# Natural gradient algorithm for cyclostationary sources

M.G. Jafari, J.A. Chambers and D.P. Mandic

A new approach to blind source separation of cyclostationary sources is introduced which incorporates a cyclic pre-whitening operation within the learning rule, and thereby provides a new member of the family of natural gradient algorithms. The technique improves the convergence properties of the natural gradient algorithm for complex valued, cyclostationary signals. Simulations show the improved convergence speed of the approach.

*Introduction:* The aim of blind source separation (BSS) is to extract mutually statistically independent source signals from observations containing only mixtures of these signals [1]. Apart from the independence hypothesis, it is typically assumed that the sources have unit variance owing to the scale ambiguity inherent to BSS. In this Letter, we further assume that the source signals are cyclostationary. The statistical properties of cyclostationary signals vary periodically with time. This periodicity may arise as a result of amplitude, phase or frequency modulation in manmade signals, while it may be due to natural variations in unprocessed data [2]. The complex linear instantaneous mixing operation at the heart of the BSS problem is modelled by $\mathbf{x}(k)=\mathbf{A}\mathbf{s}(k)$, where $\mathbf{s}(k)\in\mathbb{C}^m$ and $\mathbf{x}(k)\in\mathbb{C}^n$ represent, respectively, the complex source and observed signals, $\mathbf{A}\in\mathbb{C}^{n\times m}$ is an unknown, full column rank, complex mixing matrix, and $k$ denotes the discrete time index. The sources are then recovered with the separating system $\mathbf{y}(k)=\mathbf{W}(k)\mathbf{x}(k)$, where $\mathbf{y}(k)\in\mathbb{C}^m$ represents the complex extracted sources, and $\mathbf{W}(k)\in\mathbb{C}^{m\times n}$ is the separating matrix. In the sequel, we assume that $m=n$.

*Cyclostationarity:* The source signals are modelled as wide sense cyclostationary signals, i.e. their mean and autocorrelation function are periodic in time. Moreover, as in [3], we assume that the cyclic correlation function $r_{pq}(\tau)=E\{e^{j\beta_p k}s_p(k+\tau)s_q^*(k)\}$ of the sources has the following properties:

$$E\{e^{j\beta_p k}s_p(k+\tau)s_q^*(k)\}=0, \quad \text{if } p\neq q \tag{1}$$

$$E\{e^{j\beta_q k}s_p(k+\tau)s_p^*(k)\}=0, \quad \text{if } \beta_p\neq\beta_q \tag{2}$$

$$E\{e^{j\beta_p k}s_p(k)s_q^*(k)\}>0, \quad \forall p \tag{3}$$

where * denotes complex conjugation, and $\beta_p$ is a nonzero cycle frequency of source $p$. It follows that the cyclic correlation matrix of the sources at lag $\tau=0$, defined as $\mathbf{R}_{\mathbf{s}}^{\beta_p}(0)=E\{e^{j\beta_p k}\mathbf{s}(k)\mathbf{s}^H(k)\}$ where $(\cdot)^H$ denotes the Hermitian transpose operator, becomes

$$\mathbf{R}_{\mathbf{s}}^{\beta_p}(0)=\mathbf{I}' \tag{4}$$

where the elements of $\mathbf{I}'$, $[\mathbf{I}']_{l,g}$ are defined by

$$[\mathbf{I}']_{l,g}=\begin{cases}1, & \text{if } l\in\{1,2,\ldots,m\}, g=l=p\\ 0, & \text{otherwise}\end{cases} \tag{5}$$

In the BSS context, the second-order cyclostationary statistics of the source signals can be exploited by seeking that the output cyclic correlation matrix is of the form (4), i.e. the outputs are required to be decorrelated in the cyclostationary sense. Since BSS effectively amounts to a two-stage process in which the data are sphered and then rotated [4], we replace the classic whitening stage by cyclostationary pre-whitening. This entails finding a vector $\mathbf{z}(k)=\mathbf{V}(k)\mathbf{x}(k)$, such that

$$\lim_{k\to\infty}E\{e^{j\beta_p k}\mathbf{z}(k)\mathbf{z}^H(k)\}=\mathbf{I}'$$

$$\text{or}\quad \lim_{k\to\infty}\left(E\{e^{j\beta_p k}\mathbf{z}(k)\mathbf{z}^H(k)\}+\mathbf{J}E\{e^{j\beta_p k}\mathbf{z}(k)\mathbf{z}^H(k)\}\mathbf{J}^H\right)=\mathbf{I} \tag{6}$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{J}$ is the real permutation matrix with elements

