ exponencial com $\kappa = 0,1$, $\phi(y(k)) = \frac{\text{sign}(y(k))}{\kappa}$ e vetor \mathbf{w} inicializado como $[0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0]$. (a) comparação dos valores reais com os estimados via EPO; (b) evolução da ISI (em dB).

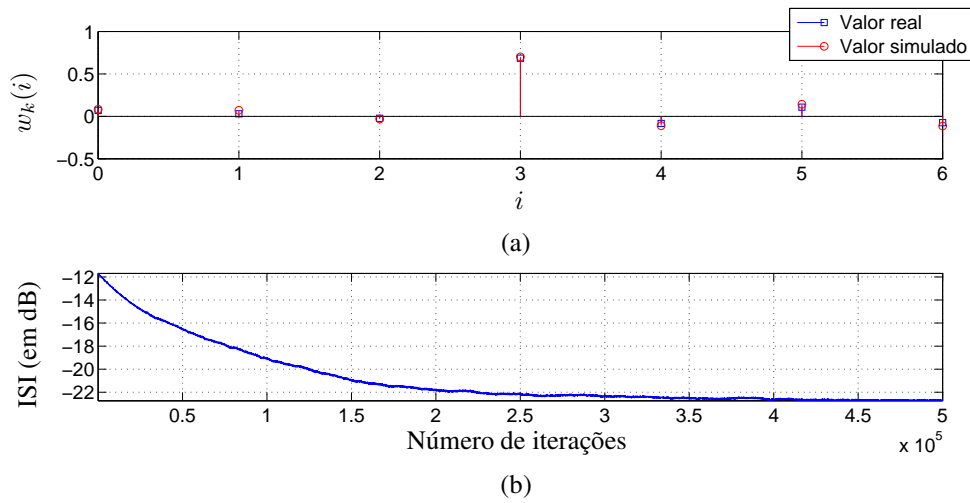


Figura 7: Análise da técnica EPO aplicada ao canal $\mathbf{h}^{(3)}$ do experimento com $M = L = 7$. Os parâmetros utilizados foram $\beta = 5 \times 10^{-5}$, $f(s)$ exponencial com $\kappa = 0,1$, $\phi(y(k)) = \frac{\text{sign}(y(k))}{\kappa}$ e vetor \mathbf{w} inicializado como $[0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0]$. (a) comparação dos valores reais com os estimados via EPO; (b) evolução da ISI (em dB).

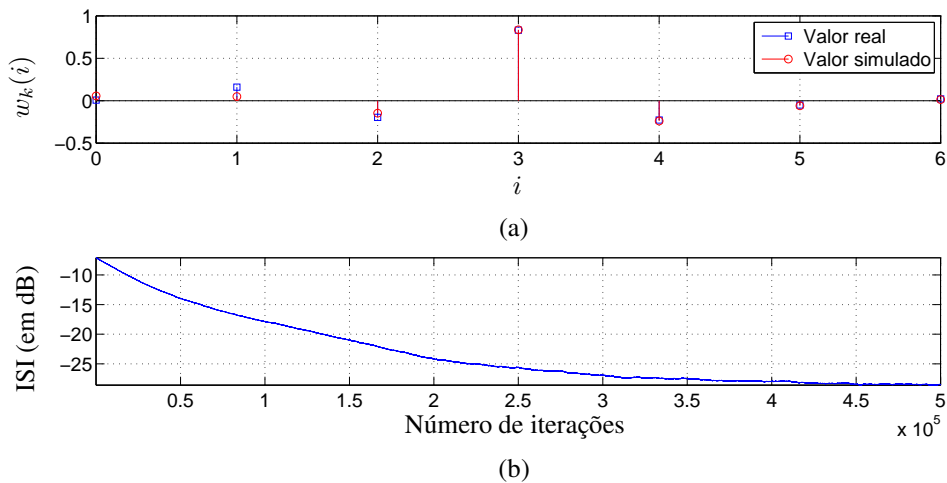


Figura 8: Análise da técnica EPO aplicada ao canal $\mathbf{h}^{(4)}$ do experimento com $M = L = 7$. Os parâmetros utilizados foram $\beta = 5 \times 10^{-5}$, $f(s)$ exponencial com $\kappa = 0,1$, $\phi(y(k)) = \frac{\text{sign}(y(k))}{\kappa}$ e vetor \mathbf{w} inicializado como $[0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0]$. (a) comparação dos valores reais com os estimados via EPO; (b) evolução da ISI (em dB).

solução correta será efetuada.

Para contornar estas características indesejáveis, elaboramos uma técnica mais sofisticada, intitulada de EAO (estimativa de alta ordem). Como transição entre a EPO e a EAO, apresentamos uma outra técnica proposta que se destina a rastrear (por meio de equações recursivas) o comportamento do transiente do algoritmo DCGN. Tanto a análise de transiente quanto a EAO não exigem que a solução final para a qual o algoritmo converge apresente baixa interferência interssimbólica.

3.3 ANÁLISE DE TRANSIENTE DO ALGORITMO DCGN

A eliminação da hipótese de bom desempenho final do algoritmo permite-nos, em alguns casos específicos, não empregar a aproximação da Eq. (19). A análise de transiente em questão recorre a apenas duas hipóteses:

- Hipótese IV: O valor de β é suficientemente pequeno para permitir desprezar a variância de $w_k(i)$, de modo que podemos interpretar a evolução dos coeficientes adaptativos como determinística;
- Hipótese V: os filtros $w_k(i)$ independem de \mathbf{x}_k .

A Hipótese IV é mais forte do que as comumente empregadas em análise de algoritmos de filtragem adaptativa supervisionada. Tal se deve à intensa não-linearidade das equações de atualização, a qual, para ser contornada, exige o emprego de suposições fortes, cuja justificativa se dará por meio da grande concordância observada com os resultados experimentais.

