ARTICLE Communicated by Steven Nowlan and Erkki Oja

Natural Gradient Works Efficiently in Learning

Shun-ichi Amari
RIKEN Frontier Research Program, Saitama 351-01, Japan

When a parameter space has a certain underlying structure, the ordinary gradient of a function does not represent its steepest direction, but the natural gradient does. Information geometry is used for calculating the natural gradients in the parameter space of perceptrons, the space of matrices (for blind source separation), and the space of linear dynamical systems (for blind source deconvolution). The dynamical behavior of natural gradient online learning is analyzed and is proved to be Fisher efficient, implying that it has asymptotically the same performance as the optimal batch estimation of parameters. This suggests that the plateau phenomenon, which appears in the backpropagation learning algorithm of multilayer perceptrons, might disappear or might not be so serious when the natural gradient is used. An adaptive method of updating the learning rate is proposed and analyzed.

1 Introduction

The stochastic gradient method (Widrow, 1963; Amari, 1967; Tsypkin, 1973; Rumelhart, Hinton, & Williams, 1986) is a popular learning method in the general nonlinear optimization framework. The parameter space is not Euclidean but has a Riemannian metric structure in many cases. In these cases, the ordinary gradient does not give the steepest direction of a target function; rather, the steepest direction is given by the natural (or contravariant) gradient. The Riemannian metric structures are introduced by means of information geometry (Amari, 1985; Murray and Rice, 1993; Amari, 1997a; Amari, Kurata, & Nagoska, 1992). This article gives the natural gradients explicitly in the case of the space of perceptrons for neural learning, the space of matrices for blind source separation, and the space of linear dynamical systems for blind multichannel source deconvolution. This is an extended version of an earlier article (Amari, 1996), including new results.

How good is natural gradient learning compared to conventional gradient learning? The asymptotic behavior of online natural gradient learning is studied for this purpose. Training examples can be used only once in online learning when they appear. Therefore, the asymptotic performance of online learning cannot be better than the optimal batch procedure where all the examples can be reused again and again. However, we prove that natural gradient online learning gives the Fisher-efficient estimator in the sense

of asymptotic statistics when the loss function is differentiable, so that it is asymptotically equivalent to the optimal batch procedure (see also Amari, 1995; Opper, 1996). When the loss function is nondifferentiable, the accuracy of asymptotic online learning is worse than batch learning by a factor of 2 (see, for example, Van den Broeck & Reimann, 1996). It was shown in Amari et al. (1992) that the dynamic behavior of natural gradient in the Boltzmann machine is excellent.

It is not easy to calculate the natural gradient explicitly in multilayer perceptrons. However, a preliminary analysis (Yang & Amari, 1997), by using a simple model, shows that the performance of natural gradient learning is remarkably good, and it is sometimes free from being trapped in plateaus, which give rise to slow convergence of the backpropagation learning method (Saad & Solla, 1995). This suggests that the Riemannian structure might eliminate such plateaus or might make them not so serious.

Online learning is flexible, because it can track slow fluctuations of the target. Such online dynamics were first analyzed in Amari (1967) and then by many researchers recently. Sompolinsky, Barkai, and Seung (1995), and Barkai, Seung, and Sompolinsky (1995) proposed an adaptive method of adjusting the learning rate (see also Amari, 1967). We generalize their idea and evaluate its performance based on the Riemannian metric of errors.

The article is organized as follows. The natural gradient is defined in section 2. Section 3 formulates the natural gradient in various problems of stochastic descent learning. Section 4 gives the statistical analysis of efficiency of online learning, and section 5 is devoted to the problem of adaptive changes in the learning rate. Calculations of the Riemannian metric and explicit forms of the natural gradients are given in sections 6, 7, and 8.

2 Natural Gradient

Let $S = \{\boldsymbol{w} \in \boldsymbol{R}^n\}$ be a parameter space on which a function $L(\boldsymbol{w})$ is defined. When S is a Euclidean space with an orthonormal coordinate system $\boldsymbol{w}$, the squared length of a small incremental vector $d\boldsymbol{w}$ connecting $\boldsymbol{w}$ and $\boldsymbol{w} + d\boldsymbol{w}$ is given by

$$|d\boldsymbol{w}|^2 = \sum_{i=1}^{n} (dw_i)^2,$$

where dw_i are the components of $d\boldsymbol{w}$. However, when the coordinate system is nonorthonormal, the squared length is given by the quadratic form

$$|d\boldsymbol{w}|^2 = \sum_{i,j} g_{ij}(\boldsymbol{w}) dw_i dw_j. \tag{2.1}$$

When S is a curved manifold, there is no orthonormal linear coordinates, and the length of $d\boldsymbol{w}$ is always written as in equation 2.1. Such a space is

a Riemannian space. We show in later sections that parameter spaces of neural networks have the Riemannian character. The $n \times n$ matrix $G = (g_{ij})$ is called the Riemannian metric tensor, and it depends in general on $\boldsymbol{w}$. It reduces to

$$g_{ij}(\boldsymbol{w}) = \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j \end{cases}$$

in the Euclidean orthonormal case, so that G is the unit matrix I in this case.

The steepest descent direction of a function $L(\boldsymbol{w})$ at $\boldsymbol{w}$ is defined by the vector $d\boldsymbol{w}$ that minimizes $L(\boldsymbol{w} + d\boldsymbol{w})$ where $|d\boldsymbol{w}|$ has a fixed length, that is, under the constraint

$$|d\boldsymbol{w}|^2 = \varepsilon^2 \tag{2.2}$$

for a sufficiently small constant ε.

Theorem 1. *The steepest descent direction of $L(\boldsymbol{w})$ in a Riemannian space is given by*

$$-\tilde{\nabla} L(\boldsymbol{w}) = -G^{-1}(\boldsymbol{w}) \nabla L(\boldsymbol{w}) \tag{2.3}$$

where $G^{-1} = (g^{ij})$ is the inverse of the metric $G = (g_{ij})$ and ∇L is the conventional gradient,

$$\nabla L(\boldsymbol{w}) = \left(\frac{\partial}{\partial w_1} L(\boldsymbol{w}), \ldots, \frac{\partial}{\partial w_n} L(\boldsymbol{w}) \right)^T,$$

the superscript T denoting the transposition.

Proof. We put

$$d\boldsymbol{w} = \varepsilon \boldsymbol{a},$$

and search for the $\boldsymbol{a}$ that minimizes

$$L(\boldsymbol{w} + d\boldsymbol{w}) = L(\boldsymbol{w}) + \varepsilon \nabla L(\boldsymbol{w})^T \boldsymbol{a}$$

under the constraint

$$|\boldsymbol{a}|^2 = \sum g_{ij} a_i a_j = 1.$$

By the Lagrangean method, we have

$$\frac{\partial}{\partial a_i} \{ \nabla L(\boldsymbol{w})^T \boldsymbol{a} - \lambda \boldsymbol{a}^T G \boldsymbol{a} \} = 0.$$

This gives

$$\nabla L(\boldsymbol{w}) = 2\lambda G\boldsymbol{a}$$

or

$$\boldsymbol{a} = \frac{1}{2\lambda} G^{-1}\nabla L(\boldsymbol{w}),$$

where λ is determined from the constraint.

We call

$$\tilde{\nabla} L(\boldsymbol{w}) = G^{-1}\nabla L(\boldsymbol{w})$$

the natural gradient of L in the Riemannian space. Thus, $-\tilde{\nabla}L$ represents the steepest descent direction of L. (If we use the tensorial notation, this is nothing but the contravariant form of $-\nabla L$.) When the space is Euclidean and the coordinate system is orthonormal, we have

$$\tilde{\nabla} L = \nabla L. \tag{2.4}$$

This suggests the natural gradient descent algorithm of the form

$$\boldsymbol{w}_{t+1} = \boldsymbol{w}_t - \eta_t \tilde{\nabla} L(\boldsymbol{w}_t), \tag{2.5}$$

where η_t is the learning rate that determines the step size.

3 Natural Gradient Learning

Let us consider an information source that generates a sequence of independent random variables $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_t, \ldots$, subject to the same probability distribution $q(\boldsymbol{z})$. The random signals $\boldsymbol{z}_t$ are processed by a processor (like a neural network) that has a set of adjustable parameters $\boldsymbol{w}$. Let $l(\boldsymbol{z}, \boldsymbol{w})$ be a loss function when signal $\boldsymbol{z}$ is processed by the processor whose parameter is $\boldsymbol{w}$. Then the risk function or the average loss is

$$L(\boldsymbol{w}) = E[l(\boldsymbol{z}, \boldsymbol{w})], \tag{3.1}$$

where E denotes the expectation with respect to $\boldsymbol{z}$. Learning is a procedure to search for the optimal $\boldsymbol{w}^*$ that minimizes $L(\boldsymbol{w})$.

The stochastic gradient descent learning method can be formulated in general as

$$\boldsymbol{w}_{t+1} = \boldsymbol{w}_t - \eta_t C(\boldsymbol{w}_t)\nabla l(\boldsymbol{z}_t, \boldsymbol{w}_t), \tag{3.2}$$

where η_t is a learning rate that may depend on t and $C(\boldsymbol{w})$ is a suitably chosen positive definite matrix (see Amari, 1967). In the natural gradient online learning method, it is proposed to put $C(\boldsymbol{w})$ equal to $G^{-1}(\boldsymbol{w})$ when the Riemannian structure is defined. We give a number of examples to be studied in more detail.

3.1 Statistical Estimation of Probability Density Function. In the case of statistical estimation, we assume a statistical model $\{p(\boldsymbol{z}, \boldsymbol{w})\}$, and the problem is to obtain the probability distribution $p(\boldsymbol{z}, \hat{\boldsymbol{w}})$ that approximates the unknown density function $q(\boldsymbol{z})$ in the best way—that is, to estimate the true $\boldsymbol{w}$ or to obtain the optimal approximation $\boldsymbol{w}$ from the observed data. A typical loss function is

$$l(\boldsymbol{z}, \boldsymbol{w}) = -\log p(\boldsymbol{z}, \boldsymbol{w}). \tag{3.3}$$

The expected loss is then given by

$$\begin{aligned} L(\boldsymbol{w}) &= -E[\log p(\boldsymbol{z}, \boldsymbol{w})] \\ &= E_q\left[\log \frac{q(\boldsymbol{z})}{p(\boldsymbol{z}, \boldsymbol{w})}\right] + H_Z, \end{aligned}$$

where H_Z is the entropy of $q(\boldsymbol{z})$ not depending on $\boldsymbol{w}$. Hence, minimizing L is equivalent to minimizing the Kullback-Leibler divergence

$$D[q(\boldsymbol{z}) \,:\, p(\boldsymbol{z}, \boldsymbol{w})] = \int q(\boldsymbol{z}) \log \frac{q(\boldsymbol{z})}{p(\boldsymbol{z}, \boldsymbol{w})} d\boldsymbol{z} \tag{3.4}$$

of two probability distributions $q(\boldsymbol{z})$ and $p(\boldsymbol{z}, \boldsymbol{w})$. When the true distribution $q(\boldsymbol{z})$ is written as $q(\boldsymbol{z}) = p(\boldsymbol{z}, \boldsymbol{w}^*)$, this is equivalent to obtain the maximum likelihood estimator $\hat{\boldsymbol{w}}$.

The Riemannian structure of the parameter space of a statistical model is defined by the Fisher information (Rao, 1945; Amari, 1985)

$$g_{ij}(\boldsymbol{w}) = E\left[\frac{\partial \log p(\boldsymbol{x}, \boldsymbol{w})}{\partial w_i} \frac{\partial \log p(\boldsymbol{x}, \boldsymbol{w})}{\partial w_j}\right] \tag{3.5}$$

in the component form. This is the only invariant metric to be given to the statistical model (Chentsov, 1972; Campbell, 1985; Amari, 1985). The learning equation (see equation 3.2) gives a sequential estimator $\hat{\boldsymbol{w}}_t$.

3.2 Multilayer Neural Network. Let us consider a multilayer feedforward neural network specified by a vector parameter $\boldsymbol{w} = (w_1, \ldots, w_n)^T \in \boldsymbol{R}^n$. The parameter $\boldsymbol{w}$ is composed of modifiable connection weights and thresholds. When input $\boldsymbol{x}$ is applied, the network processes it and calculates the outputs $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{w})$. The input $\boldsymbol{x}$ is subject to an unknown probability

distribution $q(\boldsymbol{x})$. Let us consider a teacher network that, by receiving $\boldsymbol{x}$, generates the corresponding output $\boldsymbol{y}$ subject to a conditional probability distribution $q(\boldsymbol{y} \mid \boldsymbol{x})$. The task is to obtain the optimal $\boldsymbol{w}^*$ from examples such that the student network approximates the behavior of the teacher.

Let us denote by $l(\boldsymbol{x}, \boldsymbol{w})$ a loss when input signal $\boldsymbol{x}$ is processed by a network having parameter $\boldsymbol{w}$. A typical loss is given,

$$l(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}) = \frac{1}{2}|\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{w})|^2, \tag{3.6}$$

where $\boldsymbol{y}$ is the output given by the teacher.

Let us consider a statistical model of neural networks such that its output $\boldsymbol{y}$ is given by a noisy version of $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{w})$,

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{w}) + \boldsymbol{n}, \tag{3.7}$$

where $\boldsymbol{n}$ is a multivariate gaussian noise with zero mean and unit covariance matrix I. By putting $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y})$, which is an input-output pair, the model specifies the probability density of $\boldsymbol{z}$ as

$$p(\boldsymbol{z}, \boldsymbol{w}) = cq(\boldsymbol{x}) \exp\left\{-\frac{1}{2}|\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{w})|^2\right\}, \tag{3.8}$$

where c is a normalizing constant and the loss function (see equation 3.6) is rewritten as

$$l(\boldsymbol{z}, \boldsymbol{w}) = \text{const} + \log q(\boldsymbol{x}) - \log p(\boldsymbol{z}, \boldsymbol{w}). \tag{3.9}$$

Given a sequence of examples $(\boldsymbol{x}_1, \boldsymbol{y}_1), \ldots, (\boldsymbol{x}_t, \boldsymbol{y}_t), \ldots$, the natural gradient online learning algorithm is written as

$$\boldsymbol{w}_{t+1} = \boldsymbol{w}_t - \eta_t \tilde{\nabla} l(\boldsymbol{x}_t, \boldsymbol{y}_t, \boldsymbol{w}_t). \tag{3.10}$$

Information geometry (Amari, 1985) shows that the Riemannian structure is given to the parameter space of multilayer networks by the Fisher information matrix,

$$g_{ij}(\boldsymbol{w}) = E\left[\frac{\partial \log p(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{w})}{\partial w_i} \; \frac{\partial p(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{w})}{\partial w_j}\right]. \tag{3.11}$$

We will show how to calculate $G = (g_{ij})$ and its inverse in a later section.

3.3 Blind Separation of Sources. Let us consider m signal sources that produce m independent signals $s_i(t)$, $i = 1, \ldots, m$, at discrete times $t = 1, 2, \ldots$. We assume that $s_i(t)$ are independent at different times and that the

expectations of s_i are 0. Let $r(\boldsymbol{s})$ be the joint probability density function of $\boldsymbol{s}$. Then it is written in the product form

$$r(\boldsymbol{s}) = \prod_{i=1}^{m} r_1(s_1). \tag{3.12}$$

Consider the case where we cannot have direct access to the source signals $\boldsymbol{s}(t)$ but we can observe their m instantaneous mixtures $\boldsymbol{x}(t)$,

$$\boldsymbol{x}(t) = A\boldsymbol{s}(t) \tag{3.13}$$

or

$$x_i(t) = \sum_{j=1}^{m} A_{ij} s_j(t),$$

where $A = (A_{ij})$ is an $m \times m$ nonsingular mixing matrix that does not depend on t, and $\boldsymbol{x} = (x_1, \ldots, x_m)^T$ is the observed mixtures.

Blind source separation is the problem of recovering the original signals $\boldsymbol{s}(t)$, $t = 1, 2, \ldots$ from the observed signals $\boldsymbol{x}(t)$, $t = 1, 2, \ldots$ (Jutten & Hérault, 1991). If we know A, this is trivial, because we have

$$\boldsymbol{s}(t) = A^{-1}\boldsymbol{x}(t).$$

The "blind" implies that we do not know the mixing matrix A and the probability distribution densities $r_i(s_i)$.

A typical algorithm to solve the problem is to transform $\boldsymbol{x}(t)$ into

$$\boldsymbol{y}(t) = W_t\boldsymbol{x}(t), \tag{3.14}$$

where W_t is an estimate of A^{-1}. It is modified by the following learning equation:

$$W_{t+1} = W_t - \eta_t F(\boldsymbol{x}_t, W_t). \tag{3.15}$$

Here, $F(\boldsymbol{x}, W)$ is a special matrix function satisfying

$$E[F(\boldsymbol{x}, W)] = 0 \tag{3.16}$$

for any density functions $r(\boldsymbol{s})$ in equation 3.12 when $W = A^{-1}$. For W_t of equation 3.15 to converge to A^{-1}, equation 3.16 is necessary but not sufficient, because the stability of the equilibrium is not considered here.

Let $K(W)$ be an operator that maps a matrix to a matrix. Then

$$\tilde{F}(\boldsymbol{x}, W) = K(W)F(\boldsymbol{x}, W)$$

satisfies equation 3.16 when F does. The equilibrium of F and $\tilde{F}$ is the same, but their stability can be different. However, the natural gradient does not alter the stability of an equilibrium, because G^{-1} is positive-definite.

Let $l(\boldsymbol{x}, W)$ be a loss function whose expectation

$$L(W) = E[l(\boldsymbol{x}, W)]$$

is the target function minimized at $W = A^{-1}$. A typical function F is obtained by the gradient of l with respect to W,

$$F(\boldsymbol{x}, W) = \nabla l(\boldsymbol{x}, W). \tag{3.17}$$

Such an F is also obtained by heuristic arguments. Amari and Cardoso (in press) gave the complete family of F satisfying equation 3.16 and elucidated the statistical efficiency of related algorithms.

From the statistical point of view, the problem is to estimate $W = A^{-1}$ from observed data $\boldsymbol{x}(1), \ldots, \boldsymbol{x}(t)$. However, the probability density function of $\boldsymbol{x}$ is written as

$$p_X(\boldsymbol{x}; W, r) = |W| r(W\boldsymbol{x}), \tag{3.18}$$

which is specified not only by W to be estimated but also by an unknown function r of the form 3.12. Such a statistical model is said to be semiparametric and is a difficult problem to solve (Bickel, Klassen, Ritov, & Wellner, 1993), because it includes an unknown function of infinite degrees of freedom. However, we can apply the information-geometrical theory of estimating functions (Amari & Kawanabe, 1997) to this problem.