$$[\mathbf{J}]_{l,g}=\begin{cases}1, & \text{if } l\in\{1,2,\ldots,m\}, g=m-l+1\\ 0, & \text{otherwise}\end{cases}$$

and $E\{e^{j\beta_p k}\mathbf{z}(k)\mathbf{z}^H(k)\}=\hat{\mathbf{R}}_z^{\beta_p}$ is the cyclic correlation matrix of the signals $\mathbf{z}(k)$, estimated at sample $k$. The pre-whitening learning rule in [4] is then modified to give a cyclic pre-whitening learning rule of the form

$$\mathbf{V}(k+1)=\mathbf{V}(k)+\mu(\mathbf{I}-\hat{\mathbf{R}}_z^{\beta_p}-\mathbf{J}\hat{\mathbf{R}}_z^{\beta_p}\mathbf{J}^H)\mathbf{V}(k) \tag{7}$$

where $\mu$ is a step-size parameter. The inclusion of the term $\mathbf{J}\hat{\mathbf{R}}_z^{\beta_p}\mathbf{J}^H$ in (7) ensures balance in the algorithm so that convergence is attained for all possible choices of cycle frequency. Following the approach in [4], the update (7) corresponds to the adaption of a matrix $\mathbf{V}(k)$ such that it converges to a point where $\tilde{\mathbf{R}}_z^{\beta_p}=\mathbf{I}$ with

$$\tilde{\mathbf{R}}_z^{\beta_p}=\hat{\mathbf{R}}_z^{\beta_p}+\mathbf{J}\hat{\mathbf{R}}_z^{\beta_p}\mathbf{J}^H \tag{8}$$

Effectively, (7) minimises the cost function

$$KL(\tilde{\mathbf{R}}_z^{\beta_p})=\frac{1}{2}\left(Trace(\tilde{\mathbf{R}}_z^{\beta_p})-\log\det(\tilde{\mathbf{R}}_z^{\beta_p})-m\right) \tag{9}$$

which represents the Kullback-Liebler divergence between two Gaussian distributions with covariance matrices $\tilde{\mathbf{R}}_z^{\beta_p}$ and $\mathbf{I}$. Employing the natural gradient descent method, the gradient of (9) is given by [4]

$$\Delta KL(\tilde{\mathbf{R}}_z^{\beta_p})=\tilde{\mathbf{R}}_z^{\beta_p}-\mathbf{I} \tag{10}$$

which, from (8), becomes

$$\Delta KL(\tilde{\mathbf{R}}_z^{\beta_p})=\hat{\mathbf{R}}_z^{\beta_p}+\mathbf{J}\hat{\mathbf{R}}_z^{\beta_p}\mathbf{J}^H-\mathbf{I} \tag{11}$$

leading to the learning rule (7).

*Complex cyclostationary NGA:* The natural gradient algorithm (NGA) update equation

$$\mathbf{W}(k+1)=\mathbf{W}(k)+\eta[\mathbf{I}-\mathbf{f}(\mathbf{y}(k))\mathbf{y}^T(k)]\mathbf{W}(k) \tag{12}$$

where $\eta$ is the step-size parameter, is readily extended to the complex case by selecting an appropriate phase preserving complex activation function [5] $\mathbf{f}(\mathbf{y}(k))$, and replacing the transpose operator by the Hermitian transpose operator. The learning rule (12) is then extended to include explicitly the cyclostationary whitening operation, which yields

$$\mathbf{W}(k+1)=\mathbf{W}(k)+\eta\big[(\mathbf{I}-\mathbf{f}(\mathbf{y}(k))\mathbf{y}^H(k)) + (\mathbf{I}-\hat{\mathbf{R}}_y^{\beta_p}-\mathbf{J}\hat{\mathbf{R}}_y^{\beta_p}\mathbf{J}^H)\big]\mathbf{W}(k) \tag{13}$$

where $\hat{\mathbf{R}}_y^{\beta_p}$ is the output cyclic correlation matrix estimated at the current iteration using an exponentially weighted average of the instantaneous statistics

$$\hat{\mathbf{R}}_y^{\beta_p}(k+1)=(1-\lambda)\hat{\mathbf{R}}_y^{\beta_p}(k)+\lambda[e^{j\beta_p k}\mathbf{y}(k)\mathbf{y}^H(k)] \tag{14}$$

A straightforward extension from the real to the complex case is to employ so-called split-complex nonlinearities, such as $f_i(y_i(k))=\tanh(y_{iR}(k))+j\tanh(y_{iI}(k))$, where $y_{iR}(k)$ and $y_{iI}(k)$ denote, respectively, the real and imaginary parts of $y_i(k)$, when the sources are super-Gaussian, or as $f_i(y_i(k))=y_i(k)|y_i|^2$ [4] for the sub-Gaussian case.