Por simplicidade⁴, seja o caso $M = L = 2$, cujas equações de atualização encontram-se em (9). Interessados na configuração de transiente, não utilizaremos a condição $\lim_{k \rightarrow \infty}$. Aplicando o valor esperado nas equações de atualização supracitadas, temos:

$$\begin{aligned} E[w_{k+1}(0)] &= E[w_k(0)] + \beta E[w_k(0)] - \beta E[w_k^2(0)\phi(y(k))x(k)] - \beta E[w_k^2(1)\phi(y(k))x(k)] \\ &\quad - \beta E[w_k(0)w_k(1)\phi(y(k))x(k-1)], \\ E[w_{k+1}(1)] &= E[w_k(1)] + \beta E[w_k(1)] - \beta E[w_k(0)w_k(1)\phi(y(k))x(k)] \\ &\quad - \beta E[w_k^2(0)\phi(y(k))x(k-1)] - \beta E[w_k^2(1)\phi(y(k))x(k-1)]. \end{aligned} \quad (32)$$

As Hipóteses IV e V podem ser empregadas para transformar as equações acima em:

$$\begin{aligned} w_{k+1}(0) &= w_k(0) + \beta w_k(0) - \beta w_k^2(0)E[\phi(y(k))x(k)] - \beta w_k^2(1)E[\phi(y(k))x(k)] \\ &\quad - \beta w_k(0)w_k(1)E[\phi(y(k))x(k-1)], \\ w_{k+1}(1) &= w_k(1) + \beta w_k(1) - \beta w_k(0)w_k(1)E[\phi(y(k))x(k)] \\ &\quad - \beta w_k^2(0)E[\phi(y(k))x(k-1)] - \beta w_k^2(1)E[\phi(y(k))x(k-1)]. \end{aligned} \quad (33)$$

Na EPO, os termos $E[\phi(y(k))x(k-i)]$ são calculados por meio da linearização de $\phi(\cdot)$, supondo que $k \rightarrow \infty$ e que o algoritmo tende para uma boa solução. Como tais hipóteses não se adequam à análise de transiente, pode-se delas prescindir por meio do cálculo analítico destes termos. Algumas configurações específicas de pdf das fontes e função $\phi(\cdot)$ permitem este cálculo, como a pdf uniforme com $\phi(y) = y^3$ ou uma pdf exponencial com $\phi(y(k)) = \frac{\text{sign}(y(k))}{\kappa}$ (configuração do próximo experimento). No entanto, em geral os resultados geram fórmulas muito mais complexas (e substancialmente maiores) do que as derivadas da análise EPO.

3.3.1 EXPERIMENTOS

Como exemplo de aplicação da análise de transiente, seja o caso $M = L = 2$, fontes com pdf exponencial (Eq. (23)) e $\phi(y(k)) = \frac{\text{sign}(y(k))}{\kappa}$. O termo $E[\phi(y(k))x(k)]$ pode então ser calculado por:

$$\begin{aligned} E[\phi(y(k))x(k)] &= \frac{h(0)}{4\kappa^4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\frac{c(1)}{c(0)}s_1 - \frac{c(2)}{c(0)}s_2}^{\infty} s_1 e^{-\frac{|s_0|+|s_1|+|s_2|}{\kappa}} ds_0 ds_1 ds_2 \\ &\quad + \frac{h(1)}{4\kappa^4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\frac{c(1)}{c(0)}s_1 - \frac{c(2)}{c(0)}s_2}^{\infty} s_2 e^{-\frac{|s_0|+|s_1|+|s_2|}{\kappa}} ds_0 ds_1 ds_2, \end{aligned} \quad (34)$$

onde, por motivos de concisão, $s_i = s(k-i)$.

O resultado final dos termos $E[\phi(y(k))x(k-i)]$ apresenta grande extensão, sendo por isso omitido. As Figuras 9 e 10 apresentam a comparação entre os resultados simulado e teórico a respeito dos coeficientes adaptativos e da ISI. A Figura 11 apresenta a evolução (simulada e teórica) dos coeficientes adaptativos numa outra escolha da pdf das fontes e da função $\text{score } \phi(\cdot)$. Estas figuras mostram a precisão do cálculo recursivo empregado para a estimativa média da convergência, seja em configurações com β (relativamente) pequeno e elevado, independentemente de o algoritmo tender a uma boa solução ou não.

⁴O fato de ilustrarmos a aplicação da técnica para um caso simples serve apenas para propósitos didáticos, não inutilizando seu emprego em contextos mais complexos.

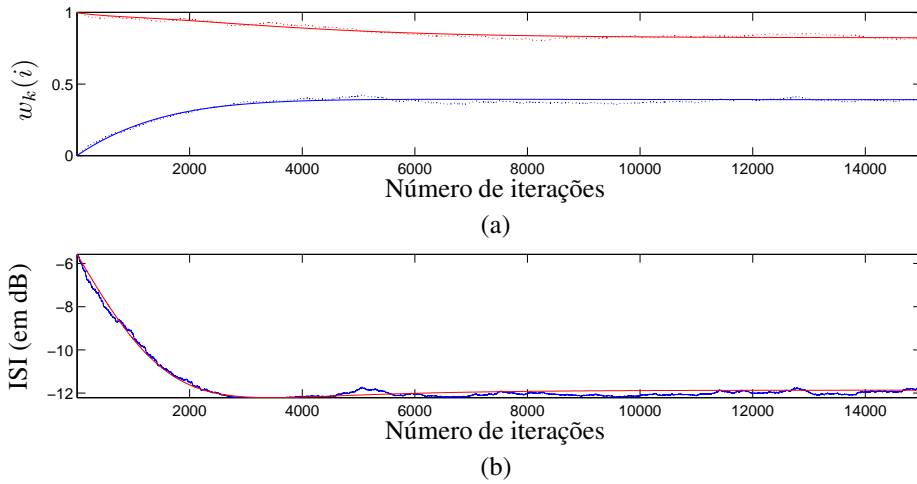


Figura 9: Aplicação da análise de transiente para o algoritmo DCGN $\mathbf{h} = [0, 95 \quad -0, 5]$, com fontes amostradas segundo uma pdf exponencial ($\kappa = 1$), $\beta = 0,5 \times 10^{-3}$ e $\phi(y(k)) = \text{sign}(y(k))$. (a) comparação da evolução real (em linha tracejada) do algoritmo com a estimada (em linha sólida). A cor vermelha refere-se a $w_k(0)$ e a azul, a $w_k(1)$. (b) comparação da evolução da ISI (em dB) real (em azul) e estimada (em vermelho).