When F is given by the gradient of a loss function (see equation 3.17), where ∇ is the gradient $\partial/\partial W$ with respect to a matrix, the natural gradient is given by

$$\tilde{\nabla} l = G^{-1} \circ \nabla l. \tag{3.19}$$

Here, G is an operator transforming a matrix to a matrix so that it is an $m^2 \times m^2$ matrix. G is the metric given to the space $Gl(m)$ of all the nonsingular $m \times m$ matrices. We give its explicit form in a later section based on the Lie group structure. The inverse of G is also given explicitly. Another important problem is the stability of the equilibrium of the learning dynamics. This has recently been solved by using the Riemannian structure (Amari, Chen, & Chichocki, in press; see also Cardoso & Laheld, 1996). The superefficiency of some algorithms has been also proved in Amari (1997b) under certain conditions.

3.4 Blind Source Deconvolution. When the original signals $\boldsymbol{s}(t)$ are mixed not only instantaneously but also with past signals as well, the prob-

lem is called blind source deconvolution or equalization. By introducing the time delay operator z^{-1},

$$z^{-1}\boldsymbol{s}(t) = \boldsymbol{s}(t-1), \tag{3.20}$$

we have a mixing matrix filter $\boldsymbol{A}$ denoted by

$$\boldsymbol{A}(z) = \sum_{k=0}^{\infty} A_k z^{-k}, \tag{3.21}$$

where A_k are $m \times m$ matrices. The observed mixtures are

$$\boldsymbol{x}(t) = \boldsymbol{A}(z)\boldsymbol{s}(t) = \sum_{k} A_k \boldsymbol{s}(t-k). \tag{3.22}$$

To recover the original independent sources, we use the finite impulse response model

$$\boldsymbol{W}(z) = \sum_{k=0}^{d} W_k z^{-1} \tag{3.23}$$

of degree d. The original signals are recovered by

$$\boldsymbol{y}(t) = \boldsymbol{W}_t(z)\boldsymbol{x}(t), \tag{3.24}$$

where $\boldsymbol{W}_t$ is adaptively modified by

$$\boldsymbol{W}_{t+1}(z) = \boldsymbol{W}_t(z) - \eta_t \nabla l\{\boldsymbol{x}_t, \boldsymbol{x}_{t-1}, \ldots, \boldsymbol{W}_t(z)\}. \tag{3.25}$$

Here, $l(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}, \ldots, \boldsymbol{W})$ is a loss function that includes some past signals. We can summarize the past signals into a current state variable in the online learning algorithm. Such a loss function is obtained by the maximum entropy method (Bell & Sejnowski, 1995), independent component analysis (Comon, 1994), or the statistical likelihood method.

In order to obtain the natural gradient learning algorithm

$$\boldsymbol{W}_{t+1}(z) = \boldsymbol{W}_t(z) - \eta_t \tilde{\nabla} l(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}, \ldots, \boldsymbol{W}_t),$$

we need to define the Riemannian metric in the space of all the matrix filters (multiterminal linear systems). Such a study was initiated by Amari (1987). It is possible to define G and to obtain G^{-1} explicitly (see section 8). A preliminary investigation into the performance of the natural gradient learning algorithm has been undertaken by Douglas, Chichocki, and Amari (1996) and Amari et al. (1997).

4 Natural Gradient Gives Fisher-Efficient Online Learning Algorithms

This section studies the accuracy of natural gradient learning from the statistical point of view. A statistical estimator that gives asymptotically the best result is said to be Fisher efficient. We prove that natural gradient learning attains Fisher efficiency.

Let us consider multilayer perceptrons as an example. We study the case of a realizable teacher, that is, the behavior of the teacher is given by $q(\boldsymbol{y} \mid \boldsymbol{x}) = p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{w}^*)$. Let $D_T = \{(\boldsymbol{x}_1, \boldsymbol{y}_1), \ldots, (\boldsymbol{x}_T, \boldsymbol{y}_T)\}$ be T-independent input-output examples generated by the teacher network having parameter $\boldsymbol{w}^*$. Then, minimizing the log loss,

$$l(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{w}) = -\log p(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{w}),$$

over the training data D_T is to obtain $\hat{\boldsymbol{w}}_T$ that minimizes the training error

$$L_{\text{train}}(\boldsymbol{w}) = \frac{1}{T}\sum_{t=1}^{T} l(\boldsymbol{x}_t, \boldsymbol{y}_t; \boldsymbol{w}). \tag{4.1}$$

This is equivalent to maximizing the likelihood $\prod_{t=1}^{T} p(\boldsymbol{x}_t, \boldsymbol{y}_t; \boldsymbol{w})$. Hence, $\hat{\boldsymbol{w}}_T$ is the maximum likelihood estimator. The Cramér-Rao theorem states that the expected squared error of an unbiased estimator satisfies

$$E[(\hat{\boldsymbol{w}}_T - \boldsymbol{w}^*)(\hat{\boldsymbol{w}}_T - \boldsymbol{w}^*)^T] \geq \frac{1}{T}G^{-1}, \tag{4.2}$$

where the inequality holds in the sense of positive definiteness of matrices. An estimator is said to be efficient or Fisher efficient when it satisfies equation 4.2 with equality for large T. The maximum likelihood estimator is Fisher efficient, implying that it is the best estimator attaining the Cramér-Rao bound asymptotically,

$$\lim_{T\to\infty} TE[(\hat{\boldsymbol{w}}_T - \boldsymbol{w}^*)(\hat{\boldsymbol{w}}_T - \boldsymbol{w}^*)^T] = G^{-1}, \tag{4.3}$$

where G^{-1} is the inverse of the Fisher information matrix $G = (g_{ij})$ defined by equation 3.11.

Examples $(\boldsymbol{x}_1, \boldsymbol{y}_1), (\boldsymbol{x}_2, \boldsymbol{y}_2) \ldots$ are given one at a time in the case of online learning. Let $\tilde{\boldsymbol{w}}_t$ be an online estimator at time t. At the next time, $t+1$, the estimator $\tilde{\boldsymbol{w}}_t$ is modified to give a new estimator $\tilde{\boldsymbol{w}}_{t+1}$ based on the current observation $(\boldsymbol{x}_t, \boldsymbol{y}_t)$. The old observations $(\boldsymbol{x}_1, \boldsymbol{y}_1), \ldots, (\boldsymbol{x}_{t-1}, \boldsymbol{y}_{t-1})$ cannot be reused to obtain $\tilde{\boldsymbol{w}}_{t+1}$, so the learning rule is written as

$$\tilde{\boldsymbol{w}}_{t+1} = \boldsymbol{m}(\boldsymbol{x}_t, \boldsymbol{y}_t, \tilde{\boldsymbol{w}}_t).$$

The process $\{\tilde{\boldsymbol{w}}_t\}$ is Markovian. Whatever learning rule $\boldsymbol{m}$ is chosen, the behavior of the estimator $\tilde{\boldsymbol{w}}_t$ is never better than that of the optimal batch estimator $\hat{\boldsymbol{w}}_t$ because of this restriction. The gradient online learning rule

$$\tilde{\boldsymbol{w}}_{t+1} = \tilde{\boldsymbol{w}}_t - \eta_t C \frac{\partial l(\boldsymbol{x}_t, \boldsymbol{y}_t; \tilde{\boldsymbol{w}}_t)}{\partial \boldsymbol{w}},$$

was proposed where C is a positive-definite matrix, and its dynamical behavior was studied by Amari (1967) when the learning constant $\eta_t = \eta$ is fixed. Heskes and Kappen (1991) obtained similar results, which ignited research into online learning. When η_t satisfies some condition, say, $\eta_t = c/t$, for a positive constant c, the stochastic approximation guarantees that $\tilde{\boldsymbol{w}}_t$ is a consistent estimator converging to $\boldsymbol{w}^*$. However, it is not Fisher efficient in general.

There arises a question of whether there exists a learning rule that gives an efficient estimator. If it exists, the asymptotic behavior of online learning is equivalent to that of the best batch estimation method. This article answers the question affirmatively, by giving an efficient online learning rule (see Amari, 1995; see also Opper, 1996).

Let us consider the natural gradient learning rule,

$$\tilde{\boldsymbol{w}}_{t+1} = \tilde{\boldsymbol{w}}_t - \frac{1}{t}\tilde{\nabla} l(\boldsymbol{x}_t, \boldsymbol{y}_t, \tilde{\boldsymbol{w}}_t). \tag{4.4}$$

Theorem 2. *Under the learning rule (see equation 4.4), the natural gradient online estimator $\tilde{\boldsymbol{w}}_t$ is Fisher efficient.*

Proof. Let us denote the covariance matrix of estimator $\tilde{\boldsymbol{w}}_t$ by

$$\tilde{V}_{t+1} = E[(\tilde{\boldsymbol{w}}_{t+1} - \boldsymbol{w}^*)(\tilde{\boldsymbol{w}}_{t+1} - \boldsymbol{w}^*)^T]. \tag{4.5}$$

This shows the expectation of the squared error. We expand

$$\frac{\partial l(\boldsymbol{x}_t, \boldsymbol{y}_t; \tilde{\boldsymbol{w}}_t)}{\partial \boldsymbol{w}} = \frac{\partial l(\boldsymbol{x}_t, \boldsymbol{y}_t; \boldsymbol{w}^*)}{\partial \boldsymbol{w}} + \frac{\partial^2 l(\boldsymbol{x}_t, \boldsymbol{y}_t; \boldsymbol{w}^*)}{\partial \boldsymbol{w} \partial \boldsymbol{w}} (\tilde{\boldsymbol{w}}_t - \boldsymbol{w}^*) + O(|\tilde{\boldsymbol{w}}_t - \boldsymbol{w}^*|^2).$$

By subtracting $\boldsymbol{w}^*$ from the both sides of equation 4.4 and taking the expectation of the square of the both sides, we have

$$\tilde{V}_{t+1} = \tilde{V}_t - \frac{2}{t}\tilde{V}_t + \frac{1}{t^2} G^{-1} + O\left(\frac{1}{t^3}\right), \tag{4.6}$$

where we used

$$E\left[\frac{\partial l(\boldsymbol{x}_t, \boldsymbol{y}_t; \boldsymbol{w}^*)}{\partial \boldsymbol{w}}\right] = 0, \tag{4.7}$$

$$E\left[\frac{\partial^2 l(\boldsymbol{x}_t, \boldsymbol{y}_t; \boldsymbol{w}^*)}{\partial \boldsymbol{w} \partial \boldsymbol{w}}\right] = G(\boldsymbol{w}^*), \tag{4.8}$$

$$G(\tilde{\boldsymbol{w}}_t) = G(\boldsymbol{w}^*) + O\left(\frac{1}{t}\right),$$

because $\tilde{\boldsymbol{w}}_t$ converges to $\boldsymbol{w}^*$ as guaranteed by stochastic approximation under certain conditions (see Kushner & Clark, 1978). The solution of equation 4.6 is written asymptotically as

$$\tilde{V}_t = \frac{1}{t} G^{-1} + O\left(\frac{1}{t^2}\right),$$

proving the theorem.

The theory can be extended to be applicable to the unrealizable teacher case, where

$$K(\boldsymbol{w}) = E\left[\frac{\partial^2}{\partial \boldsymbol{w} \partial \boldsymbol{w}} l(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{w})\right] \tag{4.9}$$

should be used instead of $G(\boldsymbol{w})$ in order to obtain the same efficient result as the optimal batch procedure. This is locally equivalent to the Newton-Raphson method. The results can be stated in terms of the generalization error instead of the covariance of the estimator, and we can obtain more universal results (see Amari, 1993; Amari & Murata, 1993).

Remark. In the cases of blind source separation and deconvolution, the models are semiparametric, including the unknown function r (see equation 3.18). In such cases, the Cramér-Rao bound does not necessarily hold. Therefore, Theorem 2 does not hold in these cases. It holds when we can estimate the true r of the source probability density functions and use it to define the loss function $l(\boldsymbol{x}, W)$. Otherwise equation 4.8 does not hold. The stability of the true solution is not necessarily guaranteed either. Amari, Chen, & Cichocki (in press) have analyzed this situation and proposed a universal method of attaining the stability of the equilibrium solution.

5 Adaptive Learning Constant

The dynamical behavior of the learning rule (see equation 3.2) was studied in Amari (1967) when η_t is a small constant η. In this case, $\boldsymbol{w}_t$ fluctuates around the (local) optimal value $\boldsymbol{w}^*$ for large t. The expected value and variance of $\boldsymbol{w}_t$ was studied, and the trade-off between the convergence speed and accuracy of convergence was demonstrated.

When the current $\boldsymbol{w}_t$ is far from the optimal $\boldsymbol{w}^*$, it is desirable to use a relatively large η to accelerate the convergence. When it is close to $\boldsymbol{w}^*$, a

small η is preferred in order to eliminate fluctuations. An idea of an adaptive change of η was discussed in Amari (1967) and was called "learning of learning rules."

Sompolinsky et al. (1995) (see also Barkai et al., 1995) proposed a rule of adaptive change of η_t, which is applicable to the pattern classification problem where the expected loss $L(\boldsymbol{w})$ is not differentiable at $\boldsymbol{w}^*$. This article generalizes their idea to a more general case where $L(\boldsymbol{w})$ is differentiable and analyzes its behavior by using the Riemannian structure.

We propose the following learning scheme:

$$\boldsymbol{w}_{t+1} = \boldsymbol{w}_t - \eta_t \tilde{\nabla} l(\boldsymbol{x}_t, \boldsymbol{y}_t; \hat{\boldsymbol{w}}_t) \tag{5.1}$$

$$\eta_{t+1} = \eta_t \exp\{\alpha[\beta l(\boldsymbol{x}_t, \boldsymbol{y}_t; \hat{\boldsymbol{w}}_t) - \eta_t]\}, \tag{5.2}$$

where α and β are constants. We also assume that the training data are generated by a realizable deterministic teacher and that $L(\boldsymbol{w}^*) = 0$ holds at the optimal value. (See Murata, Müller, Ziehe, and Amari (1996) for a more general case.) We try to analyze the dynamical behavior of learning by using the continuous version of the algorithm for the sake of simplicity,

$$\frac{d}{dt}\boldsymbol{w}_t = -\eta_t G^{-1}(\boldsymbol{w}_t)\frac{\partial}{\partial \boldsymbol{w}} l(\boldsymbol{x}_t, \boldsymbol{y}_t; \boldsymbol{w}_t), \tag{5.3}$$

$$\frac{d}{dt}\eta_t = \alpha\eta_t[\beta l(\boldsymbol{x}_t, \boldsymbol{z}_t; \boldsymbol{w}_t) - \eta_t]. \tag{5.4}$$

In order to show the dynamical behavior of $(\boldsymbol{w}_t, \eta_t)$, we use the averaged version of equations 5.3 and 5.4 with respect to the current input-output pair $(\boldsymbol{x}_t, \boldsymbol{y}_t)$. The averaged learning equation (Amari, 1967, 1977) is written as

$$\frac{d}{dt}\boldsymbol{w}_t = -\eta_t G^{-1}(\boldsymbol{w}_t)\left\langle \frac{\partial}{\partial \boldsymbol{w}} l(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{w}_t)\right\rangle, \tag{5.5}$$

$$\frac{d}{dt}\eta_t = \alpha\eta_t\{\beta\langle l(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{w}_t)\rangle - \eta_t\}, \tag{5.6}$$

where $\langle\ \ \rangle$ denotes the average over the current $(\boldsymbol{x}, \boldsymbol{y})$. We also use the asymptotic evaluations

$$\begin{aligned}\left\langle \frac{\partial}{\partial \boldsymbol{w}} l(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{w}_t)\right\rangle &= \left\langle \frac{\partial}{\partial \boldsymbol{w}} l(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{w}^*)\right\rangle + \left\langle \frac{\partial^2}{\partial \boldsymbol{w}\partial \boldsymbol{w}} l(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{w}^*)(\boldsymbol{w}_t - \boldsymbol{w}^*)\right\rangle \\ &= G^*(\boldsymbol{w}_t - \boldsymbol{w}^*), \\ \langle l(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{w}_t)\rangle &= \frac{1}{2}(\boldsymbol{w}_t - \boldsymbol{w}^*)^T G^*(\boldsymbol{w}_t - \boldsymbol{w}^*),\end{aligned}$$

where $G^* = G(\boldsymbol{w}^*)$ and we used $L(\boldsymbol{w}^*) = 0$. We then have

$$\frac{d}{dt}\boldsymbol{w}_t = -\eta_t(\boldsymbol{w}_t - \boldsymbol{w}^*), \tag{5.7}$$

$$\frac{d}{dt}\eta_t = \alpha\eta_t \left\{ \frac{\beta}{2}(\boldsymbol{w}_t - \boldsymbol{w}^*)^T G^*(\boldsymbol{w}_t - \boldsymbol{w}^*) - \eta_t \right\}. \tag{5.8}$$

Now we introduce the squared error variable,

$$e_t = \frac{1}{2}(\boldsymbol{w}_t - \boldsymbol{w}^*)^T G^*(\boldsymbol{w}_t - \boldsymbol{w}^*), \tag{5.9}$$

where e_t is the Riemannian magnitude of $\boldsymbol{w}_t - \boldsymbol{w}^*$. It is easy to show

$$\frac{d}{dt}e_t = -2\eta_t e_t, \tag{5.10}$$

$$\frac{d}{dt}\eta_t = \alpha\beta\eta_t e_t - \alpha\eta_t^2. \tag{5.11}$$

The behavior of equations 5.10 and 5.11 is interesting. The origin $(0, 0)$ is its attractor. However, the basin of attraction has a boundary of fractal structure. Anyway, starting from an adequate initial value, it has the solution of the form

$$e_t = \frac{a}{t},$$
$$\eta_t = \frac{b}{t}.$$

The coefficients a and b are determined from

$$a = 2ab$$
$$b = -\alpha\beta ab + \alpha b^2.$$

This gives

$$b = \frac{1}{2},$$
$$a = \frac{1}{\beta}\left(\frac{1}{2} - \frac{1}{\alpha}\right), \qquad \alpha > 2.$$

This proves the $1/t$ convergence rate of the generalization error, that is, the optimal order for any estimator $\hat{\boldsymbol{w}}_t$ converging to $\boldsymbol{w}^*$. The adaptive η_t shows a nice characteristic when the target teacher is slowly fluctuating or changes suddenly.

6 Natural Gradient in the Space of Perceptrons

The Riemannian metric and its inverse are calculated in this section to obtain the natural gradient explicitly. We begin with an analog simple perceptron whose input-output behavior is given by

$$y = f(\boldsymbol{w} \cdot \boldsymbol{x}) + n, \tag{6.1}$$

where n is a gaussian noise subject to $N(0, \sigma^2)$ and

$$f(u) = \frac{1 - e^{-u}}{1 + e^{-u}}. \tag{6.2}$$

The conditional probability density of y when $\boldsymbol{x}$ is applied is

$$p(y \mid \boldsymbol{x}; \boldsymbol{w}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}[y - f(\boldsymbol{w} \cdot \boldsymbol{x})]^2\right\}. \tag{6.3}$$

The distribution $q(\boldsymbol{x})$ of inputs $\boldsymbol{x}$ is assumed to be the normal distribution $N(0, I)$. The joint distribution of $(\boldsymbol{x}, y)$ is

$$p(y, \boldsymbol{x}; \boldsymbol{w}) = q(\boldsymbol{x})p(y \mid \boldsymbol{x}; \boldsymbol{w}).$$

In order to calculate the metric G of equation 3.11 explicitly, let us put

$$w^2 = |\boldsymbol{w}|^2 = \sum w_i^2 \tag{6.4}$$

where $|\boldsymbol{w}|$ is the Euclidean norm. We then have the following theorem.