*Simulations:* Two sinusoids of normalised frequency $(20\pi)^{-1}$ and $(4\pi)^{-1}$ are mixed by a real stationary channel and separated using NGA and the cyclostationary NGA approach. Since the sources and mixing are real, the exponential function in (14) simplifies to a cosine function, and $\beta$, $\eta$ and $\lambda$ are chosen, respectively, as $2(4\pi)^{-1}$, 0.005, and 0.05. The performance index (PI), as conventionally used to assess BSS algorithms [6], resulting from the application of the two methods, and averaged over 50 trials, is shown in Fig. 1, while the sources extracted with a single realisation of both techniques, together with the original signals, are plotted in Fig. 2. The results illustrate that exploitation of the cyclostationary nature of the signals leads to faster convergence rate. In particular, the cyclostationary NGA algorithm converges to a PI of 0.01 within approximately 200 samples, whereas NGA requires 400 samples. Finally, a first-order autoregressive Gaussian process with coefficient $a=0.95\exp(j/4\pi)$, amplitude modulated by $\cos(1/20\pi)$, and a complex exponential of normalised frequency $0.175/2\pi$ are mixed by a real, instantaneous, time-invariant mixing channel. Separation, carried out with complex NGA and the complex cyclostationary NGA method, with $\beta$, $\eta$ and $\lambda$ equal to $0.35/2\pi$, 0.005, and 0.05, results in the PI shown in Fig. 3. It is

evident that the NGA algorithm converges to a PI of 0.01 after 3000 samples, while the cyclostationary NGA approach needs only 1500 samples for convergence. It is noteworthy that convergence is smoother in the complex case owing to the interaction between terms within the algorithm.

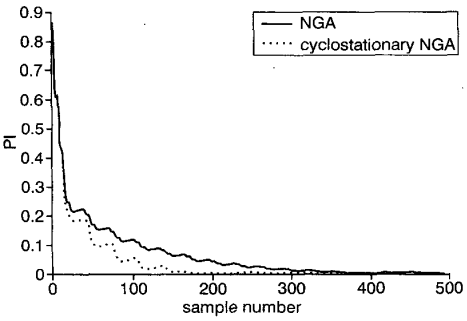


**Fig. 1** *Average performance indices obtained with NGA and cyclostationary NGA approach, when source signals and mixing matrix are real*

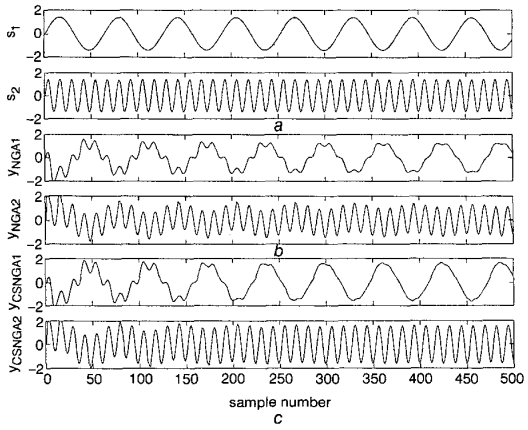


**Fig. 2** *Sources extracted with single realisation of both techniques, together with original signals*

*a* Original sinusoids
*b* Sources extracted with NGA
*c* Sources extracted with cyclostationary NGA method

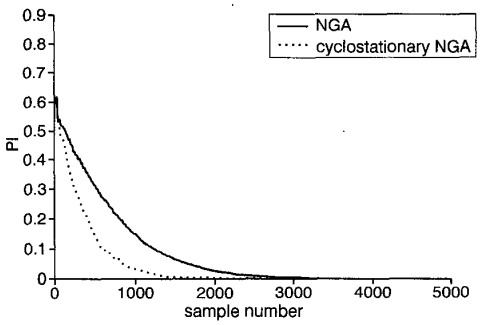


**Fig. 3** *Average performance indices obtained with complex NGA and complex cyclostationary NGA approach, when source signals are complex valued, and mixing channel is real*

*Conclusions:* An NGA-type algorithm for the separation of complex valued cyclostationary data is proposed. By exploiting the cyclostationary nature of the source signals, the technique improves the performance of the natural gradient algorithm. In particular, simulation results show that improved convergence rate is achieved when the sources are extracted by the proposed approach. Exploitation of a number of cyclic autocorrelation matrices with various lags and more comprehensive analysis are the subject of ongoing work.