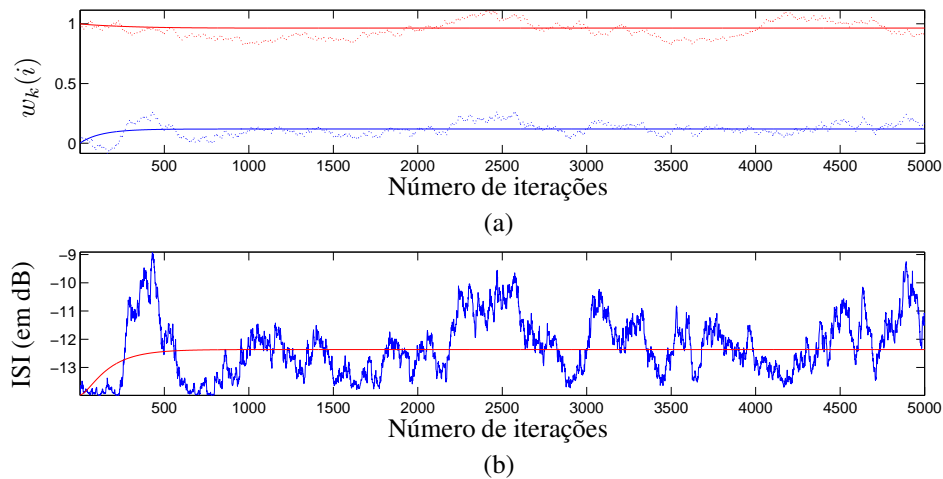


Figura 10: Aplicação da análise de transiente para o algoritmo DCGN para um canal $\mathbf{h} = [0, 2 \quad -1]$, com fontes amostradas segundo uma pdf exponencial ($\kappa = 1$), $\beta = 0,5 \times 10^{-2}$ e $\phi(y(k)) = \text{sign}(y(k))$. (a) comparação da evolução real (em linha tracejada) do algoritmo com a estimada (em linha sólida). A cor vermelha refere-se a $w_k(0)$ e a azul, a $w_k(1)$. (b) comparação da evolução da ISI (em dB) real (em azul) e estimada (em vermelho).

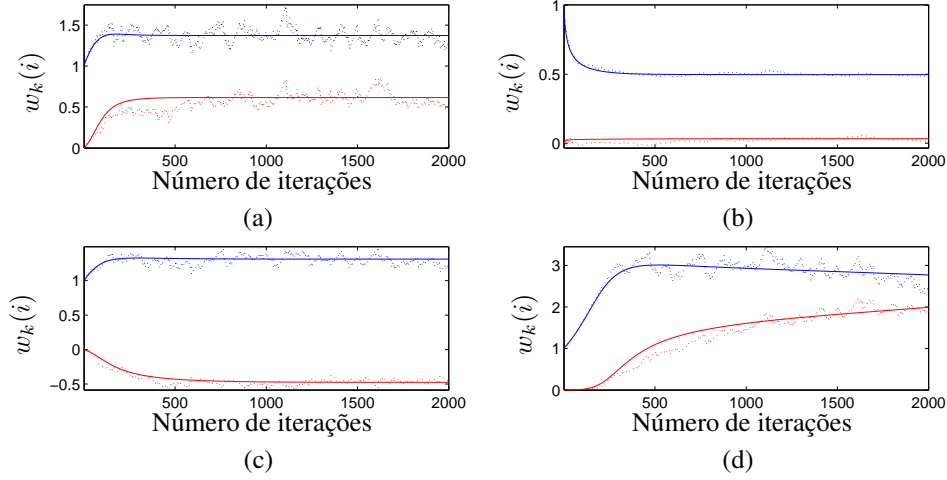


Figura 11: Evolução dos coeficientes $w_k(i)$ ($M = L = 2$) de uma simulação (linha tracejada) do algoritmo DCGN, comparada com a análise de transiente (linha sólida), com as amostras da fonte distribuídas segundo uma pdf uniforme no intervalo $[-1, 1]$ e $\phi(y(k)) = y^3(k)$ (com 2000 iterações). (a) $\beta = 10^{-2}$ e $\mathbf{h} = [-0, 5 \quad 1]$; (b) $\beta = 2 \times 10^{-3}$ e $\mathbf{h} = [-0, 2 \quad 3]$; (c) $\beta = 6 \times 10^{-3}$ e $\mathbf{h} = [-1 \quad -0, 5]$ e (d) $\beta = 5 \times 10^{-3}$ e $\mathbf{h} = [0, 4 \quad -0, 3]$.

3.4 MÉTODO EDO

A técnica de análise de transiente detalhada apresenta semelhanças com o método EDO⁵ [30], [31] de análise do comportamento da convergência de algoritmos iterativos. O método EDO é bastante geral (por exemplo, não se restringe a algoritmos que empregam atualização derivada de gradientes) e provém do amplo arcabouço fornecido pela intitulada “teoria de aproximação estocástica”. Como diz o nome, o método EDO associa a equação de atualização de um algoritmo adaptativo a uma equação diferencial ordinária, de modo que a convergência desta é acoplada à convergência do algoritmo, seja fracamente (quando o fator de aprendizado é constante, mesmo que reduzido [32]), seja fortemente (quando o fator de aprendizado tende assintoticamente a zero, segundo a proposta original [30], [31]).

Uma das vantagens do método EDO reside no emprego de amplos resultados disponíveis na literatura para equações diferenciais ordinárias, o que explica parte do interesse que desperta na comunidade científica. Embora algumas sutilezas possam prejudicar sua acurácia (por exemplo, a influência da parametrização [33]), verificamos que, para algumas configurações simples do algoritmo DCGN, o método EDO engendra resultados compatíveis com os obtidos via simulação.

Conquanto o método EDO seja obtido com alto rigor matemático, apresentaremos uma motivação informal de sua derivação. Seja a Eq. (7), abaixo transcrita por conveniência:

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \beta [\mathbf{w}_k - \phi(y(k))\mathbf{z}_k^T].$$

Sendo o primeiro termo do lado direito desta equação (\mathbf{w}_k) obtido previamente pelo algoritmo DCGN, a equação acima pode ser reescrita da forma:

$$\mathbf{w}_{k+1} = \mathbf{w}_{k-1} + \beta [\mathbf{w}_{k-1} - \phi(y(k-1))\mathbf{z}_{k-1}^T] + \beta [\mathbf{w}_k - \phi(y(k))\mathbf{z}_k^T]. \quad (35)$$

A aplicação deste procedimento recursivo K vezes gera:

$$\mathbf{w}_{k+1} = \mathbf{w}_{k-K} + \sum_{i=k-K}^k \beta [\mathbf{w}_i - \phi(y(i))\mathbf{z}_i^T]. \quad (36)$$

Uma breve manipulação matemática nos permite escrever:

$$\frac{\mathbf{w}_{k+1} - \mathbf{w}_{k-K}}{K\beta} = \frac{1}{K} \sum_{i=k-K}^k [\mathbf{w}_i - \phi(y(i))\mathbf{z}_i^T]. \quad (37)$$

O termo no lado direito pode ser compreendido como uma média e, assumindo β muito pequeno, que \mathbf{w}_i pode ser considerado praticamente constante e ergodicidade dos processos envolvidos, podemos associar a equação diferencial ordinária:

$$\frac{d\mathbf{w}(t)}{dt} = \mathbf{w}(t) - \mathbf{E} [\phi(y(k))\mathbf{z}_k^T] \quad (38)$$

⁵A sigla significa “Equações Diferenciais Ordinárias”.

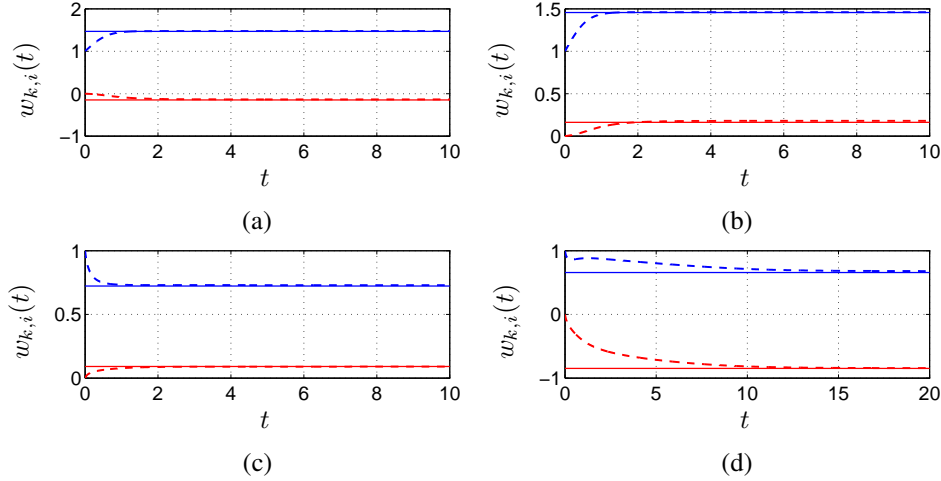


Figura 12: Comparação das funções $w_{k,i}(t)$ (em tracejado) com o valor final dos parâmetros $w_k(0)$ e $w_k(1)$ (em linha sólida; todos foram obtidos com uma simulação com 25000 iterações, obtendo-se a média das últimas 5000 iterações e $\beta = 3 \times 10^{-3}$, $f(y(k)) = y^3(k)$ e $s(k)$ com amostras iid e pdf uniforme no intervalo $[-1, 1]$). (a) filtro do canal $\mathbf{h} = [1 \ 0, 15]$; (b) filtro do canal $\mathbf{h} = [-1 \ 0, 2]$; (c) filtro do canal $\mathbf{h} = [2 \ -0, 4]$ e (d) filtro do canal $\mathbf{h} = [0, 4 \ 1, 5 \ 1]$. Para todas as simulações arbitrou-se $L = 2$, com a inicialização $\mathbf{w}_0 = [0 \ 1]$.

à evolução dos parâmetros $\mathbf{w}_k(i)$ do algoritmo. Como a abordagem supõe um valor de β muito pequeno, não é capaz de providenciar limites superiores para o mesmo. Algumas dessas aproximações, para serem precisamente elaboradas ou contornadas, requerem técnicas matemáticas avançadas, como teoria da medida e equações diferenciais estocásticas. A dinâmica evolutiva da solução da EDO resultante não pode ser associada à taxa de convergência do algoritmo, pois a EDO independe do valor de β (apenas assume-se ser este muito pequeno). Porém, o valor que esta solução assume para $t \rightarrow \infty$ é uma forma de estimar o resultado final do algoritmo.

Nossas simulações verificam que, a despeito da intensa não-linearidade das equações do algoritmo DCGN, a solução numérica da equação diferencial ordinária (Eq. (38)) gera estimativas precisas (em contextos simples), conforme apresenta a Fig. 12. As equações diferenciais não são apresentadas devido à sua grande extensão.

3.5 ESTIMATIVA ANALÍTICA PRECISA EM REGIME PERMANENTE

A análise de transiente efetuada para o algoritmo DCGN contorna duas limitações da abordagem EPO, pois, ao mesmo tempo em que não exige o conhecimento prévio do atraso final do sistema canal + equalizador, prescinde da necessidade de o algoritmo convergir para uma solução de baixa ISI.

Para estimar os valores finais do filtro equalizador, efetuamos o limite $k \rightarrow \infty$ ⁶. Supondo convergência, podemos admitir que $w_{k+1}(i) = w_k(i)$, simplificando as equações recursivas da análise de transiente e eliminando sua dependência com relação a β . Para a configuração contemplada na Eq. (33), por exemplo, resultado final é:

$$\begin{aligned}
 0 &= w_k(0) - w_k^2(0)E[\phi(y(k))x(k)] - w_k^2(1)E[\phi(y(k))x(k)] \\
 &\quad - w_k(0)w_k(1)E[\phi(y(k))x(k-1)], \\
 0 &= w_k(1) - w_k(0)w_k(1)E[\phi(y(k))x(k)] \\
 &\quad - w_k^2(0)E[\phi(y(k))x(k-1)] - w_k^2(1)E[\phi(y(k))x(k-1)].
 \end{aligned}
 \tag{39}$$

Um sistema de equações não lineares é gerado a partir das igualdades acima por meio do cálculo analítico dos termos $E[\phi(y(k))x(k-i)]$. Em algumas configurações, o cálculo das soluções deste sistema de equações é possível, exceto quando o conjunto de soluções pertence a uma variedade de dimensão não-nula (possibilidade que foge ao escopo deste trabalho). Em geral, estas soluções não são enumeráveis quando a função *score* é não-linear e os coeficientes do sistema de mistura são variáveis (sendo estes fixos, mesmo o emprego de funções *score* não lineares não necessariamente prejudica o encontro das soluções, como veremos na próxima subseção). Por outro lado, sendo esta linear, a aproximação linear não é necessária, com o resultado sendo preciso mesmo quando a hipótese de bom desempenho final do algoritmo não é satisfeita. Verificaremos a precisão da análise resultante por meio de um experimento.

⁶Equivalentemente, poderíamos, partindo da análise EDO, impor a condição $t \rightarrow \infty$; porém nesta seção damos preferência à formulação discreta.

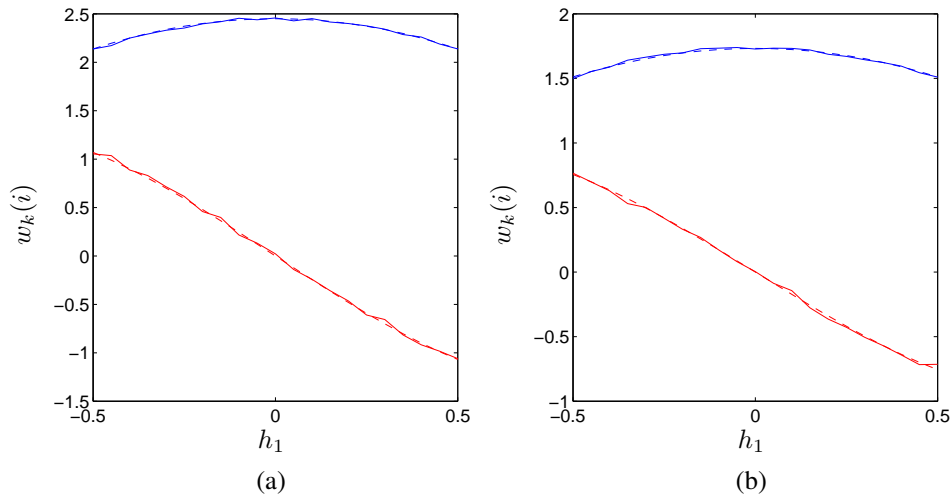


Figura 13: Comparação dos valores para os quais os coeficientes adaptativos do algoritmo DCGN tendem com os valores teóricos, com função *score* parametrizada por $f(y(k)) = \alpha y(k)$, amostras das fontes amostradas segundo uma pdf uniforme e coeficiente do canal h_1 variando no intervalo $[-1/2, 1/2]$. Em vermelho: valores de $w_k(0)$; em azul: valores de $w_k(1)$. A linha sólida apresenta os valores experimentais e a linha tracejada, os teóricos. (a) $\alpha = 0,5$ e (b) $\alpha = 1$.

3.5.1 EXPERIMENTO

Sejam fontes amostradas segundo uma pdf uniforme no intervalo $[-1, 1]$. Se o algoritmo DCGN for aplicado com a função *score*⁷ parametrizada por $f(y(k)) = \alpha y(k)$ para o caso $M = 2$, um possível conjunto de soluções para o sistema de equações⁸ é descrito por:

$$\lim_{k \rightarrow \infty} w_k(0) = \sqrt{\frac{3}{\alpha(h_0^4 + h_0^2 h_1^2 + h_1^4)}} h_1,$$

$$\lim_{k \rightarrow \infty} w_k(1) = -\sqrt{\frac{3}{\alpha(h_0^4 + h_0^2 h_1^2 + h_1^4)}} h_0.$$

Para avaliar a precisão das fórmulas acima, testamos diferentes configurações do algoritmo, com a pdf da fonte seguindo a estipulada pelo cálculo teórico e 40000 iterações do método (com a média dos coeficientes das últimas 10000 iterações para a estimativa empírica dos valores finais do algoritmo). Quatro valores de α foram testados: 1/2, 1, 3/2 e 4. O filtro do canal teve seu segundo coeficiente variado, com o primeiro fixo em $h_0 = 1$. Os resultados podem ser vistos nas Figs. 13 e 14, as quais demonstram a ampla precisão das estimativas teóricas.

Não temos conhecimento de nenhuma publicação científica que proveja soluções analíticas para os valores em regime permanente para os quais tendem o filtro equalizador do algoritmo DCGN em nenhuma configuração, como as derivadas neste trabalho. Estas soluções ajudam a elucidar o comportamento do algoritmo, bem como orientar a escolha do parâmetro α de modo a maximizar o desempenho do algoritmo em algumas configurações. Infelizmente, as equações tornam-se demasiado complexas para configurações complicadas (M e L elevados, ou funções *score* não lineares), o que torna estas fórmulas de valor menos prático do que teórico.

3.6 CRITÉRIO DE SELEÇÃO DAS SOLUÇÕES

Em geral, as soluções analíticas de sistema de equações se apresentam em grande número. Sendo espúrias boa parte destas soluções, dado que não traduzem uma característica real do algoritmo, cumpre encontrar algum critério de seleção.

Várias possibilidades foram testadas, com apenas uma se destacando pela precisão. O critério de maximização da ISI, ainda que interessante, apenas consegue selecionar uma das soluções possíveis, descartando as que poderiam ocorrer, por exemplo, com outras inicializações.

O comportamento qualitativo do sistema dinâmico associado à EDO pode ser estudado, por exemplo, via cálculo do jacobiano do lado direito da Eq. (38) [34]. Este procedimento não fornece resultados robustos, possivelmente por ignorar o efeito de β ; por isso apelamos à proposta de [35], intitulada formulação determinística discreta no tempo (DDT, do inglês *Deterministic*

⁷Escolhemos a função *score* linear porque ela permite um resultado parametrizável em função de variáveis associadas aos coeficientes do sistema de mistura, o que nos permite a confecção de gráficos mais representativos; sendo estes coeficientes conhecidos, a solução pode ser encontrada mesmo que empreguemos funções *score* não lineares, em alguns casos, como mostrará o próximo experimento.

⁸Em verdade, existem 9 soluções possíveis para o sistema resultante. Uma delas é a solução nula, facilmente descartada. As outras ocorrem em pares simétricos, dada a ambiguidade de sinal. Na próxima seção detalharemos um procedimento capaz de selecionar as soluções “realistas”.

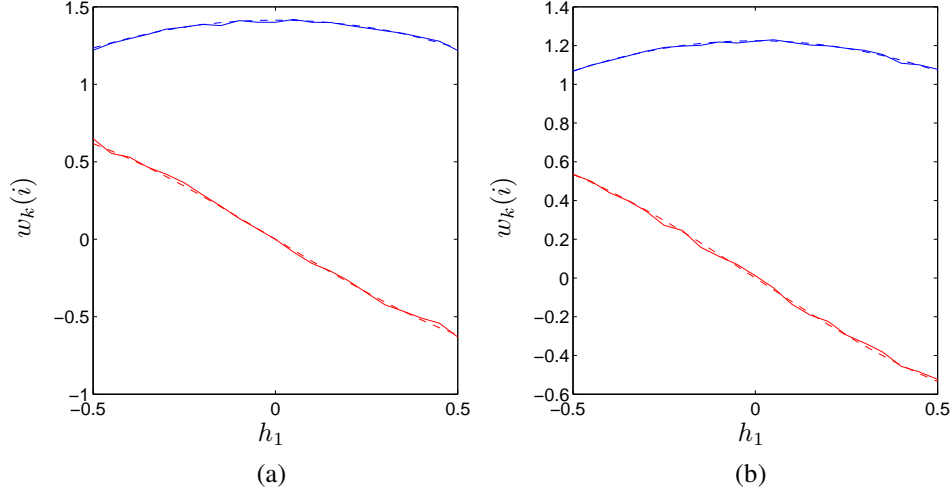


Figura 14: Comparação dos valores para os quais os coeficientes adaptativos do algoritmo DCGN tendem com os valores teóricos, com função *score* parametrizada por $f(y(k)) = \alpha y(k)$, amostras das fontes amostradas segundo uma pdf uniforme e coeficiente do canal h_1 variando no intervalo $[-1/2, 1/2]$. Em vermelho: valores de $w_k(0)$; em azul: valores de $w_k(1)$. A linha sólida apresenta os valores experimentais e a linha tracejada, os teóricos. (a) $\alpha = 1,5$ e (b) $\alpha = 2$.

Discrete-Time).

A formulação DDT atenta diretamente para a natureza discreta da atualização do algoritmo DCGN (Eq. (7)):

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \beta [\mathbf{w}_k - \phi(y(k))\mathbf{z}_k^T] = \mathbf{g}(\mathbf{w}_k), \quad (40)$$

onde a função multidimensional $\mathbf{g}(\mathbf{w}_k)$ incorpora diretamente o efeito do fator de aprendizado. Seguindo o método indireto de Lyapunov [36], o estudo dos autovalores do valor esperado do jacobiano de $\mathbf{g}(\mathbf{w}_k)$ nos pontos de equilíbrio (ou seja, $\mathbf{w}_{k+1} = \mathbf{w}_k$) pode ser empreendido para determinar a estabilidade dos mesmos, pois, na possibilidade de os módulos de todos os autovalores de $\mathbf{g}(\mathbf{w}_k)$ serem inferiores a um, podemos supor a estabilidade do ponto de equilíbrio (ou, em outras palavras, admitir a solução em questão do sistema de equações como factível).

3.6.1 EXPERIMENTO

Seja o caso em que a pdf das fontes é uniforme no intervalo $[-1, 1]$, a função *score* é $f(y(k)) = y^3(k)$, $M = L = 2$ e $\mathbf{h} = [0 \ 0, 1]$. Seja \mathbf{J}_g o jacobiano de $\mathbf{g}(\mathbf{w}_k)$. Colocando parametricamente em função de β , h_0 , h_1 , w_0 e w_1 , os elementos da matriz \mathbf{J}_g são:

$$\begin{aligned} \mathbf{J}_g(1, 1) &= -\beta h^4(0)w^4(0) - (13\beta h^4(0)w^2(0)w^2(1))/5 - (8\beta h^4(0)w^4(1))/15 - (16\beta h^3(0)h(1)w^3(0)w(1))/3 \\ &\quad - (18\beta h^3(0)h(1)w(0)w^3(1))/5 - (10\beta h^2(0)h^2(1)w^4(0))/3 - (38\beta h^2(0)h^2(1)w^2(0)w^2(1))/5 \\ &\quad - (8\beta h^2(0)h^2(1)w^4(1))/5 - (16\beta h(0)h^3(1)w^3(0)w(1))/5 - (58\beta h(0)h^3(1)w(0)w^3(1))/15 \\ &\quad - \beta h^4(1)w^4(0) - (13\beta h^4(1)w^2(0)w^2(1))/5 - (8\beta h^4(1)w^4(1))/15 + \beta + 1, \\ \mathbf{J}_g(1, 2) &= -(26\beta h^4(0)w^3(0)w(1))/15 - (32\beta h^4(0)w(0)w^3(1))/15 - (4\beta h^3(0)h(1)w^4(0))/3 \\ &\quad - (27\beta h^3(0)h(1)w^2(0)w^2(1))/5 - \beta h^3(0)h(1)w^4(1) - (76\beta h^2(0)h^2(1)w^3(0)w(1))/15 \\ &\quad - (32\beta h^2(0)h^2(1)w(0)w^3(1))/5 - (4\beta h(0)h^3(1)w^4(0))/5 - (29\beta h(0)h^3(1)w^2(0)w^2(1))/5 \\ &\quad - (5\beta h(0)h^3(1)w^4(1))/3 - (26\beta h^4(1)w^3(0)w(1))/15 - (32\beta h^4(1)w(0)w^3(1))/15, \\ \mathbf{J}_g(2, 1) &= -(32\beta h^4(0)w^3(0)w(1))/15 - (26\beta h^4(0)w(0)w^3(1))/15 - (5\beta h^3(0)h(1)w^4(0))/3 \\ &\quad - (29\beta h^3(0)h(1)w^2(0)w^2(1))/5 - (4\beta h^3(0)h(1)w^4(1))/5 - (32\beta h^2(0)h^2(1)w^3(0)w(1))/5 \\ &\quad - (76\beta h^2(0)h^2(1)w(0)w^3(1))/15 - \beta h(0)h^3(1)w^4(0) - (27\beta h(0)h^3(1)w^2(0)w^2(1))/5 \\ &\quad - (4\beta h(0)h^3(1)w^4(1))/3 - (32\beta h^4(1)w^3(0)w(1))/15 - (26\beta h^4(1)w(0)w^3(1))/15, \\ \mathbf{J}_g(2, 2) &= 1 + \beta - (8\beta h^4(0)w^4(0))/15 - (13\beta h^4(0)w^2(0)w^2(1))/5 - \beta h^4(0)w^4(1) \\ &\quad - (58\beta h^3(0)h(1)w^3(0)w(1))/15 - (16\beta h^3(0)h(1)w(0)w^3(1))/5 - (8\beta h^2(0)h^2(1)w^4(0))/5 \\ &\quad - (38\beta h^2(0)h^2(1)w^2(0)w^2(1))/5 - (10\beta h^2(0)h^2(1)w^4(1))/3 - (18\beta h(0)h^3(1)w^3(0)w(1))/5 \\ &\quad - (16\beta h(0)h^3(1)w(0)w^3(1))/3 - (8\beta h^4(1)w^4(0))/15 - (13\beta h^4(1)w^2(0)w^2(1))/5 - \beta h^4(1)w^4(1). \quad (41) \end{aligned}$$

O número de soluções ($w(0), w(1)$) do sistema de equações resultante (omitido devido à sua extensão) é igual a 25. Após retirar as soluções simétricas (oriundas da ambiguidade de sinal), trivial (nula) e complexas, temos:

$$\mathbf{S}_w = \{(0.0898, 1.4862), (1.4915, 0.1486), (-0.8504, 0.8373), (0.9200, 0.9435)\}.$$

Logo, restam-nos 4 possíveis soluções. Os autovalores do jacobiano de $\sigma(\mathbf{w}_k)$ referentes às 4 soluções (para $\beta = 0, 1$)⁹ são, respectivamente:

$$\{(0.8387, 0.6), (0.6, 0.8363), (0.6, 1.0860), (1.0792, 0.6)\}.$$

Logo, detectamos que apenas duas soluções (as duas primeiras) apresentam módulos dos autovalores inferiores a 1. E coerentemente, em nossas simulações, o algoritmo converge para uma das duas soluções (a menos de ambiguidades de sinal). O principal fator que determina a solução para a qual o algoritmo converge reside na inicialização do vetor \mathbf{w} , embora um fator de aprendizagem muito elevado possa, em alguns casos, alterar o comportamento final do algoritmo. De todo modo, o procedimento acima ilustra como podemos, de forma cega, eliminar as soluções espúrias.

4. CONCLUSÕES

Neste artigo análises acerca do algoritmo DCGN foram empreendidas e testadas para algumas configurações. Elucidamos em que sentido a desejável propriedade da equivariância pode ser associada a este algoritmo. Estimativas de primeira ordem e mais sofisticadas para os valores em regime permanente dos coeficientes adaptativos foram propostas, assim como análises de transiente via equações discretas recursivas e EDOs. Por fim, um critério de seleção das soluções obtidas pela análise teórica permitiu a eliminação das soluções que não refletem um comportamento real do algoritmo.

AGRADECIMENTOS

Os autores gostariam de agradecer o parcial apoio financeiro da FAPERJ e do CNPq.

REFERÊNCIAS

- [1] H. S. S. C. Douglas and S. Makino. “Natural Gradient Multichannel Blind Deconvolution and Speech Separation Using Causal FIR Filters”. *IEEE Transactions on Audio and Speech Processing*, vol. 13, no. 1, pp. 92–104, Jan. 2005.
- [2] M. G. A. Benveniste and G. Ruget. “Robust Identification of a Nonminimum Phase System: Blind Adjustment of a Linear Equalizer in Data Communications”. *IEEE Transactions on Automatic Control*, vol. 25, no. 3, pp. 385–399, Jun. 1980.
- [3] Y. Sato. “Two Extensional Applications of the Zero-Forcing Equalization Method”. *IEEE Transactions on Communications*, vol. 23, no. 6, pp. 684–687, Jun. 1975.
- [4] D. Godard. “Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems”. *IEEE Transactions on Communications*, vol. 28, no. 11, pp. 1867–1875, Nov. 1980.
- [5] J. Treichler and B. Agee. “A New Approach to Multipath Correction of Constant Modulus Signals”. *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol. 31, no. 2, pp. 459–472, Apr. 1983.
- [6] S. Haykin. *Adaptive Filter Theory*. Prentice Hall, New Jersey, third edition, 1996.
- [7] S. Amari and A. Cichocki. “Adaptive Blind Signal Processing - Neural Network Approaches”. *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2026–2048, Oct. 1998.
- [8] X. Sun and S. C. Douglas. “Mean Square Error Analyses of Multichannel Blind Deconvolution Algorithms”. *Proceedings of the 35th. Annual Asilomar Conference on Signals, Systems, and Computers*, vol. 1, pp. 648–652, Nov. 2001.
- [9] P. C. P. Chevalier, L. Albera and A. Ferreol. “Comparative Performance Analysis of Eight Blind Source Separation Methods on Radiocommunications Signals”. *IEEE International Joint Conference on Neural Networks (IJCNN 2004)*, pp. 273–278, Jul. 2004.
- [10] C. S. R. Y. Singh. “A Comparison of BSS Algorithms”. *IEEE International Joint Conference on Neural Networks (IJCNN 2001)*, vol. 1, pp. 932–935, Jul. 2001.
- [11] M. Fozunbal. “Performance Analysis for Blind Identification of Acoustic Channels”. *IEEE Transactions on Automatic Control*, pp. 2658–2661, Mar. 2010.
- [12] S. Abrar and A. K. Nandi. “Adaptive Minimum Entropy Equalization Algorithm”. *IEEE Communications Letters*, vol. 14, no. 10, pp. 966–968, 2010.
- [13] M. S. Radenkovic, T. Bose and B. Ramkumar. “Blind Adaptive Equalization of MIMO Systems: New Recursive Algorithms and Convergence Analysis”. *IEEE Transactions on Circuits and Systems - I: Regular Papers*, vol. 57, no. 7, pp. 1475–1488, Jul. 2010.

⁹Em nossas simulações, qualquer valor de β inferior a 0,5 não impacta a escolha das soluções factíveis. Como na prática os fatores de aprendizagem adquirem valores muito menores do que 0,5, basta-nos arbitrar um valor para β no intervalo $(0, 1/2)$.

- [14] S. Abrar and A. K. Nandi. “Blind Equalization of Square-QAM Signals: A Multimodulus Approach”. *IEEE Transactions on Communications*, vol. 58, no. 6, pp. 1674–1685, Jun. 2010.
- [15] J. B. e W. Minker. *Time-Domain Beamforming and Blind Source Separation - Speech Input in the Car Environment*. Springer, New York, first edition, 2007.
- [16] J. P. LeBlanc and C. R. J. Jr. “Global CMA Error Surface Characteristics, Source Statistic Effects: Polytopes and Manifolds”. *13th International Conference on Digital Signal Processing Proceedings*, pp. 131–134, Jul. 1997.
- [17] W. C. and C. R. J. Jr. “Characterization of the Regions of Convergence of CMA Adapted Blind Fractionally Spaced Equalizer”. *Conference Record of the Thirty-Second Asilomar Conference on Signals, Systems & Computers*, vol. 1, pp. 493–497, Nov. 1998.
- [18] R. Cusani and A. Laurenti. “Convergence Analysis of the CMA Blind Equalizer”. *IEEE Transactions on Communications*, vol. 43, no. 2, pp. 1304–1307, Feb. 1995.
- [19] A. Neves and C. Panazio. “A Class of Channels Resulting in Ill-Convergence for CMA in Decision Feedback Equalizers”. *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5736–5743, Nov. 2010.
- [20] J. Mai and A. H. Sayed. “A Feedback Approach to the Steady-State Performance of Fractionally Spaced Blind Adaptive Equalizers”. *IEEE Transactions on Signal Processing*, vol. 48, no. 1, pp. 80–91, Jan. 2000.
- [21] D. B. Haddad, M. R. Petraglia and P. B. Batalheiro. “Análises Exata e Aproximada de um Algoritmo de Deconvolução Cega”. *IX Congresso de Redes Neurais*, pp. 1–5, Out. 2009.
- [22] D. B. Haddad, M. R. Petraglia and P. B. Batalheiro. “Equivariância e Desempenho de um Algoritmo de Deconvolução Cega”. *IX Congresso de Redes Neurais*, pp. 1–5, Out. 2009.
- [23] D. B. Haddad, M. R. Petraglia and P. B. Batalheiro. “Convergência e Desempenho de um Algoritmo de Desconvolução Cega”. *Congresso Brasileiro de Automática*, pp. 1–8, Out. 2010.
- [24] A. C. S. Amari, S. C. Douglas and H. H. Yang. “Multichannel Blind Deconvolution and Equalization using the Natural Gradient”. *IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, vol. 13, no. 1, pp. 101–104, Apr. 1997.
- [25] A. C. L. Zhang and S. Amari. “Multichannel Blind Deconvolution of Nonminimum-Phase Systems Using Filter Decomposition”. *IEEE Transactions on Signal Processing*, vol. 52, no. 5, pp. 1430 – 1442, May 2004.
- [26] C. F. e S. J. Godsill. “A Bayesian Approach for Blind Separation of Sparse Sources”. *IEEE Transactions on Audio and Speech Processing*, vol. 14, no. 6, pp. 2174–2188, Nov. 2006.
- [27] R. A. K. C. R. J. J. P. LeBlanc, K. Dogancay. “Effects of Input Data Correlation on the Convergence of Blind Adaptive Equalizers”. *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 1994)*, vol. 3, pp. 313–316, Apr. 1994.
- [28] J.-F. Cardoso and B. H. Laheld. “Equivariant Adaptive Source Separation”. *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 3017 – 3030, Dec. 1996.
- [29] R. A. H. Buchner and W. Kellerman. “A Generalization of Blind Source Separation Algorithms for Convolutional Mixtures based on Second-Order Statistics”. *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 11, pp. 120–134, Jan. 2005.
- [30] L. Ljung. “On Positive Real Transfer Functions and the Convergence of Some Recursive Schemes”. *IEEE Transactions on Automatic Control*, vol. 22, no. 4, pp. 539–551, Aug. 1977.
- [31] L. Ljung. “Analysis of Recursive Stochastic Algorithms”. *IEEE Transactions on Automatic Control*, vol. 22, no. 4, pp. 551–575, Aug. 1977.
- [32] H. J. Kushner. *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*. Birkhäuser, Boston, first edition, 1990.
- [33] D. M. W. T. D. Powell and D. Ljunquist. “A Counterexample of Proof of Convergence Using Asymptotic Averaging”. *Proceedings of the 32th Conference on Decision and Control*, pp. 2060–2061, Dec. 1993.
- [34] H. K. Khalil. *Nonlinear Systems*. Prentice Hall, New Jersey, third edition, 2002.
- [35] P. J. Zufiria. “On the Discrete-Time Dynamics of the Basic Hebbian Neural-Network Node”. *IEEE Transactions on Neural Networks*, vol. 13, no. 6, pp. 1342–1352, Nov. 2002.
- [36] P. A. Regalia. *Adaptive IIR Filtering in Signal Processing and Control*. Marcel Dekker, New York, first edition, 1995.