Theorem 3. *The Fisher information metric is*

$$G(\boldsymbol{w}) = w^2 c_1(w) I + \{c_2(w) - c_1(w)\}\boldsymbol{w}\boldsymbol{w}^T, \tag{6.5}$$

where $c_1(w)$ and $c_2(w)$ are given by

$$c_1(w) = \frac{1}{4\sqrt{2\pi}\sigma^2 w^2} \int \{f^2(w\varepsilon) - 1\}^2 \exp\left\{-\frac{1}{2}\varepsilon^2\right\} d\varepsilon,$$
$$c_2(w) = \frac{1}{4\sqrt{2\pi}\sigma^2 w^2} \int \{f^2(w\varepsilon) - 1\}^2 \varepsilon^2 \exp\left\{-\frac{1}{2}\varepsilon^2\right\} d\varepsilon.$$

Proof. We have

$$\log p(y, \boldsymbol{x}; \boldsymbol{w}) = \log q(\boldsymbol{x}) - \log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}[y - f(\boldsymbol{w} \cdot \boldsymbol{x})]^2.$$

Hence,

$$\frac{\partial}{\partial w_i} \log p(y, \boldsymbol{x}; \boldsymbol{w}) = \frac{1}{\sigma^2}\{y - f(\boldsymbol{w} \cdot \boldsymbol{x})\} f'(\boldsymbol{w} \cdot \boldsymbol{x}) x_i$$
$$= \frac{1}{\sigma^2} n f'(\boldsymbol{w} \cdot \boldsymbol{x}) x_i.$$

The Fisher information matrix is given by

$$g_{ij}(\boldsymbol{w}) = E\left[\frac{\partial}{\partial w_i}\log p \frac{\partial}{\partial w_j}\log p\right]$$
$$= \frac{1}{\sigma^2}E[\{f'(\boldsymbol{w}\cdot\boldsymbol{x})\}^2 x_i x_j],$$

where $E[n^2] = \sigma^2$ is taken into account. This can be written, in the vector-matrix form, as

$$G(\boldsymbol{w}) = \frac{1}{\sigma^2}E[(f')^2\boldsymbol{x}\boldsymbol{x}^T].$$

In order to show equation 6.5, we calculate the quadratic form $\boldsymbol{r}^T G(\boldsymbol{w})\boldsymbol{r}$ for arbitrary $\boldsymbol{r}$. When $\boldsymbol{r} = \boldsymbol{w}$,

$$\boldsymbol{w}^T G\boldsymbol{w} = \frac{1}{\sigma^2}E[\{f'(\boldsymbol{w}\cdot\boldsymbol{x})\}^2(\boldsymbol{w}\cdot\boldsymbol{x})^2].$$

Since $u = \boldsymbol{w}\cdot\boldsymbol{x}$ is subject to $N(0, w^2)$, we put $u = w\varepsilon$, where ε is subject to $N(0, 1)$. Noting that

$$f'(u) = \frac{1}{2}\{1 - f^2(u)\},$$

we have,

$$\boldsymbol{w}^T G(\boldsymbol{w})\boldsymbol{w} = \frac{w^2}{4\sqrt{2\pi}\sigma^2}\int \varepsilon^2\{f^2(w\varepsilon) - 1\}^2 \exp\left\{-\frac{\varepsilon^2}{2}\right\} d\varepsilon,$$

which confirms equation 6.5 when $\boldsymbol{r} = \boldsymbol{w}$. We next put $\boldsymbol{r} = \boldsymbol{v}$, where $\boldsymbol{v}$ is an arbitrary unit vector orthogonal to $\boldsymbol{w}$ (in the Euclidean sense). We then have

$$\boldsymbol{v}^T G(\boldsymbol{w})\boldsymbol{v} = \frac{1}{4\sigma^2}E[\{f^2(\boldsymbol{w}\cdot\boldsymbol{x}) - 1\}^2(\boldsymbol{v}\cdot\boldsymbol{x})^2].$$

Since $u = \boldsymbol{w}\cdot\boldsymbol{x}$ and $v = \boldsymbol{v}\cdot\boldsymbol{x}$ are independent, and v is subject to $N(0, 1)$, we have

$$\boldsymbol{v}^T G(\boldsymbol{w})\boldsymbol{v} = \frac{1}{4\sigma^2}E[(\boldsymbol{v}\cdot\boldsymbol{x})^2]E[(f^2\{\boldsymbol{w}\cdot\boldsymbol{x}) - 1\}^2]$$
$$= \frac{1}{4\sqrt{2\pi}\sigma^2}\int \{f^2(w\varepsilon) - 1\}^2 \exp\left\{-\frac{\varepsilon^2}{2}\right\} d\varepsilon.$$

Since $G(\boldsymbol{w})$ in equation 6.5 is determined by the quadratic forms for n-independent $\boldsymbol{w}$ and $\boldsymbol{v}$'s, this proves equation 6.5.

To obtain the natural gradient, it is necessary to have an explicit form of G^{-1}. We can calculate $G^{-1}(\boldsymbol{w})$ explicitly in the perceptron case.

Theorem 4. The inverse of the Fisher information metric is

$$G^{-1}(\boldsymbol{w}) = \frac{1}{w^2 c_1(w)} I + \frac{1}{w^4}\left(\frac{1}{c_2(w)} - \frac{1}{c_1(w)}\right)\boldsymbol{w}\boldsymbol{w}^T. \tag{6.6}$$

This can easily be proved by direct calculation of GG^{-1}. The natural gradient learning equation (3.10) is then given by

$$\boldsymbol{w}_{t+1} = \boldsymbol{w}_t + \eta_t\{y_t - f(\boldsymbol{w}_t.\boldsymbol{x}_t)\} f'(\boldsymbol{w}_t \cdot \boldsymbol{x}_t) \left[\frac{1}{w_t^2 c_1(w_t)}\boldsymbol{x}_t + \frac{1}{w_t^4}\left(\frac{1}{c_2(w_t)} - \frac{1}{c_1(w_t)}\right)(\boldsymbol{w}_t \cdot \boldsymbol{x}_t)\boldsymbol{w}_t\right]. \tag{6.7}$$

We now show some other geometrical characteristics of the parameter space of perceptrons. The volume V_n of the manifold of simple perceptrons is measured by

$$V_n = \int \sqrt{|G(\boldsymbol{w})|}\, d\boldsymbol{w} \tag{6.8}$$

where $|G(\boldsymbol{w})|$ is the determinant of $G = (g_{ij})$, which represents the volume density by the Riemannian metric. It is interesting to see that the manifold of perceptrons has a finite volume.

Bayesian statistics considers that $\boldsymbol{w}$ is randomly chosen subject to a prior distribution $\pi(\boldsymbol{w})$. A choice of $\pi(\boldsymbol{w})$ is the Jeffrey prior or noninformative prior given by

$$\pi(\boldsymbol{w}) = \frac{1}{V_n}\sqrt{|G(\boldsymbol{w})|}. \tag{6.9}$$

The Jeffrey prior is calculated as follows.

Theorem 5. *The Jeffrey prior and the volume of the manifold are given, respectively, by*

$$\sqrt{|G(\boldsymbol{w})|} = \frac{w}{V_n}\sqrt{c_2(w)\{c_1(w)\}^{n-1}}, \tag{6.10}$$

$$V_n = a_{n-1}\int \sqrt{c_2(w)\{c_1(w)\}^{n-1}}\, w^n dw, \tag{6.11}$$

respectively, where a_{n-1} is the area of the unit $(n-1)$-sphere.

The Fisher metric G can also be calculated for multilayer perceptrons. Let us consider a multilayer perceptron having m hidden units with sigmoidal activation functions and a linear output unit. The input-output relation is

$$y = \sum v_i f(\boldsymbol{w}_i \cdot \boldsymbol{x}) + n,$$

or the conditional probability is

$$p(y \mid \boldsymbol{x}; \boldsymbol{v}, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_m) = c \exp\left[-\frac{1}{2}\{y - \sum v_i f(\boldsymbol{w}_i \cdot \boldsymbol{x})\}^2\right]. \tag{6.12}$$

The total parameter $\boldsymbol{w}$ consist of $\{\boldsymbol{v}, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_m\}$. Let us calculate the Fisher information matrix G. It consists of $m+1$ blocks corresponding to these $\boldsymbol{w}_i$'s and $\boldsymbol{v}$.

From

$$\frac{\partial}{\partial \boldsymbol{w}_i} \log p(y \mid \boldsymbol{x}; \boldsymbol{w}) = n v_i f'(\boldsymbol{w}_i \cdot \boldsymbol{x})\boldsymbol{x},$$

we easily obtain the block submatrix corresponding to $\boldsymbol{w}_i$ as

$$\begin{aligned} E\left[\frac{\partial}{\partial \boldsymbol{w}_i} \log p \frac{\partial}{\partial \boldsymbol{w}_i} \log p\right] &= \frac{1}{\sigma^4} E[n^2] v_i^2 E[\{f'(\boldsymbol{w}_i \cdot \boldsymbol{x})\}^2 \boldsymbol{x}\boldsymbol{x}^T] \\ &= \frac{1}{\sigma^2} v_i^2 E[\{f'(\boldsymbol{w}_i \cdot \boldsymbol{x})\}^2 \boldsymbol{x}\boldsymbol{x}^T]. \end{aligned}$$

This is exactly the same as the simple perceptron case except for a factor of $(v_i)^2$. For the off-diagonal block, we have

$$E\left[\frac{\partial}{\partial \boldsymbol{w}_i} \log p \frac{\partial}{\partial \boldsymbol{w}_j} \log p\right] = \frac{1}{\sigma^2} v_i v_j E[f'(\boldsymbol{w}_i \cdot \boldsymbol{x}) f'(\boldsymbol{w}_j \cdot \boldsymbol{x})\boldsymbol{x}\boldsymbol{x}^T].$$

In this case, we have the following form,

$$G_{\boldsymbol{w}_i \boldsymbol{w}_j} = c_{ij} I + d_{ii} \boldsymbol{w}_i \boldsymbol{w}_i^T + d_{ij} \boldsymbol{w}_i \boldsymbol{w}_j^T + d_{ji} \boldsymbol{w}_j \boldsymbol{w}_i^T + d_{jj} \boldsymbol{w}_j \boldsymbol{w}_j^T, \tag{6.13}$$

where the coefficients c_{ij} and d_{ij}'s are calculated explicitly by similar methods.

The $\boldsymbol{v}$ block and $\boldsymbol{v}$ and $\boldsymbol{w}_i$ block are also calculated similarly. However, the inversion of G is not easy except for simple cases. It requires inversion of a $2(m+1)$ dimensional matrix. However, this is much better than the direct inversion of the original $(n+1)m$-dimensional matrix of G. Yang and Amari (1997) performed a preliminary study on the performance of the natural gradient learning algorithm for a simple multilayer perceptron. The result shows that natural gradient learning might be free from the plateau phenomenon. Once the learning trajectory is trapped in a plateau, it takes a long time to get out of it.

7 Natural Gradient in the Space of Matrices and Blind Source Separation

We now define a Riemannian structure to the space of all the $m \times m$ nonsingular matrices, which forms a Lie group denoted by $Gl(m)$, for the purpose of introducing the natural gradient learning rule to the blind source separation problem. Let dW be a small deviation of a matrix from W to $W + dW$. The tangent space T_W of $Gl(m)$ at W is a linear space spanned by all such small deviations dW_{ij}'s and is called the Lie algebra.

We need to introduce an inner product at W by defining the squared norm of dW

$$ds^2 = \langle dW, dW \rangle_W = \| dW \|^2 .$$

By multiplying W^{-1} from the right, W is mapped to $WW^{-1} = I$, the unit matrix, and $W + dW$ is mapped to $(W + dW)W^{-1} = I + dX$, where

$$dX = dWW^{-1}. \tag{7.1}$$

This shows that a deviation dW at W is equivalent to the deviation dX at I by the correspondence given by multiplication of W^{-1}. The Lie group invariance requires that the metric is kept invariant under this correspondence, that is, the inner product of dW at W is equal to the inner product of dWY at WY for any Y,

$$\langle dW, dW \rangle_W = \langle dWY, dWY \rangle_{WY}. \tag{7.2}$$

When $Y = W^{-1}$, $WY = I$. This principle was used to derive the natural gradient in Amari, Cichocki, and Yang (1996); see also Yang and Amari (1997) for detail. Here we give its analysis by using dX.

We define the inner product at I by

$$\langle dX, dX \rangle_I = \sum_{i,j} (dX_{ij})^2 = \mathrm{tr}(dX^T dX). \tag{7.3}$$

We then have the Riemannian metric structure at W as

$$\langle dW, dW \rangle_W = \mathrm{tr}\{(W^{-1})^T dW^T dW W^{-1}\}. \tag{7.4}$$

We can write the metric tensor G in the component form. It is a quantity having four indices $G_{ij,kl}(W)$ such that

$$\begin{aligned} ds^2 &= \sum G_{ij,kl}(W) dW_{ij} dW_{kl}, \\ G_{ij,kl}(W) &= \sum_m \delta_{ik} W^{-1}_{jm} W^{-1}_{lm}, \end{aligned} \tag{7.5}$$

where W^{-1}_{jm} are the components of W^{-1}. While it may not appear to be straightforward to obtain the explicit form of G^{-1} and natural gradient $\tilde{\nabla}L$, in fact it can be calculated as shown below.

Theorem 6. *The natural gradient in the matrix space is given by*

$$\tilde{\nabla}L = (\nabla L)W^T W. \tag{7.6}$$

Proof. The metric is Euclidean at I, so that both $G(I)$ and its inverse, $G^{-1}(I)$, are the identity. Therefore, by mapping dW at W to dX at I, the natural gradient learning rule in terms of dX is written as

$$\frac{dX}{dt} = -\eta_t G^{-1}(I)\frac{\partial L}{\partial X} = -\eta_t \frac{\partial L}{\partial X}, \tag{7.7}$$

where the continuous time version is used. We have from equation 7.1

$$\frac{dX}{dt} = \frac{dW}{dt}W^{-1}. \tag{7.8}$$

The gradient $\partial L/\partial X$ is calculated as

$$\frac{\partial L}{\partial X} = \frac{\partial L(W)}{\partial W}\left(\frac{\partial W^T}{\partial X}\right) = \frac{\partial L}{\partial W}W^T.$$

Therefore, the natural gradient learning rule is

$$\frac{dW}{dt} = -\eta_t \frac{\partial L}{\partial W}W^T W,$$

which proves equation 7.6.

The $dX = dWW^{-1}$ forms a basis of the tangent space at W, but this is not integrable; that is, we cannot find any matrix function $X = X(W)$ that satisfies equation 7.1. Such a basis is called a nonholonomic basis. This is a locally defined basis but is convenient for our purpose. Let us calculate the natural gradient explicitly. To this end, we put

$$l(\boldsymbol{x}, W) = -\log\det|W| - \sum_{i-1}^{n}\log f_i(y_i), \tag{7.9}$$

where $\boldsymbol{y} = W\boldsymbol{x}$ and $f_i(y_i)$ is an adequate probability distribution. The expected loss is

$$L(W) = E[l(\boldsymbol{x}, W)],$$

which represents the entropy of the output $\boldsymbol{y}$ after a componentwise nonlinear transformation (Nadal & Parga, 1994; Bell & Sejnowski, 1995). The independent component analysis or the mutual information criterion also gives a similar loss function (Comon, 1994; Amari et al., 1996; see also Oja & Karhunen, 1995). When f_i is the true probability density function of the ith source, $l(\boldsymbol{x}, W)$ is the negative of the log likelihood.

The natural gradient of l is calculated as follows. We calculate the differential

$$dl = l(\boldsymbol{x}, W + dW) - l(\boldsymbol{x}, W) = -d \log \det |W| - \sum d \log f_i(y_i)$$

due to change dW. Then,

$$\begin{aligned} d \log \det |W| &= \log \det |W + dW| - \log \det |W| \\ &= \log \det |(W + dW) W^{-1}| = \log(\det |I + dX|) \\ &= \mathrm{tr} dX. \end{aligned}$$

Similarly, from $d\boldsymbol{y} = dW\boldsymbol{x}$,

$$\begin{aligned} \sum d \log f_i(y_i) &= -\varphi(\boldsymbol{y})^T dW \boldsymbol{x} \\ &= -\varphi(\boldsymbol{y})^T dX \boldsymbol{y}, \end{aligned}$$

where $\varphi(\boldsymbol{y})$ is the column vector

$$\begin{aligned} &\varphi(\boldsymbol{y}) = [\varphi_1(y_1), \ldots, \varphi_m(y_m)], \\ &\varphi_i(y_i) = -\frac{d}{dy} \log f_i(y_i). \end{aligned} \tag{7.10}$$

This gives $\partial L / \partial X$, and the natural gradient learning equation is

$$\frac{dW}{dt} = \eta_t (I - \varphi(\boldsymbol{y})^T \boldsymbol{y}) W. \tag{7.11}$$

The efficiency of this equation is studied from the statistical and information geometrical point of view (Amari & Kawanabe, 1997; Amari & Cardoso, in press). We further calculate the Hessian by using the natural frame dX,

$$d^2 l = \boldsymbol{y}^T dX^T \dot{\varphi}(\boldsymbol{y}) dX \boldsymbol{y} + \varphi(\boldsymbol{y})^T dX dX \boldsymbol{y}, \tag{7.12}$$

where $\dot{\varphi}(\boldsymbol{y})$ is the diagonal matrix with diagonal entries $d\varphi_i(y_i)/dy_i$. Its expectation can be explicitly calculated (Amari et al., in press). The Hessian is decomposed into diagonal elements and two-by-two diagonal blocks (see also Cardoso & Laheld, 1996). Hence, the stability of the above learning rule is easily checked. Thus, in terms of dX, we can solve the two fundamental problems: the efficiency and the stability of learning algorithms of blind source separation (Amari & Cardoso, in press; Amari et al., in press).

8 Natural Gradient in Systems Space

The problem is how to define the Riemannian structure in the parameter space $\{W(z)\}$ of systems, where z is the time-shift operator. This was given in Amari (1987) from the point of view of information geometry (Amari, 1985, 1997a; Murray & Rice, 1993). We show here only ideas (see Douglas et al., 1996; Amari, Douglas, Cichocki, & Yang, 1997, for preliminary studies).

In the case of multiterminal deconvolution, a typical loss function l is given by

$$l = -\log\det|W_0| - \sum_i \int p\{y_i; \boldsymbol{W}(z)\}\log f_i(y_i)dy_i, \tag{8.1}$$

where $p\{y_i; \boldsymbol{W}(z)\}$ is the marginal distribution of $\boldsymbol{y}(t)$ which is derived from the past sequence of $\boldsymbol{x}(t)$ by matrix convolution $\boldsymbol{W}(z)$ of equation 3.24. This type of loss function is obtained from maximization of entropy, independent component analysis, or maximum likelihood.