*15 March 2002*
*Electronics Letters Online No: 20020503*
*DOI: 10.1049/el:20020503*

M.G. Jafari and J.A. Chambers (*Centre for Digital Signal Processing Research, King's College, London, United Kingdom*)

E-mail: maria.jafari@kcl.ac.uk

D.P. Mandic (*Communications and Signal Processing Research Group, Imperial College, London, United Kingdom*)

**References**

1 AMARI, S., and CICHOCKI, A.: 'Adaptive blind signal processing – neural network approaches', *Proc. IEEE*, 1998, **86**, pp. 2026–2048
2 GARDNER, W.: 'Exploitation of the spectral redundancy in cyclostationary signals', *IEEE Signal Process. Mag.*, 1991, **8**, pp. 14–37
3 ABED-MERAIM, K., XIANG, Y., MANTON, J.H., and HUA, Y.: 'Blind source separation using second-order cyclostationary statistics', *IEEE Trans. Signal Process.*, 2001, **49**, pp. 694–701
4 CARDOSO, J., and LAHELD, B.: 'Equivariant adaptive source separation', *IEEE Trans. Signal Process.*, 1996, **44**, pp. 3017–3030
5 PARK, H.M., JUNG, H.Y., LEE, T.W., and LEE, S.Y.: 'On subband-based blind signal separation for noisy speech recognition', *Electron. Lett.*, 1999, **35**, pp. 2011–2012
6 CICHOCKI, A., ORSIER, B., BACK, A., and AMARI, S.: 'On-line adaptive algorithms in non-stationary environments using a modified conjugate gradient approach'. Proc. IEEE Workshop Neural Networks for Signal Processing, Amelia Island, FL, USA, 1997, pp. 316–325

## Wavelet domain natural gradient algorithm for blind source separation of non-stationary sources

M.G. Jafari and J.A. Chambers

A wavelet domain approach to blind source separation which avoids the permutation problem in transfer domain operation, and yields improved convergence rate for certain non-stationary signals, is introduced. The wavelet transform also facilitates noise reduction, further improving performance of the natural gradient algorithm.

*Introduction:* The instantaneous blind source separation (BSS) problem entails extracting a number of mutually statistically independent source signals $\mathbf{s}(k)$ from observations $\mathbf{x}(k)$ which contain mixtures of these signals, represented by $\mathbf{x}(k)=\mathbf{A}\mathbf{s}(k)$, where $\mathbf{A}$ is an unknown, full column rank, mixing matrix. Sources are then recovered with separating system $\mathbf{y}(k)=\mathbf{W}(k)\mathbf{x}(k)$, where $\mathbf{W}(k)$ is the separating matrix. The natural gradient algorithm (NGA) is a BSS technique with update equation

$$\mathbf{W}(k+1)=\mathbf{W}(k)+\eta(k)[\mathbf{I}-\mathbf{f}(\mathbf{y}(k))\mathbf{y}^T(k)]\mathbf{W}(k) \quad (1)$$

where $\mathbf{f}(\mathbf{y}(k))$ is the activation function. In general, performance of NGA degrades as probability density functions (PDFs) of the sources become more Gaussian. However, mapping certain non-stationary signals from the time to the wavelet domain results in their sample PDFs being further from a normal distribution. Hence, we address the BSS problem in the wavelet domain, and show the proposed method results in faster convergence speed than conventional NGA.

*Wavelet domain NGA:* The wavelet transform (WT)-NGA approach uses NGA to separate sources in the wavelet domain. The mixtures are divided into blocks of length $N$. When the measurement length is $M$ samples, assumed to be a multiple of $N$, the $i$th block includes samples $\mathbf{x}(iN+1), \ldots, \mathbf{x}((i+1)N)$, where $i=0, \ldots, (M/N)-1$. The wavelet transform of each data block is then evaluated, and we use concatenated coefficients resulting from a single block, thus avoiding the permutation problem in transform domain BSS. The noise level is reduced by applying thresholding, and sources are separated on-line with NGA, giving a vector of estimated signals, the inverse WT of which represents the recovered sources.

*Convergence:* Working in the time-scale domain has the advantage that certain signals are less Gaussian than in the time domain [1]. Moreover, the statistics of the wavelet coefficients can be modelled by