Post-Nonlinear Blind Extraction in the Presence of Ill-Conditioned Mixing

Wai Yie Leong and Danilo P. Mandic, *Senior Member, IEEE*

***Abstract*—An extension of blind source extraction (BSE) of one or a group of sources to the case of ill—conditioned and post-nonlinear (PNL) mixing is introduced. This is achieved by a "mixed objective" type of cost function which jointly maximizes the kurtosis of a recovered source and estimates a measure of nonlinearity within the mixing system. This helps to circumvent problems with existing BSE methods, which are limited to noiseless and linear mixing models. Simulations illustrate the performance of the proposed algorithm and its usefulness, especially in the presence of very ill-conditioned mixing systems.**

***Index Terms*—Blind source extraction, blind source separation, deflation, ill-conditioned matrix, post-nonlinear (PNL) model.**

NOMENCLATURE

β	Sign of the kurtosis.
$\hbar$	Open subset of $\mathbf{x}$.
$\triangle w_1$	Small change applied to weight $w_1 \tilde{f}(\cdot)$ nonlinear function.
η_0	Learning rate for continuous-time algorithm.
σ	(Sigma) Standard deviation.
$\Im[\cdot]$	Set of variables.
$[\cdot]^T$	Superscript $(^T)$denotes transpose operator.
$\langle\cdot\rangle$	Estimation operator.
$[\cdot]^H$	Complex conjugate, transpose.
Δw_1	Small change applied to weight w_1
$\mathbf{A}$	Mixing matrix.
$C_0(\mathbb{R})$	Nonlinear mapping.
$\mathrm{cum}(\cdot)$	Cumulant.
$\boldsymbol{E}$	Expected value.
exp	Exponential.
$\boldsymbol{I}$ or $\boldsymbol{I}_n$	Identity matric or identity matrix of dimension $n \times n$.
$J(\cdot)$	Cost function.
k	Discrete-time or number of iterations applied.
$\mathcal{K}(\mathbf{A})$	Condition number.
K_{norm}	Normalized kurtosis.
kurt	Kurtosis.
log	Natural logarithm.
m_q	Moments.
n	Number of possible inputs.
$\mathbb{R}$	Real dimensional parameter space.
s_1	First source signal.
$\mathbf{s}$	Vector variable of source signals.
$\mathrm{sign}(y)$	Sign function (= 1 for $y > 0$ and $= -1$ for $y < 0$).
tanh	Hyperbolic tangent.
x_1	First signal mixture.
$\mathbf{x}(t)$	Post-nonlinear mixtures
$\mathbf{w}_1$	First weight vector.
$\mathbf{W}$	=[w_{ij}] — Extraction matrix.
y_1	Signal extracted from $\mathbf{x}$ by $\mathbf{w}_1$.

Manuscript received January 26, 2007; revised September 30, 2007. First published April 18, 2008; current version published October 29, 2008. This paper was recommended by Associate Editor A. Kuh.

W. Leong is with the Communications and Signal Processing Group, Department of Electronics and Electrical Engineering, Imperial College London, London, SW7 2AZ, U.K., and also with Singapore Institute of Manufacturing Technology, 638075 Singapore (e-mail: waiyie@ieee.org).

D. Mandic is with the Communications and Signal Processing Group, Department of Electronics and Electrical Engineering, Imperial College London, London, SW7 2AZ, U.K. (e-mail: d.mandic@imperial.ac.uk)

Digital Object Identifier 10.1109/TCSI.2008.922022

I. INTRODUCTION

BLIND SIGNAL SEPARATION (BSS) [4], [20], [22] aims at recovering unobservable signals (sources) from their linear or nonlinear mixtures. This technique has recently attracted much interest due to its potentially wide number of applications. Despite the present progress in the theory of BSS [17], standard algorithms have been typically designed for noiseless and linear mixtures, a rather simplistic case. To the end, much effort has been dedicated to BSS for ill-conditioned and nonlinear mixing. In those cases we may as well opt to recover only a small subset of "interesting" sources in an ill-conditioned system, that is to perform blind source extraction (BSE).[1] A combination of BSE and deflation was originally proposed in [21], and has been subsequently further extended [23], [24], [27], [28], [31]–[33]. However, the main limitation of the existing BSE algorithms is that they have been specifically designed for linear instantaneous mixtures, a condition which is not realistic for most real world situations. To help mitigate some of these limitations, we set out to extend existing BSE techniques and derive criteria and algorithms for simultaneous post-nonlinear [9] extraction of arbitrary groups of n (where n denotes the total number of sources) signals of interest.

[1]As special case, the BSE problem ought to be treated different from BSS. BSS is meant to perform signal separation simultaneously, whereas BSE extracts individual signals sequentially.

There are a variety of BSE algorithms in the open literature, including those based on high-order statistics (HOS) (such as the kurtosis) and those based on second-order statistics (SOS) [30], such as the structure using a linear predictor [19]. In the latter case, it has been widely assumed that as long as the source signals exhibit different temporal structures, the minimisation of the mean squared prediction error will lead to successful linear extraction. However, BSE conducted this way may exhibit a relatively low success rate.

The need for the inverse modelling of an ill-conditioned and post-nonlinear system [9], [14], [29] arises in many real world situations, yet this case is rarely addressed when developing BSE algorithms. It is however clear that (1) sensors normally posses nonlinear transfer characteristics; (2) the effects of reflections and interfering signals may introduce an ill-conditioned mathematical model of the mixing system.

To help circumvent the problems associated with the assumption of linear mixing, an extension in the form of post-nonlinear mixing system is proven to be considerably more applicable, as it allows for nonlinear mixing features to be included within the system model. This approach has already attracted considerable interest [9], [14], [29]. Namely, in the post-nonlinear mixing model, the linear ICA theory[2] and the commonly exploited equivariance property might not be powerful enough to model the underlying nonlinear mapping, and BSS algorithms for the linear mixing model will generally fail. We therefore need to resort to nonlinear models, and make use of their more general approximation capabilities [8], [9]. One such set of algorithms for blind separating post-nonlinear mixtures using parametric nonlinear functions was proposed by Lee [26]. It was assumed that the mixing is performed in two stages: a linear mixing process followed by a nonlinear transfer function. The focus was on a parametric sigmoidal nonlinearity and on high-order polynomials. It was further shown in [6], that for general nonlinear ICA, there always exist an infinite number of solutions if the space of the nonlinear mixing functions is unlimited, and hence the independent components extracted from the observations are not necessarily the true source signals. Furthermore, in general, nonlinear ICA suffers from high computational complexity. Solving the nonlinear BSS problem based only on the independence assumption is possible only in some special cases, for example, when the mixtures are post-nonlinear (PNL), and using some weak assumptions [9].

To that cause, in this paper, we consider the BSE problem in the presence of i) PNL mixtures and ii) ill-conditioned mixing. Notice that in this case BSE should be treated differently from BSS.

II. Post-Nonlinear Mixtures

Consider n unknown zero-mean sources $\mathbf{s}(k) = [s_1(k), \ldots, s_n(k)]^T$ at a discrete time instant k. Sources are observed through a nonlinear vector mapping $\boldsymbol{f}(\cdot)$ and (possibly ill-conditioned)[3] mixing matrix $\mathbf{A}$, to give measurements $\mathbf{x}(k)$. This nonlinear mixing problem (from the unknown sources $\mathbf{s}(k)$ to the observation $\mathbf{x}(k)$) can be mathematically described as a post-nonlinear system. We can therefore assume the signals $\mathbf{x}(k)$ are nonlinear memoryless mixtures of n unknown statistically independent sources $\mathbf{s}(k)$, and that the observation process can be expressed as (Fig. 1)

$$\mathbf{x}(k) = \boldsymbol{f}(\mathbf{A}\mathbf{s}(k)) \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is an unknown mixing matrix which is assumed to be nonsingular.

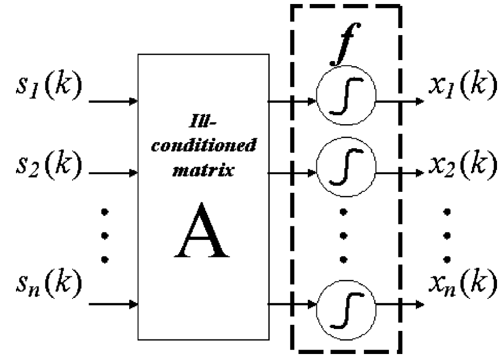

Fig. 1. Block diagram of the ill-conditioned post-nonlinear mixing model.

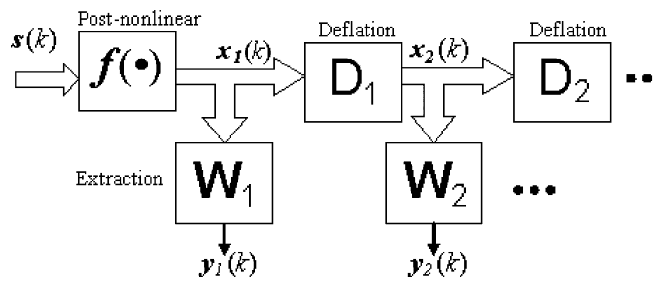

Fig. 2. General structure of the blind source extraction (BSE).

A. Blind Source Extraction Procedure

Fig. 2 shows a general structure of the BSE process which extracts one single source at a time, where there are two principal stages in this process: extraction and deflation [21]. The mixtures first undergo the extraction stage to have one source recovered; after deflation, the contribution of the extracted source is removed from the mixtures. These new "deflated" mixtures contain linear combinations of the remaining sources; the next extraction process then recovers the second source; this process repeats until the last source of interest is recovered.

Our goal is to extract the sources of interest without any prior knowledge of their distributions and the (possibly ill-conditioned and nonlinear) mixing mechanism. To that cause, we need to derive an extraction structure for which the learning rule involves both the estimation of the single processing unit, $\mathbf{w}_1(\infty) = \mathbf{w}_1$ and a procedure to estimate the nonlinear effects

[2]As a generative model, ICA aims to find the independent components from the mixture of statistically independent sources by optimising different criteria, for review, see [3], [5], [7], [18].

[3]One example of an ill-conditioned matrix can be found in [1]. This matrix can be generated as follows: $\mathbf{A} = R * H * R^{-1}$; where $R(i,j) = sin(theta)$; $R(j,i) = -R(i,j)$; $R = eye(n)$; $R(i,i) = cos(theta)$; $R(j,j) = R(i,i)$. In Matlab, this can be coded as $H = hilb(n), ind = randperm(n), theta = 2 * pi * rand, i = ind(1)$.

of ill-conditioned mixing within. This way, the extraction operation for a single signal, y_1, can be expressed as

$$y_1(k) = \mathbf{w}_1^T \mathbf{x}(k) \tag{2}$$

where $y_1(k)$ denotes the single extracted output signal and $\mathbf{x}_1 = \mathbf{x}$ for the first processing mixtures. The $\mathbf{g} = \Phi(\mathbf{w}_1^T \mathbf{A})$ denotes a global demixing vector from the sources to the outputs, where Φ is a nonlinear function explained later.

III. Proposed Extraction Algorithm

For the extraction of ill-conditioned post-nonlinear mixtures, analogous to "mixed norm" approaches to adaptive filtering [16], we propose the following "mixed contrast function" criterion based on [10], [22]

$$\begin{aligned}\boldsymbol{J}_i(y_i, \mathbf{W}) &= \left\{ \sum_{i=1}^{n} \left| \mathrm{cum}\left[y_1^4\right] \right| \right\} \\ &\quad - \left\{ -\log\left|\det(\mathbf{W})\right| - E\left\{ \log \sum_{i=1}^{n} \left[\tilde{q}_i(y_i)\right| \right\} \right\} \\ &= J_K + J_N \end{aligned} \tag{3}$$

where $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_n]$, $\tilde{q}_i(y_i(k))$, $i = 1, \ldots, n$, are true probability density functions of the source signals [25], J_K corresponds to the first term in the cost function (3) (kurtosis) and J_N to the second term (nonlinearity). The left-hand side part of (3) performs standard BSE, whereas the right hand part of (3) estimates the nonlinearity within the mixing process.

A. Nonlinearity: The Activation Function

Notice that criterion (3) represents a joint constrained optimisation problem. In order to derive a learning algorithm corresponding to (3), we shall consider separately the minimisation of each part of cost function (3). From (3), to extract only signal y_1, we have

$$\begin{aligned}\frac{\partial J_N(\mathbf{W}, y_1)}{\partial \mathbf{w}_1} &= \frac{\partial}{\partial \mathbf{w}_1} \left\{ -\log\left|\det(\mathbf{W})\right| \right. \\ &\qquad \left. -E\left\{ \log \sum_{i=1}^{n} \left[\tilde{q}_1(y_1)\right| \right\} \right\} \\ &= -\mathbf{W}^{-T} + E\left\{ \tilde{f}_1(y_1)\mathbf{x}_1^T \right\} \end{aligned} \tag{4}$$

where $\tilde{f}_1(\cdot)$ is a smooth nonlinearity and

$$\tilde{f}_1(y_1) = -\frac{\partial \log \tilde{q}_1(y_1)}{\partial y_1} = -\frac{\partial \tilde{q}_1(y_1)/\partial y_1}{\tilde{q}_1(y_1)} = -\frac{\tilde{q}_1'(y_1)}{\tilde{q}_1(y_1)}. \tag{5}$$

It is important to note that (3) holds only if the functions $\tilde{f}_1(\cdot)$ are invertible, a restriction that must be taken into account in the development of learning algorithms.

Hence, on the basis of the standard gradient descent, we obtain an approximate learning rule, given by

$$\begin{aligned}\triangle \mathbf{w}_1(k) &= -\eta_0 \frac{\partial J_N}{\partial \mathbf{w}_1} \\ &= \eta_0 \left[\Lambda(k) - \left\langle \tilde{f}_1(y_1)\left[\hat{R}_1^T(y_1)\right] \right\rangle \right] \mathbf{w}_1(k) \end{aligned} \tag{6}$$

where $\Lambda = D_1$, $\hat{R}_1(y_1) = y_1^T D_1$, $\langle \cdot \rangle$ denotes the expectation operator and η_0 the learning rate.

In a special case, for symmetric pdf distributions of sources and odd activation functions $\tilde{f}_1(y_1)$ and

$$D_1 = \mathrm{diag}\left\{ \langle |y_1| \rangle, \langle |y_2| \rangle, \ldots, \langle |y_n| \rangle \right\} \tag{7}$$

Therefore, we can obtain the median learning rule

$$\triangle \mathbf{w}_1(k) = \eta_0 \left[\Lambda(k) - \left\langle \tilde{f}_1(y_1)\left[\mathrm{sgn}(y_1)\right]^T \right\rangle \right] \mathbf{w}_1(k) \tag{8}$$

where

$$\mathrm{sgn}(y_1) \approx \tanh(y_1) = \frac{e^{y_1} - e^{-y_1}}{e^{y_1} + e^{-y_1}}. \tag{9}$$

Results in other areas show that such a median learning rule with the sgn activation function is robust to additive noise and nonlinearities [12].

B. Normalized Kurtosis-Based Cost Function

A classical measure of nonGaussianity is the kurtosis, which for zero-mean random variable $y_1(k)$ is defined as in [5]. We can represent the term $\mathrm{cum}[y_1^4(k)]$ in (3) as[4]

$$\mathrm{cum}\left[y_1^4(k)\right] = \mathrm{kurt}\left(y_1(k)\right) = E\left\{y_1^4(k)\right\} - 3\left(E\left\{y_1^2(k)\right\}\right)^2. \tag{10}$$

The normalized kurtosis, K_{norm} [19] is then obtained when the kurtosis $\mathrm{kurt}(y_1(k))$ is divided by the square of the variance $E\{y_1^2(k)\}$, to give

$$K_{\mathrm{norm}} = \frac{E\left\{|y_1|^4(k)\right\}}{E^2\left\{|y_1|^2(k)\right\}} - 3 \tag{11}$$

As a cost function for kurtosis based BSE, we may employ

$$J_K\left(\mathbf{W}(k)\right) = -\frac{1}{4}\left|\left(E\left\{y_1^2(k)\right\}\right)^2\right| = -\frac{\beta}{4}\left|\left(E\left\{y_1^2(k)\right\}\right)^2\right| \tag{12}$$

where the parameter β determines the sign of the kurtosis of the signal, within

$$\beta = \begin{cases} -1, & \text{for source signal with negative kurtosis,} \\ +1, & \text{for source signal with positive kurtosis.} \end{cases} \tag{13}$$

Applying standard gradient descent to minimize the cost function, we obtain

$$\begin{aligned}\triangle \mathbf{w}_1(k) &= -\eta_0 \frac{\partial J_K\left(\mathbf{W}(k)\right)}{\partial \mathbf{w}_1(k)} \\ &= \eta_0 \beta \frac{m_4\left(y_1(k)\right)}{m_2^3\left(y_1(k)\right)} \left[\frac{m_2\left(y_1(k)\right)}{m_4\left(y_1(k)\right)} E\left\{y_1^3(k)\mathbf{x}_1(k)\right\} \right. \\ &\qquad \left. -E\left\{y_1(k)\mathbf{x}_1(k)\right\} \right] \end{aligned} \tag{14}$$

[4]For a zero mean variable y_1, the first four univariate cumulants are thus defined as:

$\mathrm{kurt}_1(y_1) \equiv \mathrm{cum}(y_1) = E\{y_1\} = 0$, (mean);

$\mathrm{kurt}_2(y_1) \equiv \mathrm{cum}(y_1, y_1) = var(y_1) = E\{y_1^2\}$, (variance);

$\mathrm{kurt}_3(y_1) \equiv \mathrm{cum}(y_1, y_1, y_1) = E\{y_1^3\}$, (skewness);

$\mathrm{kurt}_4(y_1) \equiv \mathrm{cum}(y_1, y_1, y_1, y_1) = E\{y_1^4\} - 3E\{y_1^2\}^2$, (kurtosis).

where $\eta_0 > 0$. The term $E\{|y_1(k)|^4\}/E^3\{|y_1(k)|^2\} = m_4(y_1(k))/m_2^3(y_1(k))$ is always positive, and can be absorbed by the learning rate $\widetilde{\eta}_0 = (m_4(y_1(k))/m_2^3(y_1(k)))\eta_0 > 0$.

The moments $m_q(y_1(k)) = E\{y_{1q}(k)\}$, for $q \in \{2, 4\}$, can be estimated online as

$$\triangle m_q(y_1(k)) = \eta_0 [y_{1q}(k) - m_q(y_1(k))], \quad q \in \{2, 4\}. \tag{15}$$

Applying subsequently a stochastic approximation, we obtain an online learning rule

$$\triangle \mathbf{w}_1(k) = \eta_0 \varphi_1(y_1(k)) \mathbf{x}_1(k) \tag{16}$$

where $\eta_0 > 0$ is a learning rate and

$$\varphi_1(y_1(k)) = \beta \frac{m_4(y_1(k))}{m_2^3(y_1(k))} \left[\frac{m_2(y_1(k))}{m_4(y_1(k))} y_1^3(k) - y_1(k)\right] \tag{17}$$

is the nonlinearity. Since the positive term $m_4(y_1(k))/m_2(y_1(k))$ can be absorbed within the learning rate, we can also use the following approximation of the nonlinearity

$$\varphi_1(y_1(k)) = \beta \left[\frac{m_2(y_1(k))}{m_4(y_1(k))} y_1^3(k) - y_1(k)\right] \tag{18}$$

or

$$\varphi_2(y_1(k)) = \beta \left[\frac{1}{m_4(y_1(k))} y_1^3(k) - \frac{1}{m_2(y_1(k))} y_1(k)\right]. \tag{19}$$

For spiky signals with positive kurtosis (Super-Gaussian signals), the nonlinearity closely approximates a sigmoidal function.

As a special case, applying a simple Euler approximation to (17), update yields the discrete-time learning rule

$$\mathbf{w}_1(k+1) = \mathbf{w}_1(k) + \eta_0 \varphi_1(y_1(k)) \mathbf{x}_1(k) \tag{20}$$

where $\mathbf{x}_1(k)$ is a vector of sensor signals and $\varphi_1(\cdot)$ the nonlinearity.

C. The Proposed Blind Extraction Learning Rule

Finally, combining (5) and (20), our proposed algorithm for BSS of post-nonlinear mixtures becomes

$$\begin{aligned}\mathbf{w}_1(k+1) = \mathbf{w}_1(k) + \eta_0 \{&\varphi_1(y_1(k)) \mathbf{x}_1(k) \\ &- \left\{\left[\Lambda(k) - \left\langle \tilde{f}_1(y_1) [\mathrm{sgn}(y_1)]^T \right\rangle\right] \mathbf{w}_1(k)\right\}\}\end{aligned} \tag{21}$$

where the extracted outputs, $y_1(k) = \mathbf{w}_1^T(k)\mathbf{x}_1(k)$). This concludes the derivation of the adaptive blind source extraction algorithm based on cost function (3).

D. Deflation Learning Rule

After the successful extraction of the first source signal $y_1(k) \approx s_i(k)$, we can apply the deflation procedure which removes previously extracted signals from the mixtures. This procedure may be recursively applied to extract all source signals sequentially. This means, that for i^{th} deflation we require and online linear transformation given by

$$\mathbf{x}_{(i+1)}(k) = \mathbf{x}_i(k) - \tilde{\mathbf{d}}_i \tilde{y}_i(k), \quad (i = 1, 2, \ldots, n) \tag{22}$$

where $\tilde{y}_i = \mathrm{sgn}(y_i)$ and

$$\tilde{\mathbf{d}}_i(k+1) = \tilde{\mathbf{d}}_i(k) + \eta_0 \tilde{y}_i(k) \mathbf{x}_{i+1}^T(k), \quad (i = 1, 2, \ldots, n) \tag{23}$$

where $\tilde{\mathbf{d}}_i$ is an estimation of the i^{th} column of the identified mixing matrix $\mathbf{A}$, $y_i = \mathbf{w}_i^T \mathbf{x}_i$.

The proposed method is outlined below:

Procedure: Blind extraction and deflation of post-nonlinear mixtures

---

For post-nonlinearly mixed signals, $\mathbf{x}(\mathrm{k}) = \boldsymbol{f}(\mathbf{A}\mathbf{s}(\mathrm{k}))$, the single extracted signal is defined as $y_1(k) = \mathbf{w}_1^T(k)\mathbf{x}_1(k)$, where $\mathbf{w}_1(k)$ is randomly initialized:

For $i = 1, 2, \ldots, n$ signal

Follow the criterion

$$\boldsymbol{J}(y_i) = \left\{\sum_{i=1}^{n} \left|\mathrm{cum}\left[y_i^4\right]\right|\right\} - \left\{-\log|\det(\mathbf{W})| - E\left\{\log \sum_{i=1}^{n} [\tilde{q}_i(y_i)]\right\}\right\}$$

For $k = 1$: number of data points

1 Apply the algorithm

$$\triangle \mathbf{w}_1(k) = -\eta_0 \frac{\partial \boldsymbol{J}(\mathbf{W}(k))}{\partial \mathbf{w}_1(k)}$$

2 Perform Adaptive Extraction

$$\begin{aligned}\mathbf{w}_1(k+1) = \mathbf{w}_1(k) + \eta_0 \{&\varphi_1(y_1(k)) \mathbf{x}_1(k) \\ &- \left\{\left[\Lambda(k) - \left\langle \tilde{f}_1(y_1) [\mathrm{sgn}(y_1)]^T \right\rangle\right] \mathbf{w}_1(k)\right\}\},\end{aligned}$$

End extraction for $i = 1$ signal

Repeat for n signals, until all signals extracted

3 Deflation method

$$\mathbf{x}_{(i+1)}(k) = \mathbf{x}_i(k) - \tilde{\mathbf{d}}_i \tilde{y}_i(k);$$

where $\tilde{y}_i = \mathrm{sgn}(\tilde{y}_i)$

End extraction for n signals

IV. Experimental Results

In the experiments, simulations were based on three source signals: s_1 with binary distribution, s_2 with sine waveform and s_3 with Gaussian distribution (Fig. 3). Monte Carlo simulations with 5000 iterations of independent trials were performed. The

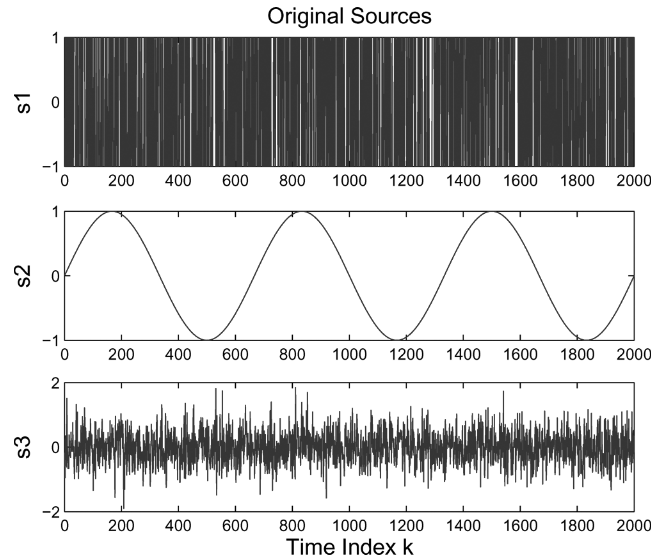

Fig. 3. Original unknown sources. s_1 with binary distribution, s_2 with sine waveform and s_3 with Gaussian distribution.

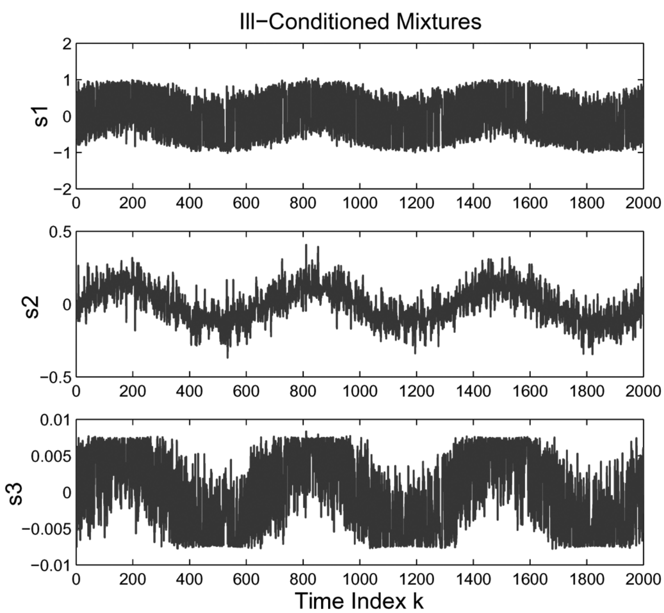

Fig. 4. Three ill-conditioned post-nonlinear mixtures.

initial values of the weights and the demixing matrix $\mathbf{W}(k)$ were randomly generated for each run. The simulations were conducted without prewhitening.

A 3×3 ill-conditioned mixing matrix[5] [1] was randomly generated (based on Fig. 1), the ill-conditioned mixing matrix $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{bmatrix} 0.9089 & -0.2110 & -0.1923 \\ -0.2110 & 0.0547 & 0.0395 \\ -0.1923 & 0.0395 & 0.0853 \end{bmatrix} \tag{24}$$

[5]The condition number of a matrix $\mathbf{A}$ is the quantity $\mathcal{K}(\mathbf{A}) = \|\mathbf{A}\|\|\mathbf{A}^{-1}\|$. It is a measure of the sensitivity of the solution of $\mathbf{As} = \mathbf{b}$ to perturbations of $\mathbf{A}$ or $\mathbf{b}$. If the condition number of $\mathbf{A}$ is 'large', $\mathbf{A}$ is said to be ill-conditioned. If the condition number is unity, $\mathbf{A}$ is said to be perfectly conditioned [15]. If $\mathbf{A}$ is normal then $\mathcal{K}(\mathbf{A}) = \lambda_{\max}(\mathbf{A})/\lambda_{\min}(\mathbf{A})$, where $\lambda_{\max}(\mathbf{A})$, $\lambda_{\min}(\mathbf{A})$ are respectively maximal and minimal (by moduli) eigenvalues of $\mathbf{A}$. If $\|\cdot\|$ is l_2 norm then $\mathcal{K}(\mathbf{A}) = \sigma_{\max}(\mathbf{A})/\sigma_{\min}(\mathbf{A})$, where $\sigma_{\max}(\mathbf{A})$, $\sigma_{\min}(\mathbf{A})$ are respectively the maximal and minimal singular values of $\mathbf{A}$ [11].

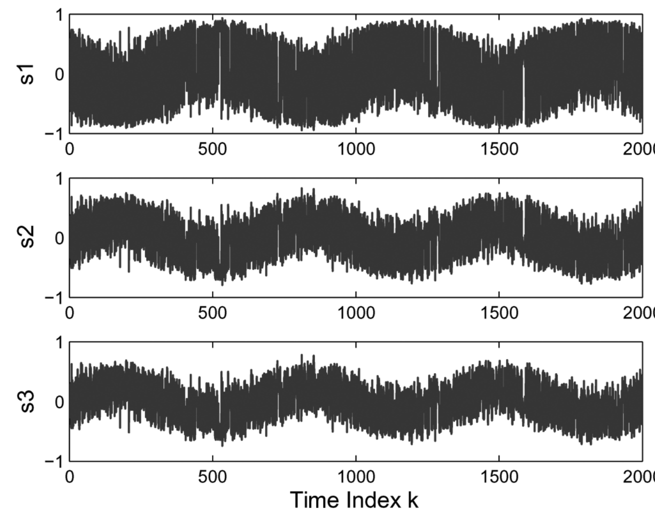

Fig. 5. Extracted signals with binary distribution (top), sine waveform (middle) and Gaussian distribution (bottom) using linear predictor [19].

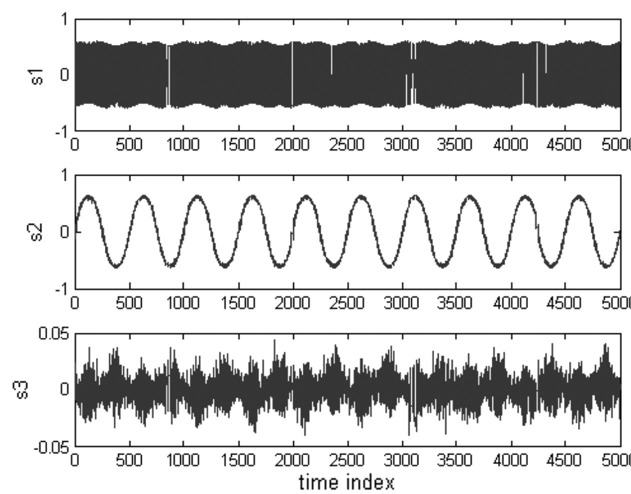

Fig. 6. Extracted signals with binary distribution (top), sine waveform (middle), and Gaussian distribution (bottom) using the proposed nonlinear predictor.

where the condition number, $\mathcal{K}(\mathbf{A}) = \lambda_{\max}/\lambda_{\min} = \sigma_{\max}/\sigma_{\min} = 212.2234$.

If, as a nonlinear function Φ from Section II, we use the sigmoid saturation type function $\tanh$, our ill-conditioned post-nonlinear mixtures (Fig. 4) can be modelled as

$$\mathbf{x}(k) = \mathbf{tanh}\left(\mathbf{As}(k)\right) \tag{25}$$

To measure the quantitative performance of the proposed algorithm, we employ the performance index (PI) defined by [2]

$$\mathrm{PI} = 10\log_{10}\left(\frac{1}{n}\left(\sum_{i=1}^{n}\frac{g_l^2}{\max\{g_1^2, g_2^2, \ldots, g_n^2\}} - 1\right)\right) \tag{26}$$

where $\mathbf{g} = \Phi(\mathbf{w}_1^T\mathbf{A}) = [g_1, g_2, \ldots, g_n]^T$. The smaller the value of PI, the better the quality of extraction.

The measure of qualitative performance were scatter plots, presented in Fig. 7, which show that the proposed method has

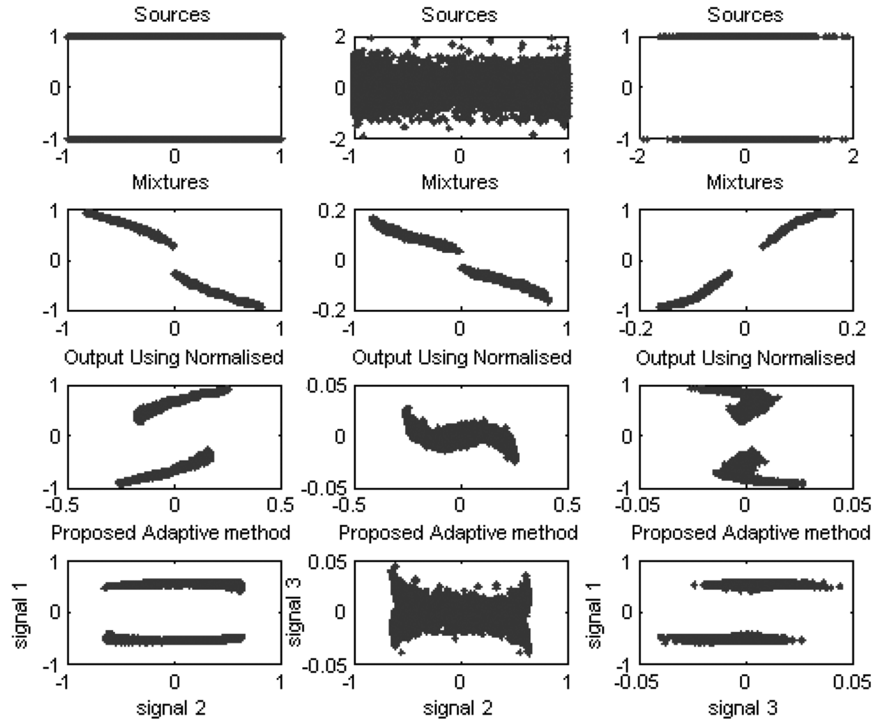

Fig. 7. Scatter plot comparing the independence of the output signals; Column 1: signal 1 and 2; Column 2: signal 2 and 3; Column 3: signal 1 and 3.

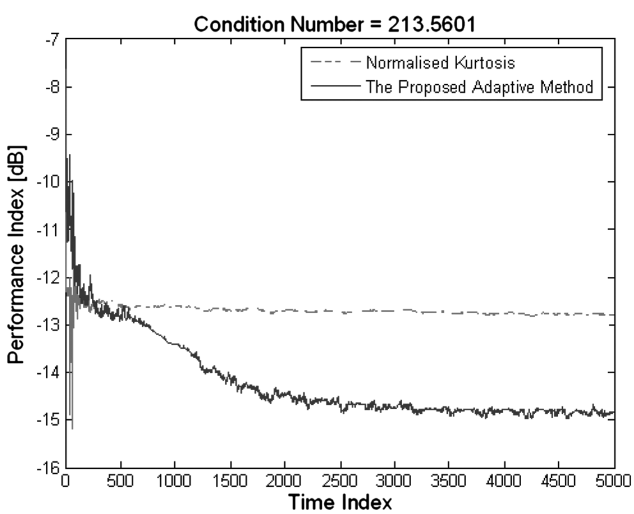

Fig. 8. Learning curve of the extraction algorithms with condition number= 213.5601 (a) - - - normalized kurtosis [19]. (b) —— The proposed adaptive method.

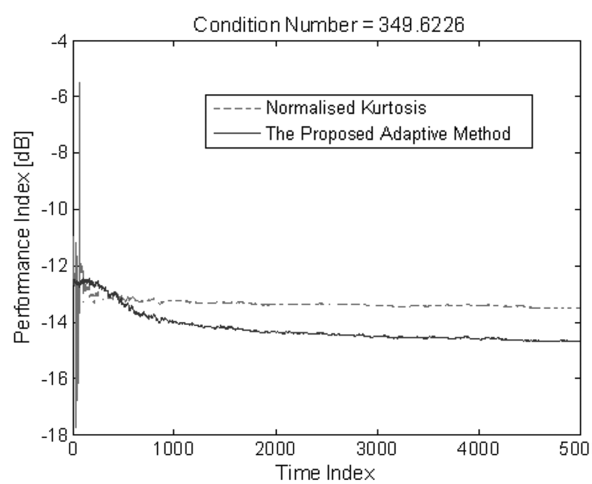

Fig. 9. Learning curve of the extraction algorithms with condition number= 349.6226. (a) - - - normalized kurtosis [19]. (b) —— The proposed adaptive method.

the potential to extract the ill-conditioned post-nonlinearly mixtures (Fig. 6), as indicated by the output scatter plots being closely matched with the original sources (Fig. 7). The proposed adaptive method also exhibits faster convergence and better performance index than the recently introduced state of the art method [19] in Fig. 8 and Fig. 9 with condition number =213.5601 and 450.4487, respectively. Fig. 10 shown the performance index for three different nonlinearities after first and second extraction in condition number=473.8132. The monomial nonlinearity $ay_1|y_1|^{n-1}$ and $asign(y_1)$ (as addressed in [13], a the scaling condition) shown performance index less than -17 dB after the first extraction.

Table I shows the Performance Index of the extracted signals with different condition numbers (1.9247, 38.7087, 190.9155, 213.5601, 363.6029, 349.6226, and 450.4487) using the normalized kurtosis method [19] and the proposed adaptive method. We observed that the proposed adaptive method outperformed the conventional normalized kurtosis method [19], and showed a natural trend, whether the normalized kurtosis method showed very inconsistent performance.

V. Conclusion

We have addressed a special class of BSS algorithms, namely ill-conditioned post-nonlinear BSE, by which we can recover a

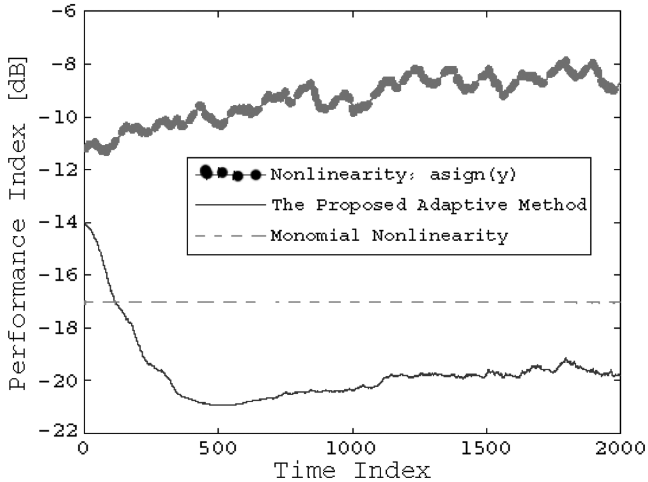

Fig. 10. A learning curve of the extraction algorithms with 3 different nonlinearities as show in [13].

TABLE I
PERFORMANCE INDEX OF THE EXTRACTED SIGNALS WITH DIFFERENT CONDITION NUMBERS USING THE NORMALIZED KURTOSIS METHOD [19] AND THE PROPOSED ADAPTIVE METHOD

Extracted Signal		
Condition Number	Performance Index (dB)	
$\mathcal{K}(\mathbf{A})$	Normalised Kurtosis [19]	Adaptive Method
1.9247	-24.8842	-25.4947
38.7087	-3.7620	-20.4420
190.9155	-15.4139	-16.4149
213.5601	-12.8912	-14.9986
349.6226	-13.5180	-14.6958
450.4487	-4.3396	-5.0357

single source or a subset of sources at a time, instead of recovering all of the sources simultaneously. The proposed adaptive algorithm does not require any prepocessing (prewhitening), and due to the design of the contrast function, it is particularly suitable for sequential blind source extraction with ill-conditioned post-nonlinear mixing matrices. Simulation results have confirmed the validity of the theoretical results and demonstrated the performance of the algorithm.

APPENDIX

By changing γ, the nonlinearity can be varied between a linear device and a hard limiter. The effects of $\gamma \rightarrow 0$ can be studied by scaling y_1 by a constant.

A convenient nonlinearity is a hyperbolic tangent function, given by

$$\Phi[y_1] = \mathrm{sgn}(y_1) \approx \tanh(\gamma y_1) = \frac{e^{\gamma y_1} - e^{-\gamma y_1}}{e^{\gamma y_1} + e^{-\gamma y_1}}. \quad (27)$$

the positive scalar γ is used to modify the shape (slope) of $\Phi(\cdot)$. In such a case

$$\lim_{\gamma^2 \rightarrow \infty} [y_1] = y_1 \quad and \quad \lim_{\gamma^2 \rightarrow 0} [\tilde{y}_1] = \gamma\sqrt{\pi/2}sign(y_1). \quad (28)$$

For sub-Gaussian source signals, the cubic nonlinear function $\Phi[y_1] = y_1^3$ has been a favorite choice. For mixtures of sub- and super-Gaussian source signals, according to the estimated kurtosis of the extracted signals, the nonlinear function can be selected from [10].

REFERENCES

[1] A. Cichocki and D. Erdogmus, MLSP 2006 data competition [Online]. Available: http://mlsp2006.conwiz.dk/, 2006
[2] A. Cichocki and S. I. Amari, "Adaptive Blind Signal and Image Processing," in *Locally Adaptive Algorithms for ICA and Their Implementations*. New York: Wiley , 2002.
[3] A. Hyvarinen, "Survey on independent component analysis," *Neural Comput. Surv.*, vol. 2, pp. 94–128, 1999.
[4] A. Hyvarinen and E. Oja, Independent Component Analysis: A Tutorial. Tech Rep. Helsinki Univ.Technology, Helsinki, Finland, Apr. 1999.
[5] A. Hyvarinen, J. Karhunen, and E. Oja*, Independent Component Analysis*. New York: Wiley, 2001.
[6] A. Hyvarinen and P. Pajunen, "Nonlinear independent component analysis: Existence and uniqueness results," *Neural Netw.*, vol. 12, pp. 429–439, 1999.
[7] A. Mansour, A. Barros, and N. Ohnishi, "Blind separation of sources: Methods, assumptions and applications," *IEICE Trans. Fundamentals Electronics, Commun. Comput. Sci.*, vol. E83-A, pp. 1498–1512, 2000.
[8] A. Taleb, "Source separation in structured nonlinear models," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2001, vol. 6, pp. 3613–3516.
[9] A. Taleb and C. Jutten, "Source separation in post-nonlinear mixtures," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2807–2820, 1999.
[10] A. J. Bell and T. J. Sejnowski*, An Information Maximization Approach to Blind Separation and Blind Deconvolution*, vol. 7, pp. 1129–1159, 1995.
[11] Golub and Van Loan*, Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press.
[12] G. R. Arce*, Nonlinear Signal Processing: A statistical Approach*, first ed. New York: Wiley, 2005.
[13] H. Mathis, "Nonlinear Functions for Blind Separation and Equalization," Ph.D. dissertation, The Swiss Federal Institute of Technology,, Zurich, Nov. 2001.
[14] H. Valpola, X. Giannakopoulos, A. Honkela, and J. Karhunen, "Nonlinear independent component analysis using ensemble learning: experiments and discussion" Tech. Rep., Helsinki Univ. Technology, Neural Networks Research Centre, Helsinki, Finland, 2000.
[15] J. Blackledge*, Digital Signal Processing: Mathematical and Computational Methods, Software Development and Applications*, 2nd ed. London, U.K.: Horwood,, 2006.
[16] J. Chambers and A. Avlonitis*, A Robust Mixed-Norm Adaptive Filter Algorithm*, vol. 4, pp. 46–48, Feb. 1997.
[17] J. Eriksson and V. Koivunen, "Identifiability and separability of linear ICA models revisited," in *Proc. Int. Workshop Independent Component Anal. Blind Signal Separation*, Apr. 2003, pp. 23–27.
[18] J. F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 44, pp. 3017–3030, 1998.
[19] W. Liu and D. P. Mandic, "A normalized kurtosis based algorithm for blind source extraction from noisy measurements," *Signal Process.*, vol. 86, pp. 1580–1585, Jul. 2006.
[20] E. Moreau and O. Macchi, "High-order contrasts for self-adaptive source separation," *Int. J. Adaptive Contr. Signal Process.*, pp. 19–46, 1996.
[21] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: A deflation approach," *Signal Process.*, vol. 49, pp. 59–83, 1995.
[22] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994.
[23] S. A. Cruces-Alvarez, "From blind signal extraction to blind instantaneous signal separation: Criteria, algorithm, and stability," *IEEE Trans. Neural Netw.*, vol. 15, pp. 859–873, Jul. 2004.
[24] S. A. Cruces-Alvarez, A. Cichoki, and S. I. Amari, "On a new blind signal extraction algorithm: Different criteria and stability analysis," *IEEE Signal Process. Lett.*, vol. 9, pp. 233–236, Aug. 2002.
[25] S. Amari, "Natural gradient works efficiently in learning," *Neural Comput.*, vol. 10, pp. 251–276, Jan. 1998.
[26] T. W. Lee, B. Koehler, and R. Orglmeister, "Blind source separation of nonlinear mixing models," *Nueral Netw. Signal Process.*, vol. 7, pp. 406–415, 1997.
[27] W. Y. Leong and D. P. Mandic, "Blind extraction of noisy events using nonlinear predictor," in *Proc. 2007 IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 6–8, 2004, pp. 657–660.

[28] W. Y. Leong and D. P. Mandic, "Adaptive blind extraction for post-nonlinearly mixed signals," in *Proc. 2006 IEEE Int. Workshop Mach. Learning Signal Process.*, Maynooth, Ireland, Sep. 6–8, 2006, pp. 91–96.
[29] W. Y. Leong and J. Homer, "EKENS: A learning on nonlinear blindly mixed signals," in *Proc. 30th Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP'05)*, Mar. 19–23, 2005, vol. 4, pp. 81–84.
[30] X. L. Li and X. D. Zhang, "Sequemtial blind extraction adopting second-order statistics," *IEEE Signal Process. Lett.*, vol. 14, pp. 58–61, 2007.
[31] Y. Q. Li and J. Wang, "Sequential blind extraction of instananeously mixed sources," *IEEE Trans. Signal Process.*, vol. 50, pp. 997–1006, May 2002.
[32] Y. Q. Li, J. Wang, and J. M. Zurada, "Blind extraction of singularly mixed source signals," *IEEE Trans. Neural Netw.*, vol. 11, no. 6, pp. 1413–1422, Nov. 2000.
[33] Z. L. Zhang and Z. Yi, "Robust extraction of specific signals with temporal structure," *Neurocomputing*, vol. 69, pp. 888–893, 2006.

Wai Yie Leong received the B.S. and the Ph.D. degrees from The University of Queensland, Australia, in 2002 and 2006, respectively, both in electrical engineering.

In 1999, she was a System Engineer at the Liong Brothers Poultry Farm. From 2002 to 2005, she was appointed as Research Assistant and Teaching Fellow of the School of Information Technology and Electrical Engineering, The University of Queensland, Australia. She is also a Teaching Fellow of St. John's College, Australia. In 2005, she joined the School of Electronics and Electrical Engineering, Imperial College London, U.K. as a Postdoctoral Research Fellow. Between her B.Eng. and Ph.D studies, she has been actively involving in research commercialization. She is now the Research Engineer of A*STAR Corporate, Singapore, and the Head of the Sensing and Conditioning Laboratory. Her research interests include blind source separation, blind extraction, smart sensor, wireless communication systems, smart antennas and biomedical engineering.

Danilo P. Mandic (SM'97) is a Reader in Signal Processing, Imperial College London. He has been working in the area of nonlinear adaptive signal processing and nonlinear dynamics. His publication record includes two research monographs Recurrent Neural Networks for Prediction, and Complex Valued Nonlinear Adaptive Filters with Wiley, an edited book on Signal Processing for Information Fusion (Springer 2007), and more than 200 publications in Signal and Image Processing. He has been a Guest Professor in KU Leuven Belgium, TUAT Tokyo, Japan, and Westminster University, U.K., and Frontier Researcher in Riken, Japan.

Dr. Mandic has been a Member of the IEEE Technical Committee on Machine Learning for Signal Processing, Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II, IEEE TRANSACTIONS ON SIGNAL PROCESSING, and International Journal of Mathematical Modelling and Algorithms. He has produced award winning papers and products resulting from his collaboration with Industry. He is a Member of the London Mathematical Society.