The gradient of l is given by

$$\nabla_m l = -(W_0^{-1})^T\delta_{0m} + \boldsymbol{\varphi}(\boldsymbol{y}_t)\boldsymbol{x}^T(t-m), \tag{8.2}$$

where

$$\nabla_m = \frac{\partial}{\partial W_m},$$

and

$$\nabla l = \sum_{m=0}^{d}(\nabla_m l)z^{-m}. \tag{8.3}$$

In order to calculate the natural gradient, we need to define the Riemannian metric G in the manifold of linear systems. The geometrical theory of the manifold of linear systems by Amari (1987) defines the Riemannian metric and a pair of dual affine connections in the space of linear systems.

Let

$$d\boldsymbol{W}(z) = \sum_m d\boldsymbol{W}_m z^{-m} \tag{8.4}$$

be a small deviation of $\boldsymbol{W}(z)$. We postulate that the inner product $\langle d\boldsymbol{W}(z), d\boldsymbol{W}(z)\rangle$ is invariant under the operation of any matrix filter $\boldsymbol{Y}(z)$,

$$\langle d\boldsymbol{W}(z), d\boldsymbol{W}(z)\rangle_{\boldsymbol{W}(z)} = \langle d\boldsymbol{W}(z)\boldsymbol{Y}(z), d\boldsymbol{W}(z)\boldsymbol{Y}(z)\rangle_{\boldsymbol{W}\boldsymbol{Y}}, \tag{8.5}$$

where $\boldsymbol{Y}(z)$ is any system matrix. If we put

$$\boldsymbol{Y}(z) = \{\boldsymbol{W}(z)\}^{-1},$$

which is a general system not necessarily belonging to FIR,

$$\boldsymbol{W}(z)\{\boldsymbol{W}(z)\}^{-1} = \boldsymbol{I}(z),$$

which is the identity system

$$\boldsymbol{I}(z) = I$$

not including any z^{-m} terms. The tangent vector $d\boldsymbol{W}(z)$ is mapped to

$$dX(z) = d\boldsymbol{W}(z)\{\boldsymbol{W}(z)\}^{-1}. \tag{8.6}$$

The inner product at I is defined by

$$\langle dX(z), dX(z)\rangle_I = \sum_{m,ij}(dX_{m,ij})^2, \tag{8.7}$$

where $dX_{m,ij}$ are the elements of matrix dX_m.

The natural gradient

$$\tilde{\nabla} l = G^{-1} \circ \nabla l$$

of the manifold of systems is given as follows.

Theorem 7. *The natural gradient of the manifold of systems is given by*

$$\tilde{\nabla} l = \nabla l(z)\boldsymbol{W}^T(z^{-1})\boldsymbol{W}(z), \tag{8.8}$$

where operator z^{-1} should be operated adequately.

The proof is omitted. It should be remarked that $\tilde{\nabla} l$ does not belong to the class of FIR systems, nor does it satisfy the causality condition either. Hence, in order to obtain an online learning algorithm, we need to introduce time delay to map it to the space of causal FIR systems. This article shows only the principles involved; details will published in a separate article by Amari, Douglas, and Cichocki.

9 Conclusions

This article introduces the Riemannian structures to the parameter spaces of multilayer perceptrons, blind source separation, and blind source deconvolution by means of information geometry. The natural gradient learning method is then introduced and is shown to be statistically efficient. This implies that optimal online learning is as efficient as optimal batch learning when the Fisher information matrix exists. It is also suggested that natural gradient learning might be easier to get out of plateaus than conventional stochastic gradient learning.

Acknowledgments

I thank A. Cichocki, A. Back, and H. Yang at RIKEN Frontier Research Program for their discussions.

References

Amari, S. (1967). Theory of adaptive pattern classifiers. *IEEE Trans., EC-16*(3), 299–307.

Amari, S. (1977). Neural theory of association and concept-formation. *Biological Cybernetics, 26*, 175–185.

Amari, S. (1985). *Differential-geometrical methods in statistics*. Lecture Notes in Statistics 28. New York: Springer-Verlag.

Amari, S. (1987). Differential geometry of a parametric family of invertible linear systems—Riemannian metric, dual affine connections and divergence. *Mathematical Systems Theory, 20*, 53–82.

Amari, S. (1993). Universal theorem on learning curves. *Neural Networks, 6*, 161–166.

Amari, S. (1995). Learning and statistical inference. In M. A. Arbib (Ed.), *Handbook of brain theory and neural networks* (pp. 522–526). Cambridge, MA: MIT Press.

Amari, S. (1996). Neural learning in structured parameter spaces—Natural Riemannian gradient. In M. C. Mozer, M. I. Jordan, & Th. Petsche (Eds.), *Advances in neural processing systems, 9*. Cambridge, MA: MIT Press.

Amari, S. (1997a). Information geometry. *Contemporary Mathematics, 203*, 81–95.

Amari, S. (1997b). *Superefficiency in blind source separation*. Unpublished manuscript.

Amari, S., & Cardoso, J. F. (In press). Blind source separation—Semi-parametric statistical approach. *IEEE Trans. on Signal Processing*.

Amari, S., Chen, T.-P., & Cichocki, A. (In press). Stability analysis of learning algorithms for blind source separation. *Neural Networks*.

Amari, S., Cichocki, A., & Yang, H. H. (1996). A new learning algorithm for blind signal separation, in *NIPS'95*, vol. 8, Cambridge, MA: MIT Press.

Amari, S., Douglas, S. C., Cichocki, A., & Yang, H. H. (1997). Multichannel blind deconvolution and equalization using the natural gradient. *Signal Processing*

Advance in Wireless Communication Workshop, Paris.
Amari, S., & Kawanabe, M. (1997). Information geometry of estimating functions in semiparametric statistical models, *Bernoulli, 3*, 29–54.
Amari, S., Kurata, K., & Nagaoka, H. (1992). Information geometry of Boltzmann machines. *IEEE Trans. on Neural Networks, 3*, 260–271.
Amari, S., & Murata, N. (1993). Statistical theory of learning curves under entropic loss criterion. *Neural Computation, 5*, 140–153.
Barkai, N., Seung, H. S., & Sompolinsky, H. (1995). Local and global convergence of on-line learning. *Phys. Rev. Lett., 75*, 1415–1418.
Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation, 7*, 1129–1159.
Bickel, P. J., Klassen, C. A. J., Ritov, Y., & Wellner, J. A. (1993). *Efficient and adaptive estimation for semiparametric models*. Baltimore: Johns Hopkins University Press.
Campbell, L. L. (1985). The relation between information theory and the differential-geometric approach to statistics. *Information Sciences, 35*, 199–210.
Cardoso, J. F., & Laheld, B. (1996). Equivariant adaptive source separation. *IEEE Trans. on Signal Processing, 44*, 3017–3030.
Chentsov, N. N. (1972). *Statistical decision rules and optimal inference* (in Russian). Moscow: Nauka [translated in English (1982), Rhode Island: AMS].
Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing, 36*, 287–314.
Douglas, S. C., Cichocki, A., & Amari, S. (1996). Fast convergence filtered regressor algorithms for blind equalization. *Electronics Letters, 32*, 2114–2115.
Heskes, T., & Kappen, B. (1991). Learning process in neural networks. *Physical Review, A44*, 2718–2762.
Jutten, C., & Hérault, J. (1991). Blind separation of sources, an adaptive algorithm based on neuromimetic architecture. *Signal Processing, 24*(1), 1–31.
Kushner, H. J., & Clark, D. S. (1978). *Stochastic approximation methods for constrained and unconstrained systems*. Berlin: Springer-Verlag.
Murata, N., & Müller, K. R., Ziehe, A., & Amari, S. (1996). Adaptive on-line learning in changing environments. In M. C. Mozer, M. I. Jordan, & Th. Petsche (Eds.), *Advaces in neural processing systems, 9*. Cambridge, MA: MIT Press.
Murray, M. K., & Rice, J. W. (1993). *Differential geometry and statistics*. New York: Chapman & Hall.
Nadal, J. P. & Parga, N. (1994). Nonlinear neurons in the low noise limit—A factorial code maximizes information transfer. *Network, 5*, 561–581.
Oja, E., & Karhunen, J. (1995). Signal separation by nonlinear Hebbian learning. In M. Palaniswami et al. (Eds.), *Computational intelligence—A dynamic systems perspective* (pp. 83–97). New York: IEEE Press.
Opper, M. (1996). Online versus offline learning from random examples: General results. *Phys. Rev. Lett., 77*, 4671–4674.
Rao, C. R. (1945). Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society, 37*, 81–91.
Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. *Parallel Distributed Processing* (Vol. 1, pp. 318–362). Cambridge, MA: MIT Press.

Saad, D., & Solla, S. A. (1995). On-line learning in soft committee machines. *Phys. Rev. E, 52*, 4225–4243.

Sompolinsky, H., Barkai, N., & Seung, H. S. (1995). On-line learning of dichotomies: Algorithms and learning curves. In J.-H. Oh et al. (Eds.), *Neural networks: The statistical mechanics perspective* (pp. 105–130). Proceedings of the CTP-PBSRI Joint Workshop on Theoretical Physics. Singapore: World Scientific.

Tsypkin, Ya. Z. (1973). *Foundation of the theory of learning systems*. New York: Academic Press.

Van den Broeck, C., & Reimann, P. (1996). Unsupervised learning by examples: On-line versus off-line. *Phys. Rev. Lett., 76*, 2188–2191.

Widrow, B. (1963). *A statistical theory of adaptation*. Oxford: Pergamon Press.

Yang, H. H., & Amari, S. (1997). *Application of natural gradient in training multilayer perceptrons*. Unpublished manuscript.

Yang, H. H., & Amari, S. (In press). Adaptive on-line learning algorithms for blind separation—Maximum entropy and minimal mutual information. *Neural Computation*.

Received January 24, 1997; accepted May 20, 1997.

This article has been cited by:

1. Mahmoud Ismail, Mina Attari, Saeid Habibi, Samir Ziada. 2018. Estimation theory and Neural Networks revisited: REKF and RSVSF as optimization techniques for Deep-Learning. *Neural Networks* . [Crossref]
2. Moussab Djerrab, Alexandre Garcia, Maxime Sangnier, Florence d'Alché-Buc. 2018. Output Fisher embedding regression. *Machine Learning* **107**:8-10, 1229-1256. [Crossref]
3. Ryo IWAKI, Hiroki YOKOYAMA, Minoru ASADA. 2018. Incremental Estimation of Natural Policy Gradient with Relative Importance Weighting. *IEICE Transactions on Information and Systems* **E101.D**:9, 2346-2355. [Crossref]
4. Thomas Fösel, Petru Tighineanu, Talitha Weiss, Florian Marquardt. 2018. Reinforcement Learning with Neural Networks for Quantum Feedback. *Physical Review X* **8**:3. . [Crossref]
5. Darío Ramos-López, Andrés R. Masegosa, Antonio Salmerón, Rafael Rumí, Helge Langseth, Thomas D. Nielsen, Anders L. Madsen. 2018. Scalable importance sampling estimation of Gaussian mixture posteriors in Bayesian networks. *International Journal of Approximate Reasoning* **100**, 115-134. [Crossref]
6. Jing Wang, Huafei Sun, Simone Fiori. 2018. A Riemannian-steepest-descent approach for optimization on the real symplectic group. *Mathematical Methods in the Applied Sciences* **41**:11, 4273-4286. [Crossref]
7. Feng Bao, Yue Deng, Mulong Du, Zhiquan Ren, Qingzhao Zhang, Yanyu Zhao, Jinli Suo, Zhengdong Zhang, Meilin Wang, Qionghai Dai. 2018. Probabilistic natural mapping of gene-level tests for genome-wide association studies. *Briefings in Bioinformatics* **19**:4, 545-553. [Crossref]
8. Fan Gu, Xue Luo, Yuqing Zhang, Yu Chen, Rong Luo, Robert L. Lytton. 2018. Prediction of geogrid-reinforced flexible pavement performance using artificial neural network approach. *Road Materials and Pavement Design* **19**:5, 1147-1163. [Crossref]
9. Victor M. H. Ong, David J. Nott, Minh-Ngoc Tran, Scott A. Sisson, Christopher C. Drovandi. 2018. Variational Bayes with synthetic likelihood. *Statistics and Computing* **28**:4, 971-988. [Crossref]
10. Gonzalo Martín-Vázquez, Toshitake Asabuki, Yoshikazu Isomura, Tomoki Fukai. 2018. Learning Task-Related Activities From Independent Local-Field-Potential Components Across Motor Cortex Layers. *Frontiers in Neuroscience* **12**. . [Crossref]
11. Meng Wang, Chuang-Bai Xiao, Zhen-Hu Ning, Tong Li, Bei Gong. 2018. Neural Networks for Compressed Sensing Based on Information Geometry. *Circuits, Systems, and Signal Processing* **10**. . [Crossref]
12. Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, Giuseppe Carleo. 2018. Neural-network quantum state tomography. *Nature Physics* **14**:5, 447-450. [Crossref]
13. Ana Rodrigues, Charles Cavalcante. 2018. Principal Curves for Statistical Divergences and an Application to Finance. *Entropy* **20**:5, 333. [Crossref]
14. Raphael Kaubruegger, Lorenzo Pastori, Jan Carl Budich. 2018. Chiral topological phases from artificial neural networks. *Physical Review B* **97**:19. . [Crossref]
15. Kendall Lowrey, Svetoslav Kolev, Jeremy Dao, Aravind Rajeswaran, Emanuel Todorov. Reinforcement learning for non-prehensile manipulation: Transfer from simulation to physical system 35-42. [Crossref]
16. Florent Bouchard, Jerome Malick, Marco Congedo. 2018. Riemannian Optimization and Approximate Joint Diagonalization for Blind Source Separation. *IEEE Transactions on Signal Processing* **66**:8, 2041-2054. [Crossref]
17. Zifeng Wang, Zheng Yu, Qing Ling, Dimitris Berberidis, Georgios B. Giannakis. 2018. Decentralized RLS With Data-Adaptive Censoring for Regressions Over Large-Scale Networks. *IEEE Transactions on Signal Processing* **66**:6, 1634-1648. [Crossref]
18. Weili Guo, Junsheng Zhao, Jinxia Zhang, Haikun Wei, Aiguo Song, Kanjian Zhang. 2018. Stability analysis of opposite singularity in multilayer perceptrons. *Neurocomputing* **282**, 192-201. [Crossref]
19. Ana Flávia P. Rodrigues, Igor M. Guerreiro, Charles Casimiro Cavalcante. 2018. Deformed exponentials and portfolio selection. *International Journal of Modern Physics C* **29**:03, 1850029. [Crossref]
20. Jiayu Chen, Vince D Calhoun, Dongdong Lin, Nora I Perrone-Bizzozero, Juan R Bustillo, Godfrey D Pearlson, Steven G Potkin, Theo G M van Erp, Fabio Macciardi, Stefan Ehrlich, Beng-Choon Ho, Scott R Sponheim, Lei Wang, Julia M Stephen, Andrew R Mayer, Faith M Hanlon, Rex E Jung, Brett A Clementz, Matcheri S Keshavan, Elliot S Gershon, John A Sweeney, Carol A Tamminga, Ole A Andreassen, Ingrid Agartz, Lars T Westlye, Jing Sui, Yuhui Du, Jessica A Turner, Jingyu Liu. 2018. Shared Genetic Risk of Schizophrenia and Gray Matter Reduction in 6p22.1. *Schizophrenia Bulletin* **13**. . [Crossref]
21. Andrew L Ferguson. 2018. Machine learning and data science in soft materials engineering. *Journal of Physics: Condensed Matter* **30**:4, 043002. [Crossref]
22. Michele Scarpiniti, Simone Scardapane, Danilo Comminiello, Raffaele Parisi, Aurelio Uncini. **69**, 57. [Crossref]

23. Shun-ichi Amari, Tomoko Ozeki, Ryo Karakida, Yuki Yoshida, Masato Okada. 2018. Dynamics of Learning in MLP: Natural Gradient and Singularity Revisited. *Neural Computation* **30**:1, 1-33. [Abstract] [Full Text] [PDF] [PDF Plus]
24. Reiji Hatsugai, Mary Inaba. Robust Reinforcement Learning with a Stochastic Value Function 519-526. [Crossref]
25. Ivan Glasser, Nicola Pancotti, Moritz August, Ivan D. Rodriguez, J. Ignacio Cirac. 2018. Neural-Network Quantum States, String-Bond States, and Chiral Topological States. *Physical Review X* **8**:1. . [Crossref]
26. Hendrik Barfuss, Klaus Reindl, Walter Kellermann. Informed Spatial Filtering Based on Constrained Independent Component Analysis 237-278. [Crossref]
27. Hongjun Wang, Jiani Hu, Weihong Deng. 2018. Face Feature Extraction: A Complete Review. *IEEE Access* **6**, 6001-6039. [Crossref]
28. Léon Bottou, Frank E. Curtis, Jorge Nocedal. 2018. Optimization Methods for Large-Scale Machine Learning. *SIAM Review* **60**:2, 223-311. [Crossref]
29. Yuki Fujimoto, Toru Ohira. A Neural Network Model with Bidirectional Whitening 47-57. [Crossref]
30. CHU Ding-li, CHEN Hong, CHEN Han-yi. 2018. Blind Source Separation based on Whale Optimization Algorithm. *MATEC Web of Conferences* **173**, 03052. [Crossref]
31. Motoya Ohnishi, Masahiro Yukawa. 2018. Online Nonlinear Estimation via Iterative L<formula> <tex>^{2}</tex> </formula>-Space Projections: Reproducing Kernel of Subspace. *IEEE Transactions on Signal Processing* 1-1. [Crossref]
32. Jorma Joutsensaari, Matthew Ozon, Tuomo Nieminen, Santtu Mikkonen, Timo Lähivaara, Stefano Decesari, M. Cristina Facchini, Ari Laaksonen, Kari E. J. Lehtinen. 2018. Identification of new particle formation events with deep learning. *Atmospheric Chemistry and Physics* **18**:13, 9597-9615. [Crossref]
33. Vahid Tavakol Aghaei, Ahmet Onat, Sinan Yıldırım. 2018. A Markov chain Monte Carlo algorithm for Bayesian policy search. *Systems Science & Control Engineering* **6**:1, 438-455. [Crossref]
34. Arslan Chaudhry, Puneet K. Dokania, Thalaiyasingam Ajanthan, Philip H. S. Torr. Riemannian Walk for Incremental Learning: Understanding Forgetting and Intransigence 556-572. [Crossref]
35. Xiaofeng Liu, B. V. K. Vijaya Kumar, Chao Yang, Qingming Tang, Jane You. Dependency-Aware Attention Control for Unconstrained Face Recognition with Image Sets 573-590. [Crossref]
36. Yuan Wang, Kirubakaran Velswamy, Biao Huang. 2018. A Novel Approach to Feedback Control with Deep Reinforcement Learning. *IFAC-PapersOnLine* **51**:18, 31-36. [Crossref]
37. San Gultekin, John Paisley. 2017. Nonlinear Kalman Filtering With Divergence Minimization. *IEEE Transactions on Signal Processing* **65**:23, 6319-6331. [Crossref]
38. Hao Xu, Huafei Sun, Tong Zhang. 2017. Geometric structure of the joint N-voM distribution manifold and its applications to sensor networks. *Mathematical Methods in the Applied Sciences* **40**:18, 6332-6347. [Crossref]
39. Hao Xu, Yueru Chen, Ruiyuan Lin, C.-C. Jay Kuo. Understanding CNN via deep features analysis 1052-1060. [Crossref]
40. Adnan Haider, Philip C. Woodland. Sequence training of DNN acoustic models with natural gradient 178-184. [Crossref]
41. Yongqiang Cheng, Xiang Li, Hongqiang Wang, Xiaoqiang Hua, Yuliang Qin. 2017. Bayesian Nonlinear Filtering via Information Geometric Optimization. *Entropy* **19**:12, 655. [Crossref]
42. Suguru Kanoga, Yasue Mitsukura. Review of Artifact Rejection Methods for Electroencephalographic Systems . [Crossref]
43. Yusuke Nomura, Andrew S. Darmawan, Youhei Yamaji, Masatoshi Imada. 2017. Restricted Boltzmann machine learning for solving strongly correlated quantum systems. *Physical Review B* **96**:20. . [Crossref]
44. Yohei Kondo, Kohei Hayashi, Shin-ichi Maeda. 2017. Sparse Bayesian linear regression with latent masking variables. *Neurocomputing* **258**, 3-12. [Crossref]
45. Todd K. Moon, Jacob H. Gunther. 2017. Contravariant Adaptation on the Manifold of Causal, FIR, Invertible Multivariable Matrix Systems. *IEEE Signal Processing Letters* **24**:10, 1572-1576. [Crossref]
46. Jemin Hwangbo, Inkyu Sa, Roland Siegwart, Marco Hutter. 2017. Control of a Quadrotor With Reinforcement Learning. *IEEE Robotics and Automation Letters* **2**:4, 2096-2103. [Crossref]
47. Lei Huang, Xianglong Liu, Yang Liu, Bo Lang, Dacheng Tao. Centered Weight Normalization in Accelerating Training of Deep Neural Networks 2822-2830. [Crossref]
48. Yoshimasa Fukuoka, Yasuyuki Okumura, Katsuyuki Fujii. Polarization rotation compensation in POLMUX-DD OFDMA-PON using ICA and adaptive filter 1-2. [Crossref]
49. Weili Guo Haikun Wei, Tianhong Liu, Aiguo Song, Kanjian Zhang. An adaptive natural gradient method with adaptive step size in multilayer perceptrons 1593-1597. [Crossref]

50. Frank Critchley, Paul Marriott. 2017. On the Limiting Behaviour of the Fundamental Geodesics of Information Geometry. *Entropy* **19**:10, 524. [Crossref]
51. M Jyothirmayi, S. Sethu Selvi. Block based Kalman filter algorithm for blind image separation using sparsity measure 1314-1320. [Crossref]
52. Claudius Strub, Gregor Schöner, Florentin Wörgötter, Yulia Sandamirskaya. 2017. Dynamic Neural Fields with Intrinsic Plasticity. *Frontiers in Computational Neuroscience* **11**. . [Crossref]
53. Zaid Albataineh, Fathi M. Salem. 2017. Adaptive Blind CDMA Receivers Based on ICA Filtered Structures. *Circuits, Systems, and Signal Processing* **36**:8, 3320-3348. [Crossref]
54. Luca Martino, Valero Laparra, Gustau Camps-Valls. Probabilistic cross-validation estimators for Gaussian process regression 823-827. [Crossref]
55. V. A. Ponomarev, M. V. Pronina, Yu. D. Kropotov. 2017. Dynamics of the EEG spectral density in the θ, α, and β bands in the visual Go/NoGo task. *Human Physiology* **43**:4, 366-376. [Crossref]
56. Karl Friston. 2017. The Emperor's new topology. *Physics of Life Reviews* **21**, 26-28. [Crossref]
57. Mahdi Nazemi, Shahin Nazarian, Massoud Pedram. High-performance FPGA implementation of equivariant adaptive separation via independence algorithm for Independent Component Analysis 25-28. [Crossref]
58. Yubo Li, Yongqiang Cheng, Xiang Li, Hongqiang Wang, Xiaoqiang Hua, Yuliang Qin. Information geometric approach for nonlinear filtering 1211-1216. [Crossref]
59. Jun Zhu, Jianfei Chen, Wenbo Hu, Bo Zhang. 2017. Big Learning with Bayesian methods. *National Science Review* **4**:4, 627-651. [Crossref]
60. Jing Wang, Ioannis Ch. Paschalidis. 2017. An Actor-Critic Algorithm With Second-Order Actor and Critic. *IEEE Transactions on Automatic Control* **62**:6, 2689-2703. [Crossref]
61. Tian Tian. Weighted orthogonal constrained maximum likelihood ICA algorithm and its application in image feature extraction 98-102. [Crossref]
62. Todd K. Moon, Jacob H. Gunther. 2017. Contravariant Adaptation on the Manifold of Invertible Matrix Transfer Functions. *IEEE Signal Processing Letters* **24**:6, 756-759. [Crossref]
63. Minh-Ngoc Tran, David J. Nott, Robert Kohn. 2017. Variational Bayes With Intractable Likelihood. *Journal of Computational and Graphical Statistics* **164**, 1-10. [Crossref]
64. Zhi-Yong Ran, Bao-Gang Hu. 2017. Parameter Identifiability in Statistical Machine Learning: A Review. *Neural Computation* **29**:5, 1151-1203. [Abstract] [Full Text] [PDF] [PDF Plus]
65. Sarah E. Marzen, Simon DeDeo. 2017. The evolution of lossy compression. *Journal of The Royal Society Interface* **14**:130, 20170166. [Crossref]
66. Athanasios S. Polydoros, Lazaros Nalpantidis. 2017. Survey of Model-Based Reinforcement Learning: Applications on Robotics. *Journal of Intelligent & Robotic Systems* **86**:2, 153-173. [Crossref]
67. Matthew Sheckells, Gowtham Garimella, Marin Kobilarov. Robust policy search with applications to safe vehicle navigation 2343-2349. [Crossref]
68. Toshio Uchiyama. Information-Theoretic Clustering and Algorithms . [Crossref]
69. Yuki Yoshida, Ryo Karakida, Masato Okada, Shun-ichi Amari. 2017. Statistical Mechanical Analysis of Online Learning with Weight Normalization in Single Layer Perceptron. *Journal of the Physical Society of Japan* **86**:4, 044002. [Crossref]
70. David M. Blei, Alp Kucukelbir, Jon D. McAuliffe. 2017. Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association* **112**:518, 859-877. [Crossref]
71. Gabriel Goh. 2017. Why Momentum Really Works. *Distill* **2**:4. . [Crossref]
72. Ali Punjani, Marcus A. Brubaker, David J. Fleet. 2017. Building Proteins in a Day: Efficient 3D Molecular Structure Estimation with Electron Cryomicroscopy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**:4, 706-718. [Crossref]
73. Tsuyoshi Ide, Takayuki Katsuki, Tetsuro Morimura, Robert Morris. 2017. City-Wide Traffic Flow Estimation From a Limited Number of Low-Quality Cameras. *IEEE Transactions on Intelligent Transportation Systems* **18**:4, 950-959. [Crossref]
74. Rahul Rahul-Mark Fonseca, Bailian Chen, Jan Dirk Jansen, Albert Reynolds. 2017. A Stochastic Simplex Approximate Gradient (StoSAG) for optimization under uncertainty. *International Journal for Numerical Methods in Engineering* **109**:13, 1756-1776. [Crossref]

75. Jiayu Chen, Vince D. Calhoun, Godfrey D. Pearlson, Nora I. Perrone-Bizzozero, Jessica A. Turner, Stefan Ehrlich, Beng-Choon Ho, Jingyu Liu. 2017. Independent component analysis of SNPs reflects polygenic risk scores for schizophrenia. *Schizophrenia Research* **181**, 83-85. [Crossref]
76. Chenyang Tao, Jianfeng Feng. 2017. Canonical kernel dimension reduction. *Computational Statistics & Data Analysis* **107**, 131-148. [Crossref]
77. Philippe Tillet, H. T. Kung, David Cox. Infomax-ICA using Hessian-free optimization 2537-2541. [Crossref]
78. Yusuke Tajiri, Hirokazu Kameoka, Tomoki Toda. A noise suppression method for body-conducted soft speech based on non-negative tensor factorization of air- and body-conducted signals 4960-4964. [Crossref]
79. Motoi Omachi, Tetsuji Ogawa, Tetsunori Kobayashi. 2017. Associative Memory Model-Based Linear Filtering and Its Application to Tandem Connectionist Blind Source Separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **25**:3, 637-650. [Crossref]
80. Deepa Thiyam, Sergio Cruces, Javier Olias, Andrzej Cichocki. 2017. Optimization of Alpha-Beta Log-Det Divergences and their Application in the Spatial Filtering of Two Class Motor Imagery Movements. *Entropy* **19**:3, 89. [Crossref]
81. Murat A. Erdogdu. Generalized Hessian approximations via Stein's lemma for constrained minimization 1-8. [Crossref]
82. Yubo Li, Yongqiang Cheng, Xiang Li, Xiaoqiang Hua, Yuliang Qin. 2017. Information Geometric Approach to Recursive Update in Nonlinear Filtering. *Entropy* **19**:2, 54. [Crossref]
83. Jair Montoya-Martínez, Jean-François Cardoso, Alexandre Gramfort. Caveats with Stochastic Gradient and Maximum Likelihood Based ICA for EEG 279-289. [Crossref]
84. Christophe Saint-Jean, Frank Nielsen. Batch and Online Mixture Learning: A Review with Extensions 267-299. [Crossref]
85. Jörn Anemüller, Hendrik Kayser. Acoustic Source Localization by Combination of Supervised Direction-of-Arrival Estimation with Disjoint Component Analysis 99-108. [Crossref]
86. Frank Critchley, Paul Marriott. Information Geometry and Its Applications: An Overview 1-31. [Crossref]
87. Pankaj Pratap Singh, R. D. Garg. On Sphering the High Resolution Satellite Image Using Fixed Point Based ICA Approach 411-419. [Crossref]
88. Wang Chuanchuan, Xu Jiaqi, Zeng Yonghu. 2017. Research on Influence of Source Number Estimation on Application of Blind Source Separation Algorithms. *Procedia Computer Science* **107**, 379-384. [Crossref]
89. Jürgen Schmidhuber. Deep Learning 338-348. [Crossref]
90. Jose Maria Giron-Sierra. Data Analysis and Classification 647-835. [Crossref]
91. Xiaomo Jiang, Sankaran Mahadevan, Yong Yuan. 2017. Fuzzy stochastic neural network model for structural system identification. *Mechanical Systems and Signal Processing* **82**, 394-411. [Crossref]
92. Abhijeet Bishnu, Vimal Bhatia. 2017. Sparse Channel Estimation for Interference Limited OFDM Systems and Its Convergence Analysis. *IEEE Access* **5**, 17781-17794. [Crossref]
93. Ryo Iwaki, Minoru Asada. Implicit Incremental Natural Actor Critic 749-758. [Crossref]
94. Trang-Anh Nghiem, Olivier Marre, Alain Destexhe, Ulisse Ferrari. Pairwise Ising Model Analysis of Human Cortical Neuron Recordings 257-264. [Crossref]
95. Gaétan Marceau-Caron, Yann Ollivier. Natural Langevin Dynamics for Neural Networks 451-459. [Crossref]
96. Florian Wenzel, Théo Galy-Fajou, Matthäus Deutsch, Marius Kloft. Bayesian Nonlinear Support Vector Machines for Big Data 307-322. [Crossref]
97. Linda S. L. Tan. 2017. Stochastic variational inference for large-scale discrete choice models using adaptive batch sizes. *Statistics and Computing* **27**:1, 237-257. [Crossref]
98. Nihat Ay, Jürgen Jost, Hông Vân Lê, Lorenz Schwachhöfer. Fields of Application of Information Geometry 295-360. [Crossref]
99. Xi-Lin Li. 2017. Preconditioned Stochastic Gradient Descent. *IEEE Transactions on Neural Networks and Learning Systems* 1-13. [Crossref]
100. Gang Chen, Colin I. J. Douch, Mengjie Zhang. 2016. Accuracy-Based Learning Classifier Systems for Multistep Reinforcement Learning: A Fuzzy Logic Approach to Handling Continuous Inputs and Learning Continuous Actions. *IEEE Transactions on Evolutionary Computation* **20**:6, 953-971. [Crossref]
101. Andrew Lamperski. Natural gradients for state and output feedback control 1984-1989. [Crossref]
102. Chunchen Liu, Lu Feng, Ryohei Fujimaki. Streaming Model Selection via Online Factorized Asymptotic Bayesian Inference 271-280. [Crossref]

103. Tomohiro Tanaka, Takafumi Moriya, Takahiro Shinozaki, Shinji Watanabe, Takaaki Hori, Kevin Duh. Automated structure discovery and parameter tuning of neural network language model based on evolution strategy 665-671. [Crossref]
104. Chunhui Li, Erchuan Zhang, Lin Jiu, Huafei Sun. 2016. Optimal control on special Euclidean group via natural gradient algorithm. *Science China Information Sciences* **59**:11. . [Crossref]
105. Linda S. L. Tan, Victor M. H. Ong, David J. Nott, Ajay Jasra. 2016. Variational inference for sparse spectrum Gaussian process regression. *Statistics and Computing* **26**:6, 1243-1261. [Crossref]
106. Alberto Prieto, Beatriz Prieto, Eva Martinez Ortigosa, Eduardo Ros, Francisco Pelayo, Julio Ortega, Ignacio Rojas. 2016. Neural networks: An overview of early research, current frameworks and new challenges. *Neurocomputing* **214**, 242-268. [Crossref]
107. Qi Liu, Tian Tan, Kai Yu. An investigation on deep learning with beta stabilizer 557-561. [Crossref]
108. Yongqiang Cheng, Xiaoqiang Hua, Hongqiang Wang, Yuliang Qin, Xiang Li. 2016. The Geometry of Signal Detection with Applications to Radar Signal Processing. *Entropy* **18**:11, 381. [Crossref]
109. Rogers F. Silva, Sergey M. Plis, Jing Sui, Marios S. Pattichis, Tulay Adali, Vince D. Calhoun. 2016. Blind Source Separation for Unimodal and Multimodal Brain Networks: A Unifying Framework for Subspace Modeling. *IEEE Journal of Selected Topics in Signal Processing* **10**:7, 1134-1149. [Crossref]
110. Hugo Jacquin, A. Rançon. 2016. Resummed mean-field inference for strongly coupled data. *Physical Review E* **94**:4. . [Crossref]
111. Shohei Shimizu. Non-Gaussian Structural Equation Models for Causal Discovery 153-184. [Crossref]
112. . Independent Component Analysis: An Overview 215-226. [Crossref]
113. Keisuke Imoto, Nobutaka Ono. Online acoustic scene analysis based on nonparametric Bayesian model 988-992. [Crossref]
114. Ulisse Ferrari. 2016. Learning maximum entropy models from finite-size data sets: A fast data-driven algorithm allows sampling from the posterior distribution. *Physical Review E* **94**:2. . [Crossref]
115. Ming Tu, Visar Berisha, Yu Cao, Jae-Sun Seo. Reducing the Model Order of Deep Neural Networks Using Information Theory 93-98. [Crossref]
116. Hwa Jeon Song, Ho Young Jung, Jeon Gue Park. 2016. Implementation of CNN in the view of mini-batch DNN training for efficient second order optimization. *Journal of the Korean society of speech sciences* **8**:2, 23-30. [Crossref]
117. Wei-Xin Li, Nuno Vasconcelos. 2016. Complex Activity Recognition Via Attribute Dynamics. *International Journal of Computer Vision* **43**. . [Crossref]
118. Pierre Baque, Timur Bagautdinov, Francois Fleuret, Pascal Fua. Principled Parallel Mean-Field Inference for Discrete Random Fields 5848-5857. [Crossref]
119. A. Cascardi, F. Micelli, M.A. Aiello. 2016. Analytical model based on artificial neural network for masonry shear walls strengthened with FRM systems. *Composites Part B: Engineering* **95**, 252-263. [Crossref]
120. Vagelis Harmandaris, Evangelia Kalligiannaki, Markos Katsoulakis, Petr Plecháč. 2016. Path-space variational inference for non-equilibrium coarse-grained systems. *Journal of Computational Physics* **314**, 355-383. [Crossref]
121. Andreas S. Stordal, Slawomir P. Szklarz, Olwijn Leeuwenburgh. 2016. A Theoretical look at Ensemble-Based Optimization in Reservoir Management. *Mathematical Geosciences* **48**:4, 399-417. [Crossref]
122. Judith Buse, Christian Beste, Elisabeth Herrmann, Veit Roessner. 2016. Neural correlates of altered sensorimotor gating in boys with Tourette Syndrome: A combined EMG/fMRI study. *The World Journal of Biological Psychiatry* **17**:3, 187-197. [Crossref]
123. Xi-Ming Li, Ji-Hong Ouyang. 2016. Tuning the Learning Rate for Stochastic Variational Inference. *Journal of Computer Science and Technology* **31**:2, 428-436. [Crossref]
124. Zhengwei Wu, Saleem A. Kassam, Visa Koivunen. Relative-gradient Bussgang-type blind equalization algorithms 2459-2463. [Crossref]
125. S. R Mir Alavi, W. Bastiaan Kleijn. Distributed linear blind source separation over wireless sensor networks with arbitrary connectivity patterns 3171-3175. [Crossref]
126. Yusuke Tajiri, Tomoki Toda, Satoshi Nakamura. Noise suppression method for body-conducted soft speech enhancement based on external noise monitoring 5935-5939. [Crossref]
127. Mario Coutino, Radmila Pribic, Geert Leus. Direction of arrival estimation based on information geometry 3066-3070. [Crossref]
128. Yannick Deville. Blind Source Separation and Blind Mixture Identification Methods 1-33. [Crossref]
129. Junhao Hua, Chunguang Li. 2016. Distributed Variational Bayesian Algorithms Over Sensor Networks. *IEEE Transactions on Signal Processing* **64**:3, 783-798. [Crossref]

130. Shanshan Li, Shaojie Chen, Chen Yue, Brian Caffo. 2016. A Parcellation Based Nonparametric Algorithm for Independent Component Analysis with Application to fMRI Data. *Frontiers in Neuroscience* **10**. . [Crossref]
131. Marc Z. Miskin, Gurdaman Khaira, Juan J. de Pablo, Heinrich M. Jaeger. 2016. Turning statistical physics models into materials design engines. *Proceedings of the National Academy of Sciences* **113**:1, 34-39. [Crossref]
132. Yong Tian, Yu Rong, Yuan Yao, Weidong Liu, Jiaxing Song. Grouped Text Clustering Using Non-Parametric Gaussian Mixture Experts 507-516. [Crossref]
133. Bo Ma, Hongwei Hu, Jianbing Shen, Yangbiao Liu, Ling Shao. 2016. Generalized Pooling for Robust Object Tracking. *IEEE Transactions on Image Processing* 1-1. [Crossref]
134. Jürgen Schmidhuber. Deep Learning 1-11. [Crossref]
135. Ryo Karakida, Masato Okada, Shun-ichi Amari. Adaptive Natural Gradient Learning Algorithms for Unnormalized Statistical Models 427-434. [Crossref]
136. Vladimir Gligorijevic, Yannis Panagakis, Stefanos Zafeiriou. Fusion and community detection in multi-layer graphs 1327-1332. [Crossref]
137. Zhijian Luo, Danping Liao, Yuntao Qian. Bound analysis of natural gradient descent in stochastic optimization setting 4166-4171. [Crossref]
138. R. H. Byrd, S. L. Hansen, Jorge Nocedal, Y. Singer. 2016. A Stochastic Quasi-Newton Method for Large-Scale Optimization. *SIAM Journal on Optimization* **26**:2, 1008-1031. [Crossref]
139. Yusuke Uchida, Shigeyuki Sakazawa, Shin'ichi Satoh. 2016. [Paper] Image Retrieval with Fisher Vectors of Binary Features. *ITE Transactions on Media Technology and Applications* **4**:4, 326-336. [Crossref]
140. Silvere Bonnabel, Axel Barrau. An intrinsic Cramér-Rao bound on SO(3) for (dynamic) attitude filtering 2158-2163. [Crossref]
141. Takafumi Moriya, Tomohiro Tanaka, Takahiro Shinozaki, Shinji Watanabe, Kevin Duh. Automation of system building for state-of-the-art large vocabulary speech recognition using evolution strategy 610-616. [Crossref]
142. Pankaj Pratap Singh, R. D. Garg. 2015. Fixed Point ICA Based Approach for Maximizing the Non-gaussianity in Remote Sensing Image Classification. *Journal of the Indian Society of Remote Sensing* **43**:4, 851-858. [Crossref]
143. Pablo Zegers. 2015. Fisher Information Properties. *Entropy* **17**:12, 4918-4939. [Crossref]
144. B. Frieden. 2015. Estimating a Repeatable Statistical Law by Requiring Its Stability During Observation. *Entropy* **17**:12, 7453-7467. [Crossref]
145. Kazunori Nakada, Kiyotaka Suzuki, Tsutomu Nakada. 2015. Single to Two Cluster State Transition of Primary Motor Cortex 4-posterior (MI-4p) Activities in Humans. *Entropy* **17**:12, 7596-7607. [Crossref]
146. Luigi Malagò, Giovanni Pistone. 2015. Natural Gradient Flow in the Mixture Geometry of a Discrete Exponential Family. *Entropy* **17**:12, 4215-4254. [Crossref]
147. Rui Vigelis, Charles Cavalcante. Information Geometry: An Introduction to New Models for Signal Processing 455-491. [Crossref]
148. Dan Xu, Quan Liu. ACIS: An Improved Actor-Critic Method for POMDPs with Internal State 369-376. [Crossref]
149. Marijn F. Stollenga, Alan J. Lockett, Jurgen Schmidhuber. The Natural Gradient as a control signal for a humanoid robot 187-193. [Crossref]
150. Jiayu Chen, Vince D. Calhoun, Alejandro Arias-Vasquez, Marcel P. Zwiers, Kimm van Hulzen, Guillén Fernández, Simon E. Fisher, Barbara Franke, Jessica A. Turner, Jingyu Liu. 2015. G-protein genomic association with normal variation in gray matter density. *Human Brain Mapping* **36**:11, 4272-4286. [Crossref]
151. Hua Yang, Hang Zhang, Liu Yang. Gradient optimized blind sources separation algorithm 1-6. [Crossref]
152. Pengcheng Xu, Yuehong Shen, Wei Jian, Wei Zhao, Cheng Peng. 2015. Maximization of Nonlinear Autocorrelation for Blind Source Separation of Non-stationary Complex Signals. *Circuits, Systems, and Signal Processing* **34**:9, 3011-3029. [Crossref]
153. James Hensman, Panagiotis Papastamoulis, Peter Glaus, Antti Honkela, Magnus Rattray. 2015. Fast and accurate approximate inference of transcript expression from RNA-seq data. *Bioinformatics* btv483. [Crossref]
154. Zhi-Yong Ran, Bao-Gang Hu. 2015. An identifying function approach for determining parameter structure of statistical learning machines. *Neurocomputing* **162**, 209-217. [Crossref]
155. Sheng-Hsiou Hsu, Luca Pion-Tonachini, Tzyy-Ping Jung, Gert Cauwenberghs. Tracking non-stationary EEG sources using adaptive online recursive independent component analysis 4106-4109. [Crossref]
156. Xin Wang, Shifeng Ou, Ying Gao, Xiaofeng Guo. A new fast nonlinear principal component analysis algorithm for blind source separation 1626-1630. [Crossref]

157. Jiayu Chen, Kent E. Hutchison, Vince D. Calhoun, Eric D. Claus, Jessica A. Turner, Jing Sui, Jingyu Liu. 2015. CREB-BDNF pathway influences alcohol cue-elicited activation in drinkers. *Human Brain Mapping* **36**:8, 3007-3019. [Crossref]

158. Zhao Junsheng, Xia Jianwei, Zhuang Guangming, Zhang Huasheng. Natural gradient learning algorithms for nonlinear systems 1979-1983. [Crossref]

159. Seiya Satoh, Ryohei Nakano. Complex-valued multilayer perceptron learning using singular regions and search pruning 1-6. [Crossref]

160. Jorge Sánchez, Javier Redolfi. 2015. Exponential family Fisher vector for image classification. *Pattern Recognition Letters* **59**, 26-32. [Crossref]

161. Panos Toulis, Edoardo M. Airoldi. 2015. Scalable estimation strategies based on stochastic approximations: classical results and new insights. *Statistics and Computing* **25**:4, 781-795. [Crossref]

162. Kazuyuki Hara, Daisuke Saito, Hayaru Shouno. Analysis of function of rectified linear unit used in deep learning 1-8. [Crossref]

163. Y. Ollivier. 2015. Riemannian metrics for neural networks II: recurrent networks and learning symbolic data sequences. *Information and Inference* **4**:2, 154-193. [Crossref]

164. Y. Ollivier. 2015. Riemannian metrics for neural networks I: feedforward networks. *Information and Inference* **4**:2, 108-153. [Crossref]

165. Junsheng Zhao, Xingjiang Yu. 2015. Adaptive natural gradient learning algorithms for Mackey–Glass chaotic time prediction. *Neurocomputing* **157**, 41-45. [Crossref]

166. Xi-ming Li, Ji-hong Ouyang, You Lu. 2015. Topic modeling for large-scale text data. *Frontiers of Information Technology & Electronic Engineering* **16**:6, 457-465. [Crossref]

167. Ali Mohammad-Djafari. 2015. Entropy, Information Theory, Information Geometry and Bayesian Inference in Data, Signal and Image Processing and Inverse Problems. *Entropy* **17**:6, 3989-4027. [Crossref]

168. Tohru Nitta. 2015. Learning Dynamics of a Single Polar Variable Complex-Valued Neuron. *Neural Computation* **27**:5, 1120-1141. [Abstract] [Full Text] [PDF] [PDF Plus]

169. Nathan Ratliff, Marc Toussaint, Stefan Schaal. Understanding the geometry of workspace obstacles in Motion Optimization 4202-4209. [Crossref]

170. Naoki Hiratani, Tomoki Fukai. 2015. Mixed Signal Learning by Spike Correlation Propagation in Feedback Inhibitory Circuits. *PLOS Computational Biology* **11**:4, e1004227. [Crossref]

171. Muhammed O. Sayin, Yasin Yilmaz, Alper Demir, Suleyman S. Kozat. 2015. The Krylov-proportionate normalized least mean fourth approach: Formulation and performance analysis. *Signal Processing* **109**, 1-13. [Crossref]

172. Xiankai Lu, Zheng Fang, Tao Xu, Haiting Zhang, Hongya Tuo. Efficient image categorization with sparse Fisher vector 1498-1502. [Crossref]

173. Nihat Ay. 2015. Information Geometry on Complexity and Stochastic Interaction. *Entropy* **17**:4, 2432-2458. [Crossref]

174. Hesam Mahmoudi, Mohammad Mehdi Homayounpour. A Persian spoken dialogue system using POMDPs 217-221. [Crossref]

175. Garvesh Raskutti, Sayan Mukherjee. 2015. The Information Geometry of Mirror Descent. *IEEE Transactions on Information Theory* **61**:3, 1451-1457. [Crossref]

176. Harold Soh, Yiannis Demiris. 2015. Spatio-Temporal Learning With the Online Finite and Infinite Echo-State Gaussian Processes. *IEEE Transactions on Neural Networks and Learning Systems* **26**:3, 522-536. [Crossref]

177. Yossi Bar-Yosef, Yuval Bistritz. 2015. Gaussian Mixture Models Reduction by Variational Maximum Mutual Information. *IEEE Transactions on Signal Processing* **63**:6, 1557-1569. [Crossref]

178. Junsheng Zhao, Haikun Wei, Chi Zhang, Weiling Li, Weili Guo, Kanjian Zhang. 2015. Natural Gradient Learning Algorithms for RBF Networks. *Neural Computation* **27**:2, 481-505. [Abstract] [Full Text] [PDF] [PDF Plus]

179. N. Ampazis, S. J. Perantonis, D. Drivaliaris. 2015. Improved Jacobian Eigen-Analysis Scheme for Accelerating Learning in Feedforward Neural Networks. *Cognitive Computation* **7**:1, 86-102. [Crossref]

180. Ye Liu, Mingfen Li, Yi Wu, Jie Jia, Liqing Zhang. 2015. Separation and Recognition of Electroencephalogram Patterns Using Temporal Independent Component Analysis. *International Journal of Pattern Recognition and Artificial Intelligence* **29**:01, 1550001. [Crossref]

181. James Hensman, Magnus Rattray, Neil D. Lawrence. 2015. Fast Nonparametric Clustering of Structured Time-Series. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **37**:2, 383-393. [Crossref]

182. Gungor Polatkan, Mingyuan Zhou, Lawrence Carin, David Blei, Ingrid Daubechies. 2015. A Bayesian Nonparametric Approach to Image Super-Resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **37**:2, 346-358. [Crossref]

183. John Paisley, Chong Wang, David M. Blei, Michael I. Jordan. 2015. Nested Hierarchical Dirichlet Processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **37**:2, 256-270. [Crossref]

184. Jacinto C. Nascimento, Miguel Barão, Jorge S. Marques, João M. Lemos. 2015. An information geometric framework for the optimization on a discrete probability spaces: Application to human trajectory classification. *Neurocomputing* **150**, 155-162. [Crossref]

185. Voot Tangkaratt, Ning Xie, Masashi Sugiyama. 2015. Conditional Density Estimation with Dimensionality Reduction via Squared-Loss Conditional Entropy Minimization. *Neural Computation* **27**:1, 228-254. [Abstract] [Full Text] [PDF] [PDF Plus]

186. Ignacio Segovia-Dominguez, Arturo Hernandez-Aguirre. An Estimation of Distribution Algorithm based on the Natural Gradient and the Boltzmann Distribution 527-534. [Crossref]

187. Carlos Klein, Emerson Vasconcelos Segundo, Viviana Mariani, Leandro Coelho. 2015. Modified Social-Spider Optimization Algorithm Applied to Electromagnetic Optimization. *IEEE Transactions on Magnetics* 1-1. [Crossref]

188. Pengcheng Xu, Yuehong Shen. 2015. A Fast Algorithm for Blind Separation of Complex Valued Signals with Nonlinear Autocorrelation. *Journal of Communications* **10**:3, 170-177. [Crossref]

189. Shinichi Shirakawa, Youhei Akimoto, Kazuki Ouchi, Kouzou Ohara. Sample Reuse in the Covariance Matrix Adaptation Evolution Strategy Based on Importance Sampling 305-312. [Crossref]

190. Jürgen Schmidhuber. 2015. Deep learning in neural networks: An overview. *Neural Networks* **61**, 85-117. [Crossref]

191. Pengcheng Xu, Zhigang Yuan, Wei Jian, Wei Zhao. 2015. Variable Step-Size Method Based on a Reference Separation System for Source Separation. *Journal of Sensors* **2015**, 1-7. [Crossref]

192. Luigi Malagò, Giovanni Pistone. Information Geometry of the Gaussian Distribution in View of Stochastic Optimization 150-162. [Crossref]

193. Nikolaus Hansen, Dirk V. Arnold, Anne Auger. Evolution Strategies 871-898. [Crossref]

194. Sergios Theodoridis. Parameter Learning 327-402. [Crossref]

195. Rostom Mabrouk, Pablo M. Rusjan, Romina Mizrahi, Mark F. Jacobs, Yuko Koshimori, Sylvain Houle, Ji Hyun Ko, Antonio P. Strafella. 2014. Image Derived Input Function for [18F]-FEPPA: Application to Quantify Translocator Protein (18 kDa) in the Human Brain. *PLoS ONE* **9**:12, e115768. [Crossref]

196. Hans-Georg Beyer. 2014. Convergence Analysis of Evolutionary Algorithms That Are Based on the Paradigm of Information Geometry. *Evolutionary Computation* **22**:4, 679-709. [Abstract] [Full Text] [PDF] [PDF Plus]

197. Jacinto C. Nascimento, Miguel Barao, Jorge S. Marques, Joao M. Lemos. 2014. Information Geometric Algorithm for Estimating Switching Probabilities in Space-Varying HMM. *IEEE Transactions on Image Processing* **23**:12, 5263-5273. [Crossref]

198. Zhiqiang Xu, Yiping Ke, Yi Wang. A Fast Inference Algorithm for Stochastic Blockmodel 620-629. [Crossref]

199. Chih-Wei Jen, Shyh-Jye Jou. 2014. Blind ICA detection based on second-order cone programming for MC-CDMA systems. *EURASIP Journal on Advances in Signal Processing* **2014**:1. . [Crossref]

200. Shahab Faiz Minhas, Patrick Gaydecki. 2014. A hybrid algorithm for blind source separation of a convolutive mixture of three speech sources. *EURASIP Journal on Advances in Signal Processing* **2014**:1. . [Crossref]

201. Jing Wang, Ioannis Ch. Paschalidis. A Hessian actor-critic algorithm 1131-1136. [Crossref]

202. Fei Zhao, Ningyun Lu, Jianhua Lu. 2014. Quality Control of Batch Processes Using Natural Gradient Based Model-Free Optimization. *Industrial & Engineering Chemistry Research* **53**:44, 17419-17428. [Crossref]

203. Wei Li, Huizhong Yang. 2014. A Non-Linear Blind Source Separation Method Based on Perceptron Structure and Conjugate Gradient Algorithm. *Circuits, Systems, and Signal Processing* **33**:11, 3573-3590. [Crossref]

204. Keith D. Gilbert, Karen L. Payton. Competitive algorithm blending for enhanced source separation 450-454. [Crossref]

205. Harold Soh, Yiannis Demiris. 2014. Incrementally Learning Objects by Touch: Online Discriminative and Generative Models for Tactile-Based Recognition. *IEEE Transactions on Haptics* **7**:4, 512-525. [Crossref]

206. Zhi-Yong Ran, Bao-Gang Hu. 2014. Determining parameter identifiability from the optimization theory framework: A Kullback–Leibler divergence approach. *Neurocomputing* **142**, 307-317. [Crossref]

207. Junsheng Zhao, Haikun Wei, Weili Guo, Kanjian Zhang. 2014. Singularities in the identification of dynamic systems. *Neurocomputing* **140**, 339-344. [Crossref]

208. Weili Guo, Haikun Wei, Junsheng Zhao, Kanjian Zhang. 2014. Averaged learning equations of error-function-based multilayer perceptrons. *Neural Computing and Applications* **25**:3-4, 825-832. [Crossref]

209. Pengcheng Xu, Yuehong Shen, Qiao Su. Blind source separation with variable step-size method based on a reference separation system 110-114. [Crossref]

210. Geng-Shen Fu, Ronald Phlypo, Matthew Anderson, Xi-Lin Li, Tulay Adali. 2014. Blind Source Separation by Entropy Rate Minimization. *IEEE Transactions on Signal Processing* **62**:16, 4245-4255. [Crossref]
211. K. J. A. Cox, P. R. Adams. 2014. Hebbian learning from higher-order correlations requires crosstalk minimization. *Biological Cybernetics* **108**:4, 405-422. [Crossref]
212. Junzhao Du, Sicong Liu, Chi Xu, Kai Wang, Hui Liu, Kewei Sha. Lightweight construction of the information potential field in wireless sensor networks 1-8. [Crossref]
213. Luigi Malagò, Giovanni Pistone. 2014. Combinatorial Optimization with Information Geometry: The Newton Method. *Entropy* **16**:8, 4260-4289. [Crossref]
214. Zhenning Zhang, Huafei Sun, Linyu Peng, Lin Jiu. 2014. A Natural Gradient Algorithm for Stochastic Distribution Systems. *Entropy* **16**:8, 4338-4352. [Crossref]
215. Danilo Comminiello, Michele Scarpiniti, Luis A. Azpicueta-Ruiz, Jeronimo Arenas-Garcia, Aurelio Uncini. 2014. Nonlinear Acoustic Echo Cancellation Based on Sparse Functional Link Representations. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **22**:7, 1172-1183. [Crossref]
216. Zonghai Sun. Gaussian Process adaptive control of nonlinear system base on online algorithm 8791-8794. [Crossref]
217. Junsheng Zhao, Haikun Wei, Weili Guo, Kanjian Zhang. Singularities in the parameter identification of nonlinear dynamic systems 6610-6614. [Crossref]
218. Zhenning Zhang, Xu Zhao, Juntong Zhou. An information geometric algorithm for multi-input and multi-output stochastic distribution control systems with output feedback vector 5340-5344. [Crossref]
219. . Probability density function based reliability evaluation of large-scale ICs 157-162. [Crossref]
220. Salem Said, Lionel Bombrun, Yannick Berthoumieu. 2014. New Riemannian Priors on the Univariate Normal Model. *Entropy* **16**:7, 4015-4031. [Crossref]
221. Xuezhi Wang, Yongqiang Cheng, Bill Moran. Nonlinear parameter estimation in statistical manifolds 101-104. [Crossref]
222. L. R. Arnaut, C. S. Obiekezie. 2014. Source Separation for Wideband Energy Emissions Using Complex Independent Component Analysis. *IEEE Transactions on Electromagnetic Compatibility* **56**:3, 559-570. [Crossref]
223. Jiayu Chen, Jingyu Liu, Vince D. Calhoun, Alejandro Arias-Vasquez, Marcel P. Zwiers, Cota Navin Gupta, Barbara Franke, Jessica A. Turner. 2014. Exploration of scanning effects in multi-site structural MRI studies. *Journal of Neuroscience Methods* **230**, 37-50. [Crossref]
224. Zhuowei Cai, Limin Wang, Xiaojiang Peng, Yu Qiao. Multi-view Super Vector for Action Recognition 596-603. [Crossref]
225. Guido Montúfar, Johannes Rauh, Nihat Ay. 2014. On the Fisher Metric of Conditional Probability Polytopes. *Entropy* **16**:6, 3207-3233. [Crossref]
226. Shinji Watanabe, Jonathan Le Roux. Black box optimization for automatic speech recognition 3256-3260. [Crossref]
227. Aurelien Schutz, Lionel Bombrun, Yannick Berthoumieu. Intrinsic prior for Bayesian classification of texture images 4359-4363. [Crossref]
228. Paul Wagner. 2014. Policy oscillation is overshooting. *Neural Networks* **52**, 43-61. [Crossref]
229. Xin Xu, Lei Zuo, Zhenhua Huang. 2014. Reinforcement learning algorithms with function approximation: Recent advances and applications. *Information Sciences* **261**, 1-31. [Crossref]
230. Sven Hoffmann, Franziska Labrenz, Maria Themann, Edmund Wascher, Christian Beste. 2014. Crosslinking EEG time–frequency decomposition and fMRI in error monitoring. *Brain Structure and Function* **219**:2, 595-605. [Crossref]
231. Wentao Fan, Nizar Bouguila. 2014. Online variational learning of generalized Dirichlet mixture models with feature selection. *Neurocomputing* **126**, 166-179. [Crossref]
232. . Introduction 1-15. [Crossref]
233. . Independent Component Analysis 59-84. [Crossref]
234. Sergey Kushnarev, Akil Narayan. 2014. Approximating the Weil--Petersson Metric Geodesics on the Universal Teichmüller Space by Singular Solutions. *SIAM Journal on Imaging Sciences* **7**:2, 900-923. [Crossref]
235. Xingjia Tang, Xiufang Zhang, Jimin Ye. 2014. Adaptive Step-Size Natural Gradient ICA Algorithm with Weighted Orthogonalization. *Circuits, Systems, and Signal Processing* **33**:1, 211-221. [Crossref]
236. Wei Li. 2014. Blind Signal Separation with Kernel Probability Density Estimation Based on MMI Criterion Optimized by Conjugate Gradient. *Journal of Communications* **9**:7, 579-587. [Crossref]

237. Nicoleta Cucu Laurenciu, Sorin D. Cotofana. Probability density function based reliability evaluation of large-scale ICs 157-162. [Crossref]
238. Zhaojun Yang, Angeliki Metallinou, Shrikanth Narayanan. 2014. Analysis and Predictive Modeling of Body Language Behavior in Dyadic Interactions from Multimodal Interlocutor Cues. *IEEE Transactions on Multimedia* 1-1. [Crossref]
239. Mark Crowley. 2014. Using Equilibrium Policy Gradients for Spatiotemporal Planning in Forest Ecosystem Management. *IEEE Transactions on Computers* **63**:1, 142-154. [Crossref]
240. Amitabh Mishra, Suryadip Chakraborty, Hailong Li, Dharma Agrawal. Error minimization and energy conservation by predicting data in wireless body sensor networks using artificial neural network and analysis of error 165-170. [Crossref]
241. Yohei Nakada, Makio Wakahara, Takashi Matsumoto. 2014. Online Bayesian Learning With Natural Sequential Prior Distribution. *IEEE Transactions on Neural Networks and Learning Systems* **25**:1, 40-54. [Crossref]
242. Lei Chen, Liyi Zhang, Yanju Guo, Yong Huang, Jingyi Liang. 2014. Blind Source Separation Based on Covariance Ratio and Artificial Bee Colony Algorithm. *Mathematical Problems in Engineering* **2014**, 1-12. [Crossref]
243. N.Y. Lu, F. Zhao, J.H. Lu, R.Y. Qi. 2014. Quality control of batch process using natural gradient based model-free optimization. *IFAC Proceedings Volumes* **47**:3, 8335-8340. [Crossref]
244. Jiayu Chen, Vince D. Calhoun, Godfrey D. Pearlson, Nora Perrone-Bizzozero, Jing Sui, Jessica A. Turner, Juan R. Bustillo, Stefan Ehrlich, Scott R. Sponheim, José M. Cañive, Beng-Choon Ho, Jingyu Liu. 2013. Guided exploration of genomic risk for gray matter abnormalities in schizophrenia using parallel independent component analysis with reference. *NeuroImage* **83**, 384-396. [Crossref]
245. Giovanni Pistone. 2013. Examples of the Application of Nonparametric Information Geometry to Statistical Physics. *Entropy* **15**:12, 4042-4065. [Crossref]
246. Jun Zhang. 2013. Nonparametric Information Geometry: From Divergence Function to Referential-Representational Biduality on Statistical Manifolds. *Entropy* **15**:12, 5384-5418. [Crossref]
247. Masashi Sugiyama. 2013. Machine Learning with Squared-Loss Mutual Information. *Entropy* **15**:12, 80-112. [Crossref]
248. Wentao Fan, Nizar Bouguila. 2013. Online Learning of a Dirichlet Process Mixture of Beta-Liouville Distributions Via Variational Inference. *IEEE Transactions on Neural Networks and Learning Systems* **24**:11, 1850-1862. [Crossref]
249. Yusuke Uchida, Shigeyuki Sakazawa. Image Retrieval with Fisher Vectors of Binary Features 23-28. [Crossref]
250. Dimitrios Korkinof, Yiannis Demiris. Online quantum mixture regression for trajectory learning by demonstration 3222-3229. [Crossref]
251. Ning Zhao, FangMin Li, XueDong Liu. 2013. Blind estimation of number of motion multi-human targets in wireless pyroelectric infrared sensor networks. *Infrared Physics & Technology* **61**, 208-215. [Crossref]
252. Seiya Satoh, Ryohei Nakano. 2013. Fast and Stable Learning Utilizing Singular Regions of Multilayer Perceptron. *Neural Processing Letters* **38**:2, 99-115. [Crossref]
253. Jen-Tzung Chien, Hiroshi Sawada, Shoji Makino. Adaptive processing and learning for audio source separation 1-6. [Crossref]
254. Zhenning Zhang, Huafei Sun, Linyu Peng. 2013. Natural gradient algorithm for stochastic distribution systems with output feedback. *Differential Geometry and its Applications* **31**:5, 682-690. [Crossref]
255. Cyril Furtlehner. 2013. Approximate inverse Ising models close to a Bethe reference point. *Journal of Statistical Mechanics: Theory and Experiment* **2013**:09, P09020. [Crossref]
256. Ying Tian, Wenli Du, Feng Qian. 2013. Fault Detection and Diagnosis for Non-Gaussian Processes with Periodic Disturbance Based on AMRA-ICA. *Industrial & Engineering Chemistry Research* **52**:34, 12082-12107. [Crossref]
257. Y. Bengio, A. Courville, P. Vincent. 2013. Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**:8, 1798-1828. [Crossref]
258. Pengwen Chen, Hung Hung, Osamu Komori, Su-Yun Huang, Shinto Eguchi. 2013. Robust Independent Component Analysis via Minimum <formula formulatype="inline"><tex Notation="TeX">$\gamma $</tex></formula>-Divergence Estimation. *IEEE Journal of Selected Topics in Signal Processing* **7**:4, 614-624. [Crossref]
259. Todd K. Moon, Jacob H. Gunther. Signal processing using vector space methods: An introspective 296-301. [Crossref]
260. Ching-Hung, Meng-Tzu Huang, Chih-Min Lin. Fuzzy neural-based learning rate adjustment for gradient based blind source separation 1450-1455. [Crossref]
261. Yongqiang Cheng, Xuezhi Wang, Mark Morelande, Bill Moran. 2013. Information geometry of target tracking sensor networks. *Information Fusion* **14**:3, 311-326. [Crossref]

262. K. Neumann, C. Strub, J.J. Steil. 2013. Intrinsic plasticity via natural gradient descent with application to drift compensation. *Neurocomputing* **112**, 26-33. [Crossref]
263. Babak Ahmadi, Kristian Kersting, Martin Mladenov, Sriraam Natarajan. 2013. Exploiting symmetries for scaling loopy belief propagation and relational training. *Machine Learning* **92**:1, 91-132. [Crossref]
264. Ani Eloyan, Ciprian M. Crainiceanu, Brian S. Caffo. 2013. Likelihood-based population independent component analysis. *Biostatistics* **14**:3, 514-527. [Crossref]
265. Rui Nian, Fang Liu, Bo He. 2013. An Early Underwater Artificial Vision Model in Ocean Investigations via Independent Component Analysis. *Sensors* **13**:7, 9104-9131. [Crossref]
266. Luigi Malago, Matteo Matteucci, Giovanni Pistone. Natural gradient, fitness modelling and model selection: A unifying perspective 486-493. [Crossref]
267. Kazuyuki Masutomi, Yuichi Nagata, Isao Ono. Extending distance-weighted exponential natural evolution strategy for function optimization in uncertain environments 2122-2129. [Crossref]
268. Xiaomin Duan, Huafei Sun, Linyu Peng, Xinyu Zhao. 2013. A natural gradient descent algorithm for the solution of discrete algebraic Lyapunov equations based on the geodesic distance. *Applied Mathematics and Computation* **219**:19, 9899-9905. [Crossref]
269. Takumi Kobayashi. BFO Meets HOG: Feature Extraction Based on Histograms of Oriented p.d.f. Gradients for Image Classification 747-754. [Crossref]
270. Scott R Kuindersma, Roderic A Grupen, Andrew G Barto. 2013. Variable risk control via stochastic optimization. *The International Journal of Robotics Research* **32**:7, 806-825. [Crossref]
271. Steve Young, Milica Gasic, Blaise Thomson, Jason D. Williams. 2013. POMDP-Based Statistical Spoken Dialog Systems: A Review. *Proceedings of the IEEE* **101**:5, 1160-1179. [Crossref]
272. Kenichi Kumatani, Rita Singh, Friedrich Faubel, John McDonough, Youssef Oualil. Joint constrained maximum likelihood regression for overlapping speech recognition 121-125. [Crossref]
273. Olivier Buffet. Policy-Gradient Algorithms 127-152. [Crossref]
274. Taiji Suzuki, Masashi Sugiyama. 2013. Sufficient Dimension Reduction via Squared-Loss Mutual Information Estimation. *Neural Computation* **25**:3, 725-758. [Abstract] [Full Text] [PDF] [PDF Plus]
275. HANG TAN, XIANHE HUANG, YING TANG, HUACHUN TAN. 2013. NEW ICA ALGORITHMS BASED ON SPECIAL LINEAR GROUP. *International Journal of Modeling, Simulation, and Scientific Computing* **04**:01, 1250028. [Crossref]
276. Andres El-Fakdi, Marc Carreras. 2013. Two-step gradient-based reinforcement learning for underwater robotics behavior learning. *Robotics and Autonomous Systems* **61**:3, 271-282. [Crossref]
277. Jimin Ye, Haihong Jin, Qingrui Zhang. 2013. Adaptive weighted orthogonal constrained algorithm for blind source separation. *Digital Signal Processing* **23**:2, 514-521. [Crossref]
278. ShiCheng Zhang, HuaFei Sun, ChunHui Li. 2013. Principal whitened gradient-projection algorithm for distribution control. *Science China Information Sciences* **56**:3, 1-8. [Crossref]
279. Ran Tao, Bing-zhao Li, Hua-fei Sun. 2013. Research Progress of The Algebraic and Geometric Signal Processing. *Defence Technology* **9**:1, 40-47. [Crossref]
280. . Topic-based Generative Models for Text Information Access 129-177. [Crossref]
281. K. Dvijotham, E. Todorov. Linearly Solvable Optimal Control 119-141. [Crossref]
282. Ani Eloyan, Sujit K. Ghosh. 2013. A semiparametric approach to source separation using independent component analysis. *Computational Statistics & Data Analysis* **58**, 383-396. [Crossref]
283. Wentao Fan, N. Bouguila. An online approach for learning non-Gaussian mixture models with localized feature selection 1-6. [Crossref]
284. Benedikt Loesch, Bin Yang. 2013. Cramér-Rao Bound for Circular and Noncircular Complex Independent Component Analysis. *IEEE Transactions on Signal Processing* **61**:2, 365-379. [Crossref]
285. Shun-ichi Amari. 2013. Dreaming of mathematical neuroscience for half a century. *Neural Networks* **37**, 48-51. [Crossref]
286. Rostom Mabrouk, François Dubeau, Layachi Bentabet. 2013. Dynamic Cardiac PET Imaging: Extraction of Time-Activity Curves Using ICA and a Generalized Gaussian Distribution Model. *IEEE Transactions on Biomedical Engineering* **60**:1, 63-71. [Crossref]
287. Ulrich Paquet, Noam Koenigstein. One-class collaborative filtering with random graphs 999-1008. [Crossref]
288. Mounira Maazaoui, Karim Abed-Meraim, Yves Grenier. 2012. Blind source separation for robot audition using fixed HRTF beamforming. *EURASIP Journal on Advances in Signal Processing* **2012**:1. . [Crossref]

289. Wentao Fan, Nizar Bouguila. Online Variational Learning for a Dirichlet Process Mixture of Dirichlet Distributions and its Application 362-367. [Crossref]

290. Youhei Akimoto, Yuichi Nagata, Isao Ono, Shigenobu Kobayashi. 2012. Theoretical Foundation for CMA-ES from Information Geometry Perspective. *Algorithmica* **64**:4, 698-716. [Crossref]

291. Filip Jurcicek. Reinforcement learning for spoken dialogue systems using off-policy natural gradient method 7-12. [Crossref]

292. Christopher M. Sandiko, Elmer R. Magsino. A Blind Source Separation of instantaneous acoustic mixtures using Natural Gradient Method 124-129. [Crossref]

293. A. Abdul Hadi, S. Sanei. A New Frequency Domain Blind Separation and Localization of Anechoic Mixtures Incorporating Spatial Notch Filtering 232-237. [Crossref]

294. Philip S. Thomas, Andrew G. Barto. Motor primitive discovery 1-8. [Crossref]

295. Ivo Grondman, Lucian Busoniu, Gabriel A. D. Lopes, Robert Babuska. 2012. A Survey of Actor-Critic Reinforcement Learning: Standard and Natural Policy Gradients. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **42**:6, 1291-1307. [Crossref]

296. He Xin, Shi Yingchun. An Improved fastICA Based on the Negative Entropy for Voiceprint Identification 677-680. [Crossref]

297. Xu Huang. An improved FastICA algorithm for blind signal separation and its application 1-4. [Crossref]

298. Masayuki Karasuyama, Masashi Sugiyama. 2012. Canonical dependency analysis based on squared-loss mutual information. *Neural Networks* **34**, 46-55. [Crossref]

299. Seyedmahdad Mirsamadi, Shabnam Ghaffarzadegan, Hamid Sheikhzadeh, Seyed Mohammad Ahadi, Amir Hossein Rezaie. 2012. Efficient Frequency Domain Implementation of Noncausal Multichannel Blind Deconvolution for Convolutive Mixtures of Speech. *IEEE Transactions on Audio, Speech, and Language Processing* **20**:8, 2365-2377. [Crossref]

300. Johannes Fürnkranz, Eyke Hüllermeier, Weiwei Cheng, Sang-Hyeun Park. 2012. Preference-based reinforcement learning: a formal framework and a policy iteration algorithm. *Machine Learning* **89**:1-2, 123-156. [Crossref]

301. Mathieu N. Galtier, Olivier D. Faugeras, Paul C. Bressloff. 2012. Hebbian Learning of Recurrent Connections: A Geometrical Perspective. *Neural Computation* **24**:9, 2346-2383. [Abstract] [Full Text] [PDF] [PDF Plus]

302. Satoshi Hara, Yoshinobu Kawahara, Takashi Washio, Paul von Bünau, Terumasa Tokunaga, Kiyohumi Yumoto. 2012. Separation of stationary and non-stationary sources with a generalized eigenvalue problem. *Neural Networks* **33**, 7-20. [Crossref]

303. Wentao Fan, Nizar Bouguila. 2012. Online variational learning of finite Dirichlet mixture models. *Evolving Systems* **3**:3, 153-165. [Crossref]

304. Nihat Ay, Holger Bernigau, Ralf Der, Mikhail Prokopenko. 2012. Information-driven self-organization: the dynamical system approach to autonomous robot behavior. *Theory in Biosciences* **131**:3, 161-179. [Crossref]

305. Anders Johansson, Kai Ramsch, Martin Middendorf, David J.T. Sumpter. 2012. Tuning positive feedback for signal detection in noisy dynamic environments. *Journal of Theoretical Biology* **309**, 88-95. [Crossref]

306. Georg Martius, J. Michael Herrmann. 2012. Variants of guided self-organization for robot control. *Theory in Biosciences* **131**:3, 129-137. [Crossref]

307. Shao Zhifei, Er Meng Joo. 2012. A survey of inverse reinforcement learning techniques. *International Journal of Intelligent Computing and Cybernetics* **5**:3, 293-311. [Crossref]

308. R. Vullings. Probabilistic source separation for robust electrocardiography 6492-6495. [Crossref]

309. Wentao Fan, Nizar Bouguila. Online variational finite Dirichlet mixture model and its applications 448-453. [Crossref]

310. Shao Zhifei, Er Meng Joo. A review of inverse reinforcement learning theory and recent advances 1-8. [Crossref]

311. Jun-Wei Qiu, John K. Zao, Yu-Hsiang Chou. A geometrically faithful memetic algorithm for searching sparse representations of EEG signals 1-7. [Crossref]

312. Yuhua Dong, Junxing Zhang. Variable step-size blind source separation algorithm based on orthogonal gradient 304-307. [Crossref]

313. Filip Jurčíček, Blaise Thomson, Steve Young. 2012. Reinforcement learning for parameter estimation in statistical spoken dialogue systems. *Computer Speech & Language* **26**:3, 168-192. [Crossref]

314. MUhammad Tahir Akhtar, Tzyy-Ping Jung, Scott Makeig, Gert Cauwenberghs. Recursive independent component analysis for online blind source separation 2813-2816. [Crossref]

315. Marco Signoretto, Emanuele Olivetti, Lieven De Lathauwer, Johan A. K. Suykens. 2012. Classification of Multichannel Signals With Cumulant-Based Kernels. *IEEE Transactions on Signal Processing* **60**:5, 2304-2314. [Crossref]

316. Hideaki Shimazaki, Shun-ichi Amari, Emery N. Brown, Sonja Grün. 2012. State-Space Analysis of Time-Varying Higher-Order Spike Correlation for Multiple Neural Spike Train Data. *PLoS Computational Biology* **8**:3, e1002385. [Crossref]

317. Takuma Ono, Nobutaka Ono, Shigeki Sagayama. User-guided independent vector analysis with source activity tuning 2417-2420. [Crossref]

318. Dimitrije Marković, Claudius Gros. 2012. Intrinsic Adaptation in Autonomous Recurrent Neural Networks. *Neural Computation* **24**:2, 523-540. [Abstract] [Full Text] [PDF] [PDF Plus]

319. Jake Gunther. 2012. Learning Echo Paths During Continuous Double-Talk Using Semi-Blind Source Separation. *IEEE Transactions on Audio, Speech, and Language Processing* **20**:2, 646-660. [Crossref]

320. Dezhong Peng, Zhang Yi, Yong Xiang, Haixian Zhang. 2012. A Globally Convergent MC Algorithm With an Adaptive Learning Rate. *IEEE Transactions on Neural Networks and Learning Systems* **23**:2, 359-365. [Crossref]

321. Ryu Takeda, Kazuhiro Nakadai, Toru Takahashi, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno. 2012. Efficient Blind Dereverberation and Echo Cancellation Based on Independent Component Analysis for Actual Acoustic Signals. *Neural Computation* **24**:1, 234-272. [Abstract] [Full Text] [PDF] [PDF Plus]

322. James Martens, Ilya Sutskever. Training Deep and Recurrent Networks with Hessian-Free Optimization 479-535. [Crossref]

323. Federico Raimondo, Juan E. Kamienkowski, Mariano Sigman, Diego Fernandez Slezak. 2012. CUDAICA: GPU Optimization of Infomax-ICA EEG Analysis. *Computational Intelligence and Neuroscience* **2012**, 1-8. [Crossref]

324. Jen-Tzung Chien, Hsin-Lung Hsieh. 2012. Convex Divergence ICA for Blind Source Separation. *IEEE Transactions on Audio, Speech, and Language Processing* **20**:1, 302-313. [Crossref]

325. Yoshua Bengio. Practical Recommendations for Gradient-Based Training of Deep Architectures 437-478. [Crossref]

326. Zhu Shou Zhong, Liu Zheng, Jiang Wen Li, Guo Kun. 2012. The Key technology of Blind Source Separation of Satellite-Based AIS. *Procedia Engineering* **29**, 3737-3741. [Crossref]

327. Kazuyuki Hara, Kentaro Katahira, Kazuo Okanoya, Masato Okada. Theoretical Analysis of Function of Derivative Term in On-Line Gradient Descent Learning 9-16. [Crossref]

328. Yann A. LeCun, Léon Bottou, Genevieve B. Orr, Klaus-Robert Müller. Efficient BackProp 9-48. [Crossref]

329. Yongqiang Cheng, Xuezhi Wang, Terry Caelli, Xiang Li, Bill Moran. 2011. Optimal Nonlinear Estimation for Localization of Wireless Sensor Networks. *IEEE Transactions on Signal Processing* **59**:12, 5674-5685. [Crossref]

330. Inkyung Ahn, Jooyoung Park. 2011. Drug scheduling of cancer chemotherapy based on natural actor-critic approach. *Biosystems* **106**:2-3, 121-129. [Crossref]

331. Tianqi Zhang, Lizhong Li, Gang Zhang, Chenqiang Gao, Ruiling Hou. Use estimation of Performance Index to realize adaptive blind source separation 2322-2326. [Crossref]

332. Lianfang Cai, Xuemin Tian. Improved post-nonlinear independent component analysis method based on Gaussian Mixture Model 274-279. [Crossref]

333. Guoxu Zhou, Shengli Xie, Zuyuan Yang, Jun-Mei Yang, Zhaoshui He. 2011. Minimum-Volume-Constrained Nonnegative Matrix Factorization: Enhanced Ability of Learning Parts. *IEEE Transactions on Neural Networks* **22**:10, 1626-1637. [Crossref]

334. Kefeng Wang, Xu Xu, Chonghui Guo. A stable ICA algorithm based on exponent density and Gaussian parametric density mixture models 291-296. [Crossref]

335. Ji Ce, Tang Baocheng, Wang Jinyu. Blind source separation based on variable step length natural gradient algorithm 261-264. [Crossref]

336. Anthony Lombard, Yuanhang Zheng, Herbert Buchner, Walter Kellermann. 2011. TDOA Estimation for Multiple Sound Sources in Noisy and Reverberant Environments Using Broadband Independent Component Analysis. *IEEE Transactions on Audio, Speech, and Language Processing* **19**:6, 1490-1503. [Crossref]

337. Atsushi Miyamoto, Kazuho Watanabe, Kazushi Ikeda, Masa-aki Sato. Phase diagrams of a variational Bayesian approach with ARD prior in NIRS-DOT 1230-1236. [Crossref]

338. Simone Fiori. 2011. Solving Minimal-Distance Problems over the Manifold of Real-Symplectic Matrices. *SIAM Journal on Matrix Analysis and Applications* **32**:3, 938-968. [Crossref]

339. . Bibliography . [Crossref]

340. Wei Xia, Xuesong Liu, Bin Wang, Liming Zhang. 2011. Independent Component Analysis for Blind Unmixing of Hyperspectral Imagery With Additional Constraints. *IEEE Transactions on Geoscience and Remote Sensing* **49**:6, 2165-2179. [Crossref]

341. Nobusumi Fukushima, Yuichi Nagata, Sigenobu Kobayashi, Isao Ono. Proposal of distance-weighted exponential natural evolution strategies 164-171. [Crossref]

342. Pontus Johannisson, Henk Wymeersch, Martin Sjödin, A. Serdar Tan, Erik Agrell, Peter A. Andrekson, Magnus Karlsson. 2011. Convergence Comparison of the CMA and ICA for Blind Polarization Demultiplexing. *Journal of Optical Communications and Networking* **3**:6, 493. [Crossref]

343. Rupert C.J. Minnett, Andrew T. Smith, William C. Lennon, Robert Hecht-Nielsen. 2011. Neural network tomography: Network replication from output surface geometry. *Neural Networks* **24**:5, 484-492. [Crossref]

344. Luigi Malago, Matteo Matteucci, Giovanni Pistone. Stochastic Natural Gradient Descent by estimation of empirical covariances 949-956. [Crossref]

345. Gen Hori. Natural gradient approach in orthogonal matrix optimization using cayley transform 2116-2119. [Crossref]

346. S Fiori. 2011. Extended Hamiltonian Learning on Riemannian Manifolds: Theoretical Aspects. *IEEE Transactions on Neural Networks* **22**:5, 687-700. [Crossref]

347. Jianwei Wu. 2011. Estimating source kurtosis directly from observation data for ICA. *Signal Processing* **91**:5, 1150-1156. [Crossref]

348. Haihong Zhang, Cuntai Guan, Yuanqing Li. 2011. A linear discriminant analysis method based on mutual information maximization. *Pattern Recognition* **44**:4, 877-885. [Crossref]

349. Guoxu Zhou, Zuyuan Yang, Shengli Xie, Jun-Mei Yang. 2011. Online Blind Source Separation Using Incremental Nonnegative Matrix Factorization With Volume Constraint. *IEEE Transactions on Neural Networks* **22**:4, 550-560. [Crossref]

350. Daniel Povey, Lukáš Burget, Mohit Agarwal, Pinar Akyazi, Feng Kai, Arnab Ghoshal, Ondřej Glembek, Nagendra Goel, Martin Karafiát, Ariya Rastrow, Richard C. Rose, Petr Schwarz, Samuel Thomas. 2011. The subspace Gaussian mixture model—A structured model for speech recognition. *Computer Speech & Language* **25**:2, 404-439. [Crossref]

351. Zhucheng Li, Rong Liu. A novel adaptive step-size algorithm of blind source separation based on non-linear function 4675-4678. [Crossref]

352. Francesco Nesta, Ted S Wada, Biing-Hwang Juang. 2011. Batch-Online Semi-Blind Source Separation Applied to Multi-Channel Acoustic Echo Cancellation. *IEEE Transactions on Audio, Speech, and Language Processing* **19**:3, 583-599. [Crossref]

353. Francesco Nesta, Piergiorgio Svaizer, Maurizio Omologo. 2011. Convolutive BSS of Short Mixtures by ICA Recursively Regularized Across Frequencies. *IEEE Transactions on Audio, Speech, and Language Processing* **19**:3, 624-639. [Crossref]

354. Masashi Sugiyama, Makoto Yamada, Paul von Bünau, Taiji Suzuki, Takafumi Kanamori, Motoaki Kawanabe. 2011. Direct density-ratio estimation with dimensionality reduction via least-squares hetero-distributional subspace search. *Neural Networks* **24**:2, 183-198. [Crossref]

355. Shai Shalev-Shwartz, Yoram Singer, Nathan Srebro, Andrew Cotter. 2011. Pegasos: primal estimated sub-gradient solver for SVM. *Mathematical Programming* **127**:1, 3-30. [Crossref]

356. Csaba Szepesvári. Reinforcement Learning Algorithms for MDPs . [Crossref]

357. . Model Design and Selection Considerations 37-63. [Crossref]

358. Taiji Suzuki, Masashi Sugiyama. 2011. Least-Squares Independent Component Analysis. *Neural Computation* **23**:1, 284-301. [Abstract] [Full Text] [PDF] [PDF Plus] [Supplemental Material]

359. Takuya Yoshioka, Tomohiro Nakatani, Masato Miyoshi, Hiroshi G. Okuno. 2011. Blind Separation and Dereverberation of Speech Mixtures by Joint Optimization. *IEEE Transactions on Audio, Speech, and Language Processing* **19**:1, 69-84. [Crossref]

360. Michael R. Bastian, Jacob H. Gunther, Todd K. Moon. 2011. A Simplified Natural Gradient Learning Algorithm. *Advances in Artificial Neural Systems* **2011**, 1-9. [Crossref]

361. Jorge Igual, Jehad Ababneh, Raul Llinares, Carmen Igual. Using Particle Swarm Optimization for Minimizing Mutual Information in Independent Component Analysis 484-491. [Crossref]

362. Amin Kheradmand, Hamid Sheikhzadeh, Kamraan Raahemifar, Ebrahim Ghanavati. Blind Source Separation in nonminimum-phase systems based on filter decomposition 332-337. [Crossref]

363. Baeksuk Chu, Jooyoung Park, Daehie Hong. 2010. Tunnel ventilation controller design using an RLS-based natural actor-critic algorithm. *International Journal of Precision Engineering and Manufacturing* **11**:6, 829-838. [Crossref]

364. Tao Yu, John H L Hansen. 2010. Discriminative Training for Multiple Observation Likelihood Ratio Based Voice Activity Detection. *IEEE Signal Processing Letters* **17**:11, 897-900. [Crossref]

365. Zhao Zhang, Man Jiang, Ning Ye. 2010. Effective multiplicative updates for non-negative discriminative learning in multimodal dimensionality reduction. *Artificial Intelligence Review* **34**:3, 235-260. [Crossref]

366. Blaise Thomson, Steve Young. 2010. Bayesian update of dialogue state: A POMDP framework for spoken dialogue systems. *Computer Speech & Language* **24**:4, 562-588. [Crossref]

367. Dijun Luo, Heng Huang, Chris Ding, Feiping Nie. 2010. On the eigenvectors of p-Laplacian. *Machine Learning* **81**:1, 37-51. [Crossref]
368. . References 285-302. [Crossref]
369. Sven Hoffmann, Michael Falkenstein. 2010. Independent component analysis of erroneous and correct responses suggests online response control. *Human Brain Mapping* **31**:9, 1305-1315. [Crossref]
370. JiangNan Yuan, JiangHong Shi, BiYu Tang, HuiHuang Chen. An Adaptive Feedback Interference Cancelling Algorithm Based on Independent Component Analysis for Wireless Repeaters 1-4. [Crossref]
371. Ying Tang, Jianping Li. 2010. Normalized natural gradient in independent component analysis. *Signal Processing* **90**:9, 2773-2777. [Crossref]
372. Shun-ichi Amari. 2010. Information geometry in optimization, machine learning and statistical inference. *Frontiers of Electrical and Electronic Engineering in China* **5**:3, 241-260. [Crossref]
373. Hirofumi Nakajima, Kazuhiro Nakadai, Yuji Hasegawa, Hiroshi Tsujino. 2010. Blind Source Separation With Parameter-Free Adaptive Step-Size Method for Robot Audition. *IEEE Transactions on Audio, Speech, and Language Processing* **18**:6, 1476-1485. [Crossref]
374. Zuyuan Yang, Guoxu Zhou, Shuxue Ding, Shengli Xie. Blind source separation by nonnegative matrix factorization with minimum-volume constraint 117-119. [Crossref]
375. Peng Li, Rui Li. Using higher order statistics and time structure to separate source signals 629-632. [Crossref]
376. Yue-Der Lin, Chih-Yu Hsu, Hung-Yun Chen, Kuo-Kun Tseng. An efficient ICA approach based on neural network framework for biomedical applications 1-8. [Crossref]
377. Peng Li, Rui Li. Using higher order statistics and time structure to separate source signals 292-295. [Crossref]
378. Wei Xia, Bin Wang, Liming Zhang. Constrained independent component analysis for hyperspectral unmixing 1293-1296. [Crossref]
379. Darko Zikic, Ali Kamen, Nassir Navab. Natural gradients for deformable registration 2847-2854. [Crossref]
380. Yunfeng Xue, Yujie Han, Yuja Wang. A Concrete Independent Component Analysis Algorithm Based on the Estimating Function 4590-4593. [Crossref]
381. Keyan Zahedi, Nihat Ay, Ralf Der. 2010. Higher Coordination With Less Control—A Result of Information Maximization in the Sensorimotor Loop. *Adaptive Behavior* **18**:3-4, 338-355. [Crossref]
382. Andres El-Fakdi, Marc Carreras, Enric Galceran. Two steps natural actor critic learning for underwater cable tracking 2267-2272. [Crossref]
383. M. Lankarany, M. H. Savoji. Knowledge based blind deconvolution of non-minimum phase FIR systems 1-5. [Crossref]
384. Simone Fiori. 2010. Learning by Natural Gradient on Noncompact Matrix-Type Pseudo-Riemannian Manifolds. *IEEE Transactions on Neural Networks* **21**:5, 841-852. [Crossref]
385. Ryu Takeda, Kazuhiro Nakadai, Toru Takahashi, Kazunori Komatani, Tetsuya Ogata, Hiroshi G Okuno. Upper-limit evaluation of robot audition based on ICA-BSS in multi-source, barge-in and highly reverberant conditions 4366-4371. [Crossref]
386. Francisco das Chagas de Souza, Orlando José Tobias, Rui Seara, Dennis R. Morgan. 2010. A PNLMS Algorithm With Individual Activation Factors. *IEEE Transactions on Signal Processing* **58**:4, 2036-2047. [Crossref]
387. Byungchan Kim, Jooyoung Park, Shinsuk Park, Sungchul Kang. 2010. Impedance Learning for Robotic Contact Tasks Using Natural Actor-Critic Algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **40**:2, 433-443. [Crossref]
388. . Bibliography 267-281. [Crossref]
389. Yu Zhu, Ruina Wu, Aiguo Li, Shengping He. The Recognition of Vibration Sensor's Signals Based on ICA 957-961. [Crossref]
390. Wei-Tao Zhang, Shun-Tian Lou, Yan-Liang Zhang. 2010. Robust nonlinear power iteration algorithm for adaptive blind separation of independent signals. *Digital Signal Processing* **20**:2, 541-551. [Crossref]
391. Y. Panagakis, C. Kotropoulos, G.R. Arce. 2010. Non-Negative Multilinear Principal Component Analysis of Auditory Temporal Modulations for Music Genre Classification. *IEEE Transactions on Audio, Speech, and Language Processing* **18**:3, 576-588. [Crossref]
392. Klaus Reindl, Yuanhang Zheng, Walter Kellermann. Speech enhancement for binaural hearing aids based on blind source separation 1-6. [Crossref]
393. V. Zarzoso, P. Comon. 2010. Robust Independent Component Analysis by Iterative Maximization of the Kurtosis Contrast With Algebraic Optimal Step Size. *IEEE Transactions on Neural Networks* **21**:2, 248-261. [Crossref]

394. V. Zarzoso, A. Hyvärinen. Iterative algorithms 179-225. [Crossref]
395. A Hamza, E. H. Meftah, S. Chitroub, R. Touhami. ICA-based blind symbol detection for compound system MIMO-OFDM in CDMA context 345-349. [Crossref]
396. Francesco Nesta, Maurizio Omologo. Cooperative Wiener-ICA for source localization and Separation by distributed microphone arrays 1-4. [Crossref]
397. Zhang Xuxiu, Qiu Tianshuang. Blind multiuser detection based on Improved Infomax and FastICA 476-479. [Crossref]
398. Csaba Szepesvári. 2010. Algorithms for Reinforcement Learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* **4**:1, 1-103. [Crossref]
399. Addisson Salazar, Luis Vergara, Arturo Serrano, Jorge Igual. 2010. A general procedure for learning mixtures of independent component analyzers. *Pattern Recognition* **43**:1, 69-85. [Crossref]
400. J.-F. Cardoso. Likelihood 107-154. [Crossref]
401. HyukJoon Jang, Hong Jeong. A Study on Source Separation Using Orthogonality between Independent Speeches in Spectrogram 338-341. [Crossref]
402. Matthieu Geist, Olivier Pietquin. Revisiting Natural Actor-Critics with Value Function Approximation 207-218. [Crossref]
403. Xueshun Wang, Shao-hua Yu, JinYou Dai, Ting Luo, Xueshun Wang, JinYou Dai, Ting Luo. A Multiple Constraint Quality of Service Routing Algorithm Base on Dominating Tree 1-4. [Crossref]
404. Yunfeng Xue, Feng Ju, Yujia Wang, Jie Yang. 2009. A source adaptive independent component analysis algorithm through solving the estimating equation. *Expert Systems with Applications* **36**:10, 12306-12313. [Crossref]
405. Chun-Nan Hsu, Han-Shen Huang, Yu-Ming Chang, Yuh-Jye Lee. 2009. Periodic step-size adaptation in second-order gradient descent for single-pass on-line structured learning. *Machine Learning* **77**:2-3, 195-224. [Crossref]
406. Ryu Takeda, Kazuhiro Nakadai, Torn Takahashi, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno. Automatic estimation of reverberation time with robot speech to improve ICA-based robot audition 250-255. [Crossref]
407. Manuele Bicego, Elżbieta Pe˛kalska, David M.J. Tax, Robert P.W. Duin. 2009. Component-based discriminative classification for hidden Markov models. *Pattern Recognition* **42**:11, 2637-2648. [Crossref]
408. Shalabh Bhatnagar, Richard S. Sutton, Mohammad Ghavamzadeh, Mark Lee. 2009. Natural actor–critic algorithms. *Automatica* **45**:11, 2471-2482. [Crossref]
409. Nizar Bouguila, Ola Amayri. 2009. A discrete mixture-based kernel for SVMs: Application to spam and image categorization. *Information Processing & Management* **45**:6, 631-642. [Crossref]
410. Yoshitatsu Matsuda, Kazunori Yamaguchi. 2009. Linear Multilayer ICA Using Adaptive PCA. *Neural Processing Letters* **30**:2, 133-144. [Crossref]
411. Hao Cheng, Bin Tang, Jingjing Du, Xiaojun Chen, Gaoyi Zhang. 2009. Mixing vector estimation in single channel blind source separation of angle modulated signal sources based on cumulant system of equations. *Signal Processing* **89**:10, 1932-1940. [Crossref]
412. Ryu Takeda, Kazuhiro Nakadai, Toru Takahashi, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno. Step-size parameter adaptation of multi-channel semi-blