INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING
Int. J. Adapt. Control Signal Process. 2006; **20**:321–336
Published online 3 April 2006 in Wiley InterScience (www.interscience.wiley.com) DOI:10.1002/acs.902

Convergence analysis of blind image deconvolution via dispersion minimization

C. Vural[1,*,†] and W. A. Sethares[2]

[1] *Electrical-Electronics Engineering, Sakarya University, Esentepe, Sakarya 54187, Turkey*
[2] *Electrical and Computer Engineering Department, University of Wisconsin-Madison, 1415 Engineering Drive, Madison, WI 53706, U.S.A.*

SUMMARY

A new non-linear adaptive filter called *blind image deconvolution via dispersion minimization* has recently been proposed for restoring noisy blurred images blindly. This is essentially a two-dimensional version of the constant modulus algorithm that is well known in the field of blind equalization. The two-dimensional extension has been shown capable of reconstructing noisy blurred images using partial *a priori* information about the true image and the point spread function in a variety of situations by means of simulations. This paper analyses the behaviour of the algorithm by investigating the static properties of the cost function and the dynamic convergence of the parameter estimates. The theoretical results are supported with computer simulations. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: blind image deconvolution; non-linear adaptive filtering; constant modulus algorithm; convergence analysis

1. INTRODUCTION

The new method for reconstructing noisy blurred images [1] called *blind image deconvolution via dispersion minimization*, is a two-dimensional (2-D) extension of the constant modulus algorithm (CMA) [2, 3]. Simulations show that the method is useful in recovering noisy images that are blurred with an unknown linear shift invariant (LSI) point spread function (PSF). This paper analyses the behaviour of the algorithm.

The problem is analogous to that of blind equalization. An unknown signal (the message in a communications setting, the unknown underlying image in the present setting) is passed through a linear system (the channel in communications systems, the blur or the PSF in the present setting) and then further corrupted by noise. The result is a signal (the received signal in the communications system, the observed image in the present setting), and the goal is to build a

*Correspondence to: C. Vural, Electrical-Electronics Engineering, Sakarya University, Esentepe, Sakarya 54187, Turkey.
† E-mail: cvural@sakarya.edu.tr

Received 21 March 2005
Revised 25 November 2005
Accepted 3 February 2006

non-linear system that deconvolves the signal in an attempt to recover the underlying image (or the unknown message).

In the blind equalization setting, the CMA algorithm, which uses a cost function that minimizes the dispersion of the signal about some constant γ called *dispersion constant*, is one successful approach. CMA is applicable whenever the unknown input arises from a finite alphabet, and the reader is referred to Reference [4] and the references therein for a comprehensive introduction to the CMA and its analysis in the context of (1-D) adaptive equalization. Since the pixels in a digitized picture are drawn from a 'finite alphabet' (usually 256 levels, though sometimes as few as two),[‡] the dispersion cost function may also be useful in the deblurring and denoising of images. The goal in this paper is to analyse the dispersion minimization algorithm in two-dimensions.

The analysis proceeds in two steps, generalizing the basic procedure in Reference [5], though our analysis does not ignore the presence of additive noise. The static analysis studies the locations of the minimum points of the cost function and shows that under suitable conditions, there is a minimum of the cost surface located near the statistically optimum Wiener solution. This is significant because minima of the dispersion cost can be found without the knowledge of the input, whereas the Wiener solution requires the knowledge of the input. The dynamic analysis investigates the stochastic dynamics of the new method. Results are given that describe the mean convergence of the adaptive parameters once the trajectories are in the neighbourhood of the Wiener solution. This can be used to study the amount of excess noise caused by truncation error due to the (finite) length of the deconvolution filter.

In the remainder of the paper, it will be assumed that the adaptive filter coefficients are close to the global minimum of the dispersion cost function. Under this assumption, a closed-form expression for the coefficients of the adaptive filter is derived in the static analysis. Convergence and consistency of the coefficients are investigated in the dynamic analysis. Only the independence assumption (see References [6, 7]) is required, and results show that for a given PSF and step-size, there is an optimum support for the adaptive filter. This result can be used as a guideline in designing blind image deconvolution algorithms.

The paper is organized as follows: the method is described in detail in Section 2, properties of the prediction error function are examined in Section 3, and the static convergence analysis is performed in Section 4. Section 5 investigates the dynamic convergence behaviour of the 2-D CMA algorithm and Section 6 provides computer simulation results. The final section concludes.

2. BLIND IMAGE DECONVOLUTION VIA DISPERSION MINIMIZATION

Consider the single-input single-output (SISO) LSI system depicted in Figure 1, in which $f(m,n)$ and $h(m,n)$ represent the (m,n)th pixel of the zero-mean independent identically distributed (i.i.d.) true image and the PSF of the degrading system. The zero-mean i.i.d. additive noise $v(m,n)$ is assumed to be independent of both the true image and the system. The output of the model is the observed $M \times N$ noisy blurred image $g(m,n)$, which can be written as the

[‡]Many fax machines and laser printers use 2-level quantization.

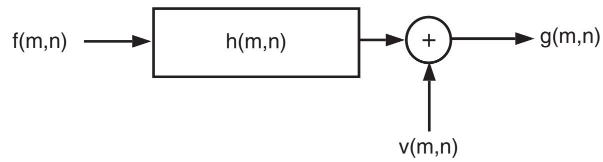

Figure 1. Linear image degradation model.

convolution of $h(m,n)$ and $f(m,n)$ as

$$g(m,n) = \sum_{k=-A}^{A} \sum_{l=-B}^{B} h(k,l) f(m-k, n-l) + v(m,n) \tag{1}$$

for $m = 1, \ldots, M$ and $n = 1, \ldots, N$, where $[-A, A] \times [-B, B]$ is the support of the PSF.

In linear image restoration, the PSF $h(m,n)$ is assumed to be known, and the true image $f(m,n)$ is estimated using any of a number of well-known linear estimation methods [8, 9]. However, the PSF of the degrading system is usually unknown in most real imaging applications. Hence, the true image must be estimated directly from the noisy blurred observed image $g(m,n)$ without knowing the true image or the PSF. While the underlying true image is unknown, certain statistical properties are known; typically the pixel values must be one of a small number of possibilities. This process is called *blind image deconvolution*.

As shown in Reference [1], ambiguities in both absolute gain and delay are inherent to blind image deconvolution. That is, scaling the true image pixel values by α and the PSF coefficients by α^{-1} simultaneously does not change the observed image, where α is a real fixed non-zero gain. Similarly, advancing the true image by an integer-valued vector while delaying the PSF by the same vector has no effect on the observed image. Keeping these ambiguities in mind, blind image deconvolution problem can be stated more clearly as follows:

> Obtain an estimate of the form $\hat{f}(m,n) = \alpha f(m - m_0, n - n_0)$ for some real $\alpha \neq 0$ and some integers m_0, n_0 from the observed image $g(m,n)$ without knowledge of the true image $f(m,n)$ and the PSF $h(m,n)$.

Kundur and Hatzinakos [10, 11] provide excellent tutorials, which explain different blind image deconvolution methods that can be categorized into two major groups: (i) those which estimate the PSF *a priori* independent of the true image so as to use it later with one of the linear image restoration methods, and (ii) those which estimate the PSF and the true image simultaneously. Algorithms belonging to the first class are computationally simple, but they are limited to situations in which the PSF has a special parametric form, and the true image has certain features. Algorithms belonging to the second class, which are computationally more complex must be used for more general situations. The method proposed in Reference [1] belongs to the second class.

Figure 2 shows a block diagram of the blind algorithm based on dispersion minimization, in which a 2-D linear finite impulse response (FIR) filter $w(m,n)$ is used to deblur the observed noisy blurred image $g(m,n)$. The coefficients of the deblurring filter are the adaptive parameters, and are updated via a gradient algorithm. The adaptive filter output at the jth iteration at

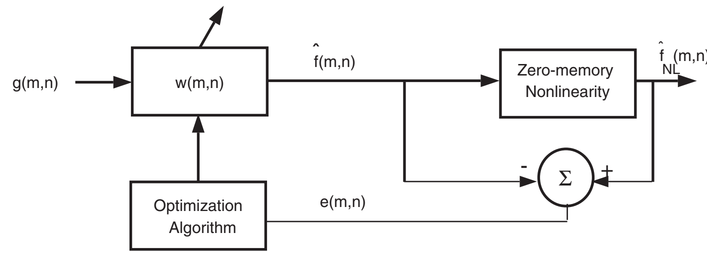

Figure 2. Block diagram of the blind image deconvolution via dispersion minimization.

pixel (m, n) can be written as

$$\hat{f}_j(m,n) = \sum_{k=-C}^{C} \sum_{l=-D}^{D} w_j(k,l) g(m-k, n-l) \tag{2}$$

where $[-C, C] \times [-D, D]$ is the support of the adaptive filter, $w_j(m, n)$ are the adaptive filter coefficients at the jth iteration. $\hat{f}_j(m, n)$ can also be written as

$$\hat{f}_j(m,n) = \sum_{k=-P}^{P} \sum_{l=-Q}^{Q} y_j(k,l) f(m-k, n-l) + \sum_{k=-C}^{C} \sum_{l=-D}^{D} w_j(k,l) v(m-k, n-l) \tag{3}$$

where $y_j(m, n) := h(m, n) * w_j(m, n)$ is the global impulse response at the jth iteration, $P = A + C$ and $Q = B + D$. If the true image $f(m, n)$ were known, then the difference between $\hat{f}_j(m, n)$ and $f(m, n)$ could be used to provide an efficient update of the unknown parameters. In blind image deconvolution, however, the true image is unavailable. As in blind equalization, one possibility is to attempt to minimize the dispersion of $\hat{f}_j(m, n)$ using the constant modulus (CM) cost J_{CM} defined by

$$J_{\text{CM}} := E[(\hat{f}_j^2(m,n) - \gamma)^2] \tag{4}$$

where γ is the *dispersion constant*. This constant γ is one piece of statistical information that must be available about the true image, and it is given by

$$\gamma = \frac{E[f^4(m,n)]}{E[f^2(m,n)]} \tag{5}$$

It is clear from Equation (4) that J_{CM} penalizes the deviations of $\hat{f}_j^2(m, n)$ from the dispersion constant γ. Since it is not possible to minimize an expected value directly, the method uses an instantaneous estimate of J_{CM} given by

$$J := \tfrac{1}{4}(\hat{f}_j^2(m,n) - \gamma)^2 \tag{6}$$

to obtain an implementable algorithm. Because the true image $f(m, n)$ is unknown, a desired image (an artificial true image) must somehow be generated so that an error term that drives the adaptive algorithm can be obtained. Note that the function of the zero-memory non-linearity (the rightmost term in Figure 2) is to generate a desired image $\hat{f}_{\text{NL}}(m, n)$ from $\hat{f}(m, n)$. The zero-memory non-linearity is chosen such that the error term $e(m, n) := \hat{f}_{\text{NL}} - \hat{f}(m, n)$ corresponds to negative of the gradient of J. Hence, the coefficients of the adaptive filter coefficients are updated according to

$$w_{j+1}(k,l) = w_j(k,l) - \mu \frac{\mathrm{d}J}{\mathrm{d}w_j(k,l)}$$
$$= w_j(k,l) - \mu\phi(\hat{f}_j(m,n))g(m-k,n-l) \tag{7}$$

for $-C \leqslant k \leqslant C$, $-D \leqslant l \leqslant D$, where μ is a small positive step-size and $\phi(\hat{f}_j(m,n)) := [\hat{f}_j^2(m,n) - \gamma]\hat{f}_j(m,n)$ is called the *prediction error function*. Note that both J_{CM} and J depend on the iteration j, though this dependence was not included in the notation for the sake of simplicity.

Writing Equation (7) for $k = -C, \ldots, C$ and $l = -D, \ldots, D$ gives

$$\mathbf{w}_{j+1} = \mathbf{w}_j - \mu\mathbf{g}(m,n)\phi(\hat{f}_j(m,n)) \tag{8}$$

where $\mathbf{w}_j$ is the lexicographically ordered coefficient vector of the adaptive filter at the jth iteration defined as

$$\mathbf{w}_j := [w_j(-C,-D), \ldots, w_j(C,D)]^{\mathrm{T}} \tag{9}$$

and $\mathbf{g}(m,n)$ is the lexicographically ordered adaptive filter input vector at pixel (m,n) defined as

$$\mathbf{g}(m,n) := [g(m+C,n+D), \ldots, g(m-C,n-D)]^{\mathrm{T}} \tag{10}$$

Note that as explained in Reference [1], $\mathbf{g}(m,n)$ can be rewritten as

$$\mathbf{g}(m,n) = H^{\mathrm{T}}\mathbf{f}(m,n) + \mathbf{v}(m,n) \tag{11}$$

where H is a suitable $(2P+1)(2Q+1) \times (2C+1)(2D+1)$ blur matrix whose coefficients are constructed from the PSF $h(m,n)$, and where $\mathbf{f}(m,n)$ and $\mathbf{v}(m,n)$ are the lexicographically ordered true image vector and the additive noise vector for pixel (m,n) given by

$$\mathbf{f}(m,n) := [f(m+P,n+Q), \ldots, f(m-P,n-Q)]^{\mathrm{T}} \tag{12}$$

$$\mathbf{v}(m,n) := [v(m+C,n+D), \ldots, v(m-C,n-D)]^{\mathrm{T}} \tag{13}$$

The output of the adaptive filter given in Equation (2) can be written in a compact form using the above definitions as

$$\hat{f}_j(m,n) = \mathbf{w}_j^{\mathrm{T}}H^{\mathrm{T}}\mathbf{f}(m,n) + \mathbf{w}_j^{\mathrm{T}}\mathbf{v}(m,n)$$
$$= \mathbf{f}^{\mathrm{T}}(m,n)\mathbf{y}_j + \mathbf{w}_j^{\mathrm{T}}\mathbf{v}(m,n) \tag{14}$$

where $\mathbf{y}_j := H\mathbf{w}_j$ is the global impulse response vector at the jth iteration, i.e.

$$\mathbf{y}_j := [y_j(-P,-Q), \ldots, y_j(P,Q)] \tag{15}$$

When convergence occurs, coefficients of the adaptive filter provide an approximate inverse of the PSF. Furthermore, the output of the adaptive filter $\hat{f}(m,n)$ is an estimate of the true image. The output of the zero-memory non-linearity at the jth $\hat{f}_{\mathrm{NL}}(m,n)$ is given by

$$\hat{f}_{\mathrm{NL}}(m,n) = \hat{f}(m,n) + \phi(\hat{f}(m,n))$$
$$= \hat{f}^3(m,n) + (1-\gamma)\hat{f}(m,n)$$

3. PROPERTIES OF THE PREDICTION ERROR FUNCTION

This section states two lemmas that provide some useful properties of the prediction error function $\phi(\cdot)$, and proofs are provided in the Appendices. The first lemma shows that when the algorithm attains one of its isolated minima, the expected value of the update term in the algorithm is zero and the second derivative of the cost function is positive definite.

Lemma 1

Let $\tilde{w}_j(m,n)$ be the coefficients of the adaptive filter that make J_{CM} attain one of its isolated minima. Then, output of the adaptive filter $\tilde{f}_j(m,n)$ satisfies

(a) $E[\phi(\tilde{f}_j(m,n))\mathbf{g}(m,n)] = \mathbf{0}$
(b) $H^{\mathrm{T}}\tilde{F}H + \tilde{V}$ is positive definite

where

$$\tilde{F} := E[\mathbf{f}(m,n)\phi'(\tilde{f}_j(m,n))\mathbf{f}^{\mathrm{T}}(m,n)] \tag{16}$$

$$\tilde{V} := E[\mathbf{v}(m,n)\phi'(\tilde{f}_j(m,n))\mathbf{v}^{\mathrm{T}}(m,n)] \tag{17}$$

and $\phi'(\cdot)$ is derivative of $\phi(\cdot)$.

The second lemma will be used in the static analysis to provide a closed-form expression for the coefficients of the adaptive filter near the global minimum of J_{CM}.

Lemma 2

When the adaptive filter has sufficiently large support, the true image has a uniform probability density function, and the additive noise is small, then the prediction error function has the following properties

(a) $E[\phi(f(m,n))f(k,l)] = 0$
(b) $E[\phi'(f(m,n))f^2(k,l))] > 0$

for all $1 \leqslant m, k, \leqslant M$ and $1 \leqslant n, l \leqslant N$.

4. STATIC CONVERGENCE ANALYSIS

This section derives a closed-form expression for the coefficients of the adaptive filter near the global minimum of J_{CM}. Note that in the statement of Theorem 1, $\mathbf{h}_{(m_0,n_0)}$ represents a unit vector $[0,\ldots,0,1,\ldots,0]^{\mathrm{T}}$, where the non-zero coefficient is in the (m_0,n_0)th position which must satisfy $-P \leqslant m_0 \leqslant P$, $-Q \leqslant n_0 \leqslant Q$.

Theorem 1

At the minimum near the global minimum of J_{CM}, the coefficients of the adaptive filter can be expressed as

$$\tilde{\mathbf{w}} = E[\phi'(f(m,n))f^2(m,n)]R^{-1}H^{\mathrm{T}}\mathbf{h}_{(0,0)} \tag{18}$$

where

$$R = H^{\mathrm{T}}FH + V$$

$$F := E[\mathbf{f}(m,n)\phi'(f(m,n))\mathbf{f}^{\mathrm{T}}(m,n)] \tag{19}$$

$$V := E[\mathbf{v}(m,n)\phi'(f(m,n))\mathbf{v}^{\mathrm{T}}(m,n)] \tag{20}$$

if H has full column rank and the true image $f(m,n)$ has a uniform probability density function.

Proof

Writing the Taylor series expansion of $\phi(\tilde{f}(m,n))$ around $f(m,n)$ and ignoring second- and higher-order terms give

$$\phi(\tilde{f}(m,n)) = \phi(f(m,n)) + \phi'(f(m,n))[\tilde{f}(m,n) - f(m,n)] \tag{21}$$

By Lemmas 1 and 2, multiplying both sides of Equation (21) by $\mathbf{g}(m,n)$ and taking expectations result in

$$E[\mathbf{g}(m,n)\phi'(f(m,n))[\tilde{f}(m,n) - f(m,n)]] = \mathbf{0} \tag{22}$$

Substituting $\mathbf{g}(m,n) = H^{\mathrm{T}}\mathbf{f}(m,n) + \mathbf{v}(m,n)$ and $\tilde{f}(m,n) = \mathbf{g}^{\mathrm{T}}(m,n)\tilde{\mathbf{w}}$, using the definitions of F and V given in (19) and (20) in Equation (22) give

$$R\tilde{\mathbf{w}} = H^{\mathrm{T}}E[\mathbf{f}(m,n)e'(f(m,n))f(m,n)] \tag{23}$$

Note that the RHS of Equation (23) can be written as

$$H^{\mathrm{T}}E[\mathbf{f}(m,n)\phi'(m,n)f(m,n)] = E[\phi'(m,n)f^2(m,n)]H^{\mathrm{T}}\mathbf{h}_{(0,0)}$$

since $f(m,n)$ are i.i.d. Hence, Equation (23) reduces to

$$R\tilde{\mathbf{w}} = E[\phi'(m,n)f^2(m,n)]H^{\mathrm{T}}\mathbf{h}_{(0,0)} \tag{24}$$

Consequently, $\tilde{\mathbf{w}}$ is given by Equation (18) if and only if R is non-singular. Since R is non-negative definite by structure, non-singularity implies that it is positive definite. Note that F is a diagonal matrix whose entries are of the form

$$E[\phi'(f(m,n))f(i,j)f(k,l)] = \begin{cases} E[\phi'(f(m,n))f^2(m,n)] & \text{if } m=i=k \text{ and } n=j=l \\ \sigma_f^2 E[\phi'(f(m,n))] & \text{if } m\neq i=k \text{ and } n\neq j=l \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

which are positive by Lemma 2(b). Hence, F is positive definite. If H is full column rank, then $R = H^{\mathrm{T}}FH$ is also positive definite. It is sufficient now to show that V is non-negative definite to complete the proof. Note that since $v(m,n)$ and $f(m,n)$ are independent, V can be written as

$$\begin{aligned} V &= E[\mathbf{v}(m,n)\mathbf{v}^{\mathrm{T}}(m,n)]E[\phi'(f(m,n))] \\ &= \sigma_v^2(3\sigma_f^2 - \gamma)I \end{aligned}$$

Table I. Values of $3\sigma_f^2 - \gamma$ for a zero-mean uniformly distributed image having different grey levels.

Level	σ_f^2	γ	$3\sigma_f^2 - \gamma$
2	1.000000	1.00000	2.000000
4	5.000000	8.20000	6.800000
8	21.55000	37.0000	27.65000
16	85.00000	152.200	102.8000
32	340.9900	613.000	409.9700
64	1365.010	2456.20	1638.830
128	5460.850	9829.00	6553.570
256	21844.44	39320.0	26213.33

Because $\sigma_v^2 > 0$, it is sufficient to show that $(3\sigma_f^2 - \gamma) \geqslant 0$. The true image $f(m,n)$ was assumed to have a zero-mean uniform probability distribution. For such an image, Table I shows values of $3\sigma_f^2 - \gamma$ for different grey levels from which it is easy see that $(3\sigma_f^2 - \gamma) \geqslant 0$ and the proof is complete. □

Theorem 1 can be used to find a necessary and sufficient condition such that $\tilde{\mathbf{w}} = \mathbf{w}_0$, where $\mathbf{w}_0$ is the optimum Wiener filter. As shown in Reference [1], the optimum Wiener filter $\mathbf{w}_0$ which minimizes $E[(\hat{f}^2(m,n) - f(m,n))^2]$ is given by

$$\mathbf{w}_0 = (H^{\mathrm{T}}H + \lambda I)^{-1} H^{\mathrm{T}} \mathbf{h}_{(0,0)} \tag{26}$$

where $\lambda = \sigma_v^2/\sigma_f^2$. Comparing Equations (18) and (26), a necessary and sufficient condition for $\tilde{\mathbf{w}} = \mathbf{w}_0$ is that

$$H^{\mathrm{T}}H + \lambda I = \frac{1}{E[\phi'(f(m,n))f^2(m,n)]}(H^{\mathrm{T}}FH + V) \tag{27}$$

The distortion introduced by the new method due to finite support of the adaptive filter and additive noise is

$$\begin{aligned} D_1 &:= E[(\tilde{f}(m,n) - f(m,n))^2] \\ &= \sigma_f^2 \|\tilde{\mathbf{y}} - \mathbf{h}_{(0,0)}\|_2^2 + \sigma_v^2 \|\tilde{\mathbf{w}}\|_2^2 \end{aligned}$$

where $\|\cdot\|_2$ represents the l_2-norm of a vector. D_1 decreases as the support of the adaptive filter $(-C, C) \times (D, D)$ gets larger because the global minimum of J_{CM} will be closer to the optimum Wiener solution. Hence, D_1 is inversely proportional to the support of the adaptive filter.

5. DYNAMIC CONVERGENCE ANALYSIS

The dynamic convergence analysis is presented in this section. Since algorithm (7) uses an instantaneous estimate of J_{CM}, it suffers from gradient noise. Rather than converging to the global minimum of J_{CM}, the adaptive filter coefficients exhibit a random motion about the minimum. Hence, it will be useful to work with the *excess error vector* at the jth iteration defined as

$$\boldsymbol{\varepsilon}_j := \mathbf{w}_j - \tilde{\mathbf{w}} \tag{28}$$

where $\tilde{\mathbf{w}}$ is the global minimum of J_{CM}. During the presentation it will be assumed that $\mathbf{w}_j$ is near $\tilde{\mathbf{w}}$ and that the true image $f(m,n)$ and the additive noise $v(m,n)$ are independent of $\boldsymbol{\varepsilon}_j$. It is the objective of this section to provide mean convergence behaviour of the adaptive filter and evaluate the correlation matrix of $\boldsymbol{\varepsilon}_j$. The following theorems are readily proved using the independence assumption.

Theorem 2

Suppose that the largest eigenvalue of $\tilde{R} = H^{\mathrm{T}}\tilde{F}H + \tilde{V}$ is $\lambda_{\max}$. Then, mean convergence behaviour of the adaptive filter near $\tilde{\mathbf{w}}$ satisfies

$$E[\boldsymbol{\varepsilon}_j] = [I - \mu(H^{\mathrm{T}}\tilde{F}H + \tilde{V})]^j E[\boldsymbol{\varepsilon}_0] \tag{29}$$

If the step-size satisfies $0 < \mu < (2/\lambda_{\max})$, then $E[\mathbf{w}_j] \to \tilde{\mathbf{w}}$.

Proof

Let $\hat{f}_j(m,n)$ and $\tilde{f}(m,n)$ be the outputs of the adaptive filter resulting from $\mathbf{w}_j$ and $\tilde{\mathbf{w}}$, respectively. Then, $\hat{f}_j(m,n)$ can be written as

$$\hat{f}_j(m,n) = \tilde{f}(m,n) + [\hat{f}_j(m,n) - \tilde{f}(m,n)] \tag{30}$$

Substituting $\hat{f}_j(m,n) = \mathbf{f}^{\mathrm{T}}(m,n)H\mathbf{w}_j + \mathbf{w}_j^{\mathrm{T}}\mathbf{v}(m,n)$, $\tilde{f}(m,n) = \mathbf{f}^{\mathrm{T}}(m,n)H\tilde{\mathbf{w}} + \tilde{\mathbf{w}}^{\mathrm{T}}\mathbf{v}(m,n)$ in (30) and using the definition of ε_j give

$$\hat{f}_j(m,n) = \tilde{f}(m,n) + [\mathbf{f}^{\mathrm{T}}(m,n)H + \mathbf{v}^{\mathrm{T}}(m,n)]\varepsilon_j \tag{31}$$

Near the global minimum of J_{CM}, $\|\varepsilon_j\|_2$ is small. So is $\hat{f}_j(m,n) - \tilde{f}(m,n)$. Writing the Taylor series expansion of $\phi(\hat{f}_j(m,n))$ around $\tilde{f}(m,n)$ and ignoring second- and higher-order terms result in

$$\phi(\hat{f}_j(m,n)) = \phi(\tilde{f}(m,n)) + \phi'(\tilde{f}(m,n))[\hat{f}_j(m,n) - \tilde{f}(m,n)] \tag{32}$$

Using (31), Equation (32) can be written as

$$\phi(\hat{f}_j(m,n)) = \phi(\tilde{f}(m,n)) + \phi'(\tilde{f}(m,n))[\mathbf{f}^{\mathrm{T}}(m,n)H + \mathbf{v}^{\mathrm{T}}(m,n)]\varepsilon_j \tag{33}$$

Recall that the adaptive filter is updated via

$$\mathbf{w}_{j+1} = \mathbf{w}_j - \mu\mathbf{g}(m,n)\phi(\hat{f}_j(m,n)) \tag{34}$$

Substituting (32) in (34) and subtracting $\tilde{\mathbf{w}}$ from both sides give

$$\varepsilon_{j+1} = \varepsilon_j - \mu[\mathbf{g}(m,n)\phi(\tilde{f}(m,n)) + \mathbf{g}(m,n)\phi'(\tilde{f}(m,n))\mathbf{g}^{\mathrm{T}}(m,n)\varepsilon_j] \tag{35}$$

Taking expectations of both sides of (35), using the independence assumption and definitions of $\tilde{F}$ and $\tilde{V}$ lead to

$$E[\varepsilon_{j+1}] = [I - \mu(H^{\mathrm{T}}\tilde{F}H + \tilde{V})]E[\varepsilon_j] \tag{36}$$

which can be iterated to obtain

$$E[\varepsilon_j] = [I - \mu(H^{\mathrm{T}}\tilde{F}H + \tilde{V})]^j E[\varepsilon_0] \tag{37}$$

Therefore, $E[\varepsilon_j] \to 0$, which means that $E[\mathbf{w}_j] \to \tilde{\mathbf{w}}$ as long as $0 < \mu < (2/\lambda_{\max})$ [6], which concludes the proof. □

The next result evaluates the correlation matrix of ε_j.

Theorem 3

The adaptive filter vector $\mathbf{w}_j \to \tilde{\mathbf{w}}$ is not consistent. At the equilibrium near the global minimum of J_{CM}, the correlation matrix of ε_j is uniquely determined by the following Lyapunov equation:

$$\tilde{R}R_\varepsilon + R_\varepsilon\tilde{R} = \mu\tilde{R}_{\mathrm{g}} \tag{38}$$

where $\tilde{R}_{\mathrm{g}} = H^{\mathrm{T}}\tilde{G}H + \tilde{T}$ and

$$\tilde{G} := E[\mathbf{f}(m,n)\phi^2(f(m,n))\mathbf{f}^{\mathrm{T}}(m,n)] \tag{39}$$

$$\tilde{T} := E[\mathbf{v}(m,n)\phi^2(f(m,n))\mathbf{v}^{\mathrm{T}}(m,n)] \tag{40}$$

if $0 < \mu < (1/\lambda_{\max})$, where $\lambda_{\max}$ is the maximum of eigenvalue of $\tilde{R} := H^{\mathrm{T}}\tilde{F}H + \tilde{V}$.

Proof

From the definition of ε_{j+1} given in Equation (35), $E[\varepsilon_{j+1}\varepsilon_{j+1}^{\mathrm{T}}]$ can be written as

$$E[\varepsilon_{j+1}\varepsilon_{j+1}^{\mathrm{T}}] = E[\varepsilon_j\varepsilon_j^{\mathrm{T}}] - \mu(I_1 + I_2 - \mu I_3) \tag{41}$$

where $\mathbf{z} := \mathbf{g}(m,n)\phi(\tilde{f}(m,n)) + \mathbf{g}(m,n)\phi'(\tilde{f}(m,n))\mathbf{g}^{\mathrm{T}}(m,n)\varepsilon_j$ and

$$I_1 := E[\mathbf{z}\varepsilon_j^{\mathrm{T}}] \tag{42}$$

$$I_2 := E[\varepsilon_j\mathbf{z}^{\mathrm{T}}] \tag{43}$$

$$I_3 := E[\mathbf{z}\mathbf{z}^{\mathrm{T}}] \tag{44}$$

I_1 and I_2 can be simplified using Lemma 1 as

$$I_1 = (H^{\mathrm{T}}\tilde{F}H + \tilde{V})R_{\varepsilon_j} \tag{45}$$

$$I_2 = R_{\varepsilon_j}(H^{\mathrm{T}}\tilde{F}H + \tilde{V}) \tag{46}$$

where $R_{\varepsilon_j} = E[\varepsilon_j\varepsilon_j^{\mathrm{T}}]$. The dominant term in I_3 is $\mathbf{g}(m,n)\phi^2(\tilde{f}(m,n))\mathbf{g}^{\mathrm{T}}(m,n)$. Hence, I_3 can be approximated as

$$\begin{aligned} I_3 &= E[\mathbf{g}(m,n)\phi^2(\tilde{f}(m,n))\mathbf{g}^{\mathrm{T}}(m,n)] \\ &= H^{\mathrm{T}}\tilde{G}H + \tilde{T} \\ &= \tilde{R}_{\mathrm{g}} \end{aligned} \tag{47}$$

Substituting (45), (46) and (47) in (41) and rearranging terms give

$$E[\varepsilon_{j+1}\varepsilon_{j+1}^{\mathrm{T}}] = E[\varepsilon_j\varepsilon_j^{\mathrm{T}}] - \mu(\tilde{R}R_\varepsilon + R_\varepsilon\tilde{R} - \mu\tilde{R}_{\mathrm{g}}) \tag{48}$$

Let R_ε be the unique positive definite solution of the Lyapunov equation

$$\tilde{R}R_\varepsilon + R_\varepsilon\tilde{R} = \mu\tilde{R}_{\mathrm{g}} \tag{49}$$

Substituting (49) in (48) and subtracting R_ε from both sides lead to

$$R_{\varepsilon_{j+1}} - R_\varepsilon = (I - 2\mu\tilde{R})(R_{\varepsilon_j} - R_\varepsilon) \tag{50}$$

Consequently,

$$\|R_{\varepsilon_{j+1}} - R_\varepsilon\| = \|I - 2\mu\tilde{R}\|\|R_{\varepsilon_j} - R_\varepsilon\| \tag{51}$$

which can be iterated to obtain

$$\|R_{\varepsilon_j} - R_\varepsilon\| = \|I - 2\mu\tilde{R}\|^j\|R_{\varepsilon_0} - R_\varepsilon\| \tag{52}$$

If $0 < \mu < (1/\lambda_{\max})$, then $\|I - 2\mu\tilde{R}\| < 1$ and

$$\lim_{j\to\infty}\|R_{\varepsilon_j} - R_\varepsilon\| = 0 \tag{53}$$

Therefore, $\lim_{j\to\infty}R_{\varepsilon_j} \to R_\varepsilon$. Since R_ε is positive definite, $\mathbf{w}_j \to \tilde{\mathbf{w}}$ is not consistent and the proof is complete. □

Adaptive filter coefficients exhibit a random motion around the global minimum at convergence. Hence, the distortion due to gradient noise is

$$\begin{aligned} D_2 &:= E[\|\mathbf{y} - \tilde{\mathbf{y}}\|^2] \\ &= E[\|H\varepsilon\|^2] \\ &= E[\mathrm{tr}(\varepsilon^{\mathrm{T}}H^{\mathrm{T}}H\varepsilon)] \\ &= \mathrm{tr}(E[\varepsilon\varepsilon^{\mathrm{T}}H^{\mathrm{T}}H]) \\ &= \mathrm{tr}(R_\varepsilon H^{\mathrm{T}}H) \end{aligned}$$

When, $\tilde{f}(m,n) \approx f(m,n)$, then $\tilde{F} = F$ and $\tilde{R}_{\mathrm{g}} = R_{\mathrm{g}}$, where

$$R_{\mathrm{g}} = H^{\mathrm{T}}GH + T$$

$$G := E[\mathbf{f}(m,n)\phi^2(f(m,n))\mathbf{f}^{\mathrm{T}}(m,n)]$$

$$T := E[\mathbf{v}(m,n)\phi^2(f(m,n))\mathbf{f}^{\mathrm{T}}(m,n)]$$

For this case,

$$RR_\varepsilon + R_\varepsilon R = \mu R_{\mathrm{g}} \tag{54}$$

If $E[\phi'(f(m,n))f^2(m,n)] = \sigma_f^2 E[\phi'(f(m,n))]$, then $R = \sigma_f^2 E[\phi'(f(m,n))] H^T H$. Hence,

$$D_2 = \operatorname{tr}\left(\frac{1}{2\sigma_f^2 E[\phi'(f(m,n))]}\mu R_g\right) = \frac{1}{2\sigma_f^2 E[\phi'(f(m,n))]}\mu \operatorname{tr}(R_g) \tag{55}$$

When $E[\phi'(f(m,n))f^2(m,n)] \neq \sigma_f^2 E[\phi'(f(m,n))]$, (55) can still be used to approximate D_2. Note that R_g is a diagonal matrix of dimension $(2C+1)(2D+1) \times (2C+1)(2D+1)$, where $(-C, C) \times (-D, D)$ is the support of the adaptive filter. Hence, D_2 increases as the support of the adaptive filter gets larger.

The total distortion is $D = D_1 + D_2$. D_1 is inversely proportional to the support of the adaptive filter while D_2 is proportional to it. Consequently, there should be an optimum support once the step-size μ is fixed.

6. SIMULATION RESULTS

The theoretical results are supported with computer simulations in this section. The classical 8-bit grey-scale *pepper* image was chosen a test image. The procedure described [1] was used to obtain true images with several grey levels, which fulfil most of the assumptions made about the true image. For the sake of simplicity, a square support was used for the adaptive filter in all simulations. A 2-D impulse function was used to initialize the adaptive filter.

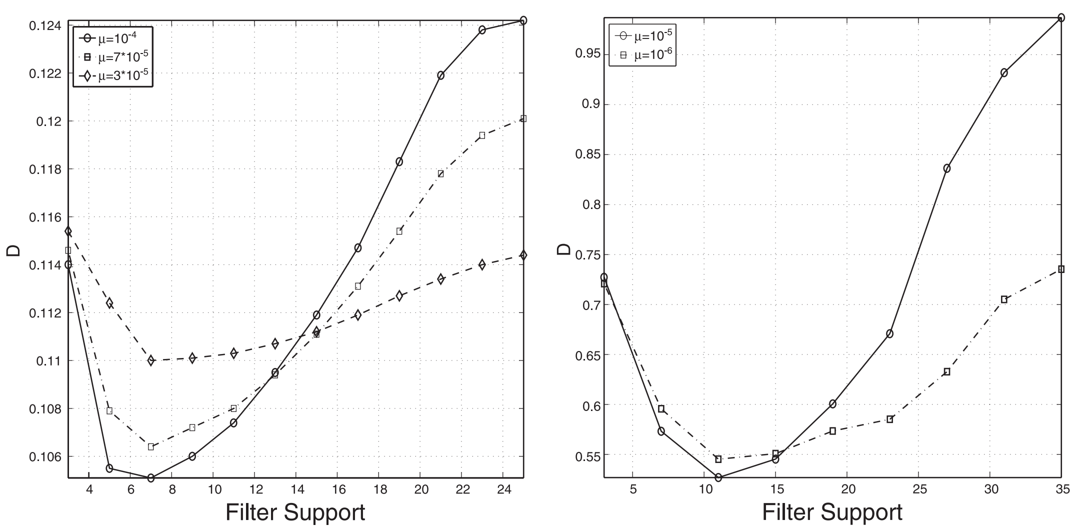

Figure 3. Dependence of D on the filter support for the motion blur. $L = 2$ (left) and $L = 4$ (right).

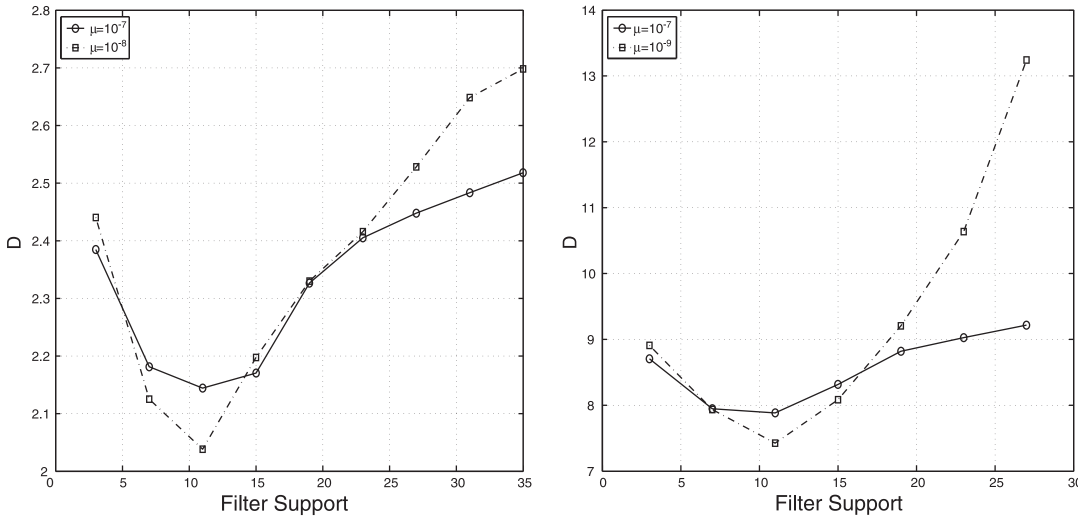

Figure 4. Dependence of D on the filter support for the motion blur. $L = 8$ (left) and $L = 16$ (right).

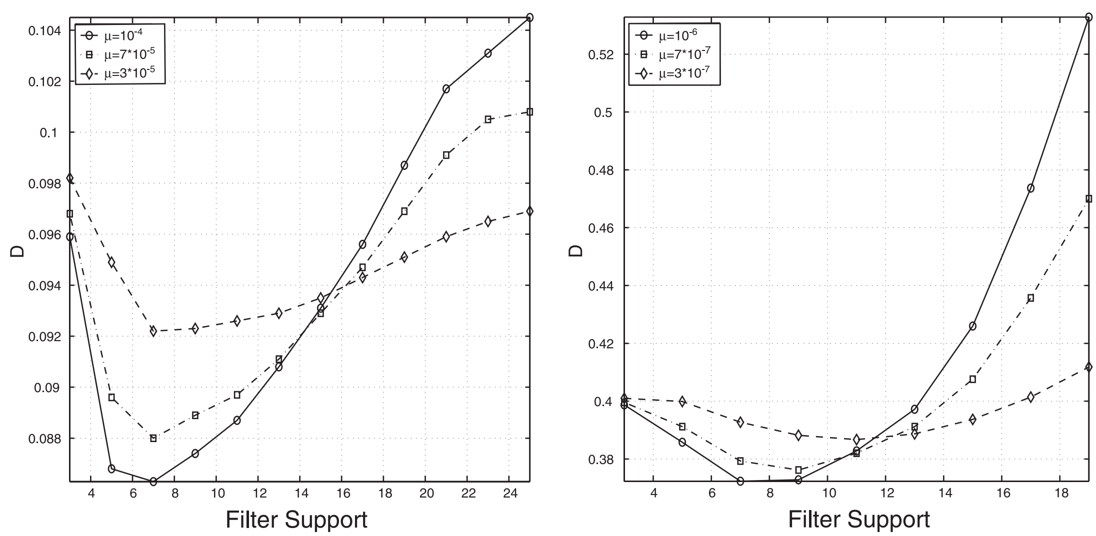

Figure 5. Dependence of D on the adaptive filter support for the scatter blur. $L = 2$ (left) and $L = 4$ (right).

The theory is first validated for a 1×5 horizontal motion blur at 50 dB blurred signal-to-noise ratio (BSNR). Relationships between the total distortion D given by

$$D = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (f(m,n) - \hat{f}(m,n))^2 \tag{56}$$

and the support of the adaptive filters are illustrated in Figures 3 and 4 for several grey-level true images. Observe that for each level there is an optimum support for a given step-size.

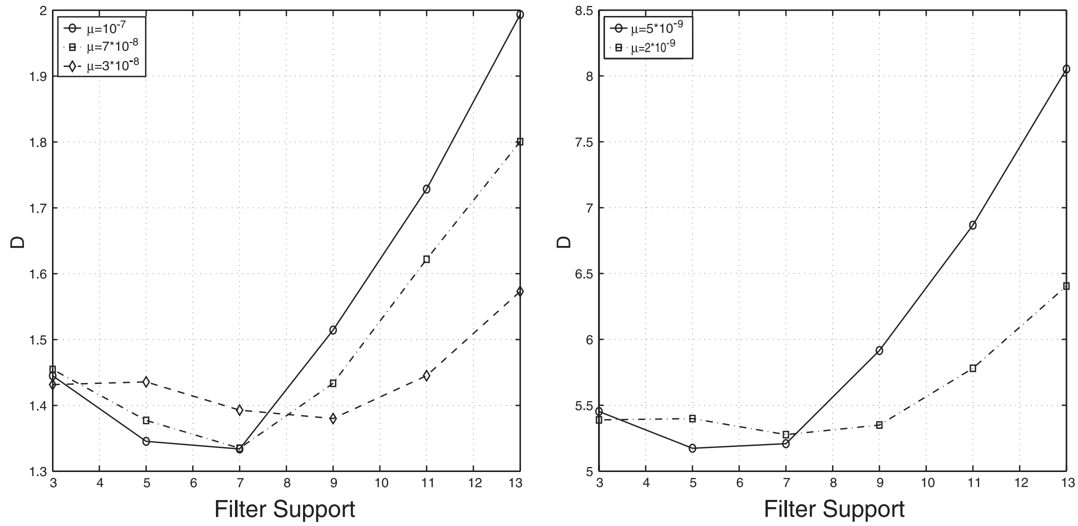

Figure 6. Dependence of D on the adaptive filter support for the scatter blur. $L = 8$ (left) and $L = 16$ (right).

Next, the theory is validated for a 5×5 scatter blur with $\beta = 1$ at 50 dB BSNR. Dependences of the total distortion D on the adaptive filter supports are given in Figures 5 and 6 for several grey levels. As in the horizontal motion blur case, there is an optimum support for the adaptive filter for each level.

7. CONCLUSIONS

The static and dynamic convergence behaviour of the blind image deconvolution method proposed in Reference [1] was studied in this paper. The analysis is based on the first-order approximation of the CM cost under the *independence assumption*. Analysis has shown that under suitable conditions, the global minimum of the CM cost is equal to the statistically optimum Wiener solution. In addition, for a given blur and step-size there is an optimum support for the adaptive filter to deblur the observed noisy blurred image. This striking result is contrary to the common folklore that increasing FIR filter support results in increased performance for the adaptive algorithms. Computer simulation results were provided to validate the theory in a variety of settings.

It must be noted before concluding the paper that an autoregressive (AR) adaptive filter could be used to implement the new method as well. In AR implementation, there is no need to find the optimum support experimentally provided that support of the blur is known. Furthermore, distortion introduced due to the finite support of the adaptive filter could be made zero in theory. In addition, AR implementation provides an estimate of the blur (not an approximate inverse of the blur as is the case in FIR implementation) at convergence. Despite these important advantages, AR implementation has one limitation that makes its use difficult in real applications: derivation of the algorithm and its implementation are not trivial. For further details, see Reference [12].

APPENDIX A: PROOF OF LEMMA 1

(a) If $\tilde{w}_j(m,n)$ are the adaptive filter coefficients that make $J_{\rm CM}$ attain one of its isolated minima, then the gradient $\nabla J_{\rm CM}$ is equal to zero vector and the Hessian $\mathscr{H}$ is positive definite, i.e.

$$\nabla J_{\rm CM} = \begin{bmatrix} \dfrac{\partial J_{\rm CM}}{\partial \tilde{w}_j(-C,-D)} \\ \vdots \\ \dfrac{\partial J_{\rm CM}}{\partial \tilde{w}_j(C,D)} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \tag{A1}$$

and

$$\mathscr{H} = \begin{bmatrix} \dfrac{\partial^2 J_{\rm CM}}{\partial \tilde{w}_j^2(-C,-D)} & \cdots & \dfrac{\partial^2 J_{\rm CM}}{\partial \tilde{w}_j(-C,-D)\partial \tilde{w}_j(C,D)} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 J_{\rm CM}}{\partial \tilde{w}_j(C,D)\partial \tilde{w}_j(-C,-D)} & \cdots & \dfrac{\partial^2 J_{\rm CM}}{\partial \tilde{w}_j^2(C,D)} \end{bmatrix} \tag{A2}$$

is positive definite. For $-C \leqslant k \leqslant C$ and $-D \leqslant l \leqslant D$,

$$\begin{aligned} \frac{\partial J_{\rm CM}}{\partial \tilde{w}_j(k,l)} &= \frac{\partial}{\partial \tilde{w}_j(k,l)} E[(\tilde{f}_j^2(m,n) - \gamma)^2] \\ &= 4E[\phi(\tilde{f}_j(m,n))g(m-k,n-l)] \end{aligned}$$

Therefore, by Equation (A1)

$$E[\phi(\tilde{f}_j(m,n))g(m-k,n-l)] = 0 \tag{A3}$$

Writing Equation (A3) for $-C \leqslant k \leqslant C$ and $-D \leqslant l \leqslant D$ results in Lemma 1(a).
(b) Also, for $-C \leqslant k, r \leqslant C$ and $-D \leqslant l, s \leqslant D$

$$\begin{aligned} \frac{\partial^2 J_{\rm CM}}{\partial \tilde{w}_j(k,l)\partial \tilde{w}_j(r,s)} &= \frac{\partial^2}{\partial \tilde{w}_j(k,l)\partial \tilde{w}_j(r,s)} E[(\tilde{f}_j^2(m,n) - \gamma)^2] \\ &= 4E[g(m-k,n-l)\phi'(\tilde{f}_j(m,n))g(m-r,n-s)] \end{aligned} \tag{A4}$$

It is clear from (A4) that the Hessian $\mathscr{H}$ can be written as

$$\mathscr{H} = 4E[\mathbf{g}(m,n)\phi'(\tilde{f}_j(m,n))]\mathbf{g}^{\rm T}(m,n)] \tag{A5}$$

Substituting $\mathbf{g}(m,n) = H^{\rm T}\mathbf{f}(m,n) + \mathbf{v}(m,n)$ in (A5), using definitions of $\tilde{F}$ and $\tilde{V}$ given by (16) and (17), and noting that $f(m,n)$ and $v(m,n)$ are independent, give

$$\mathscr{H} = 4E[H^{\rm T}\tilde{F}H + \tilde{V}]$$

Since $\mathscr{H}$ is positive definite, $H^{\rm T}\tilde{F}H + \tilde{V}$ is also positive definite.

APPENDIX B: PROOF OF LEMMA 2

Suppose that the adaptive filter has sufficiently large support, the true image has a uniform probability density function, and the amount of additive noise is small, then for sufficiently large j, $\tilde{\mathbf{w}}_j = \mathbf{w}_0$ and $\tilde{f}_j(m,n) \approx f(m,n)$, where $\mathbf{w}_0$ is the optimum Wiener filter. Hence,
(a)

$$E[\phi(f(m,n))f(k,l)] = E[\phi(\tilde{f}_j(m,n))\sum_r \sum_s \tilde{w}_j(r,s)g(k-r,l-s)]$$
$$= \sum_r \sum_s \tilde{w}_j(r,s)E[\phi(\tilde{f}_j(m,n))g(k-r,l-s)] = 0$$

by Lemma 1(a).
(b) Since the Hessian $\mathscr{H}$ is positive definite

$$\sum_r \sum_s \sum_u \sum_t \tilde{w}_j(r,s)\tilde{w}_j(u,t)E[g(k-r,l-s)\phi'(\tilde{f}(m,n))g(k-u,l-t)] > 0$$

As a consequence, for sufficiently large j

$$E[\phi'(f(m,n))f^2(k,l)] = E[\phi'(\tilde{f}(m,n))\sum_r \sum_s \tilde{w}_j(r,s)g(k-r,l-s)$$
$$\times \sum_u \sum_t \tilde{w}_j(u,t)g(k-u,l-t)]$$
$$= \sum_r \sum_s \sum_u \sum_t \tilde{w}_j(r,s)\tilde{w}_j(u,t)$$
$$\times E[g(k-r,l-s)\phi'(\tilde{f}(m,n))g(k-u,l-t)] > 0$$

REFERENCES

1. Vural C, Sethares WA. Blind image deconvolution via dispersion minimization. *Digital Signal Processing* 2006; **16**(2):137–148.
2. Godard D. Self-recovering equalization and carrier tracking in two dimensional data communication systems. *IEEE Transactions on Communication* 1980; **28**(11):1867–1875.
3. Treichler JR, Agee BG. A new approach to multipath correction of constant modulus signals. *IEEE Transactions on Communication* 1983; **31**(2):459–473.
4. Johnson JR *et al*. Blind equalization using the constant modulus criterion: a review. *Proceedings of the IEEE* 1998; **86**(10):1927–1950.
5. Li Y, Liu KJR. Static and dynamic convergence behavior of adaptive blind equalizers. *IEEE Transactions on Signal Processing* 1996; **44**(11):2736–2745.
6. Haykin S. *Adaptive Filter Theory* (3rd edn). Prentice-Hall: Englewood Cliffs, NJ, 1996.
7. Macchi O. *Adaptive Processing*. Wiley: New York, NY, 1995.
8. Bahnam MA, Katsaggelos AK. Digital image restoration. *Signal Processing Magazine* 1997; **14**(2):24–41.
9. Sezan MI, Tekalp AM. Survey of recent developments in digital image restoration. *Optical Engineering* 1990; **29**(5):393–404.
10. Kundur D, Hatzinakos D. Blind image deconvolution. *IEEE Signal Processing Magazine* 1996; **13**(3):43–64.
11. Kundur D, Hatzinakos D. Blind image deconvolution revisited. *IEEE Signal Processing Magazine* 1996; **13**(6):61–63.
12. Vural C, Sethares WA. Recursive blind image deconvolution via dispersion minimization. *International Journal of Adaptive Control and Signal Processing* 2005; **19**(8):601–622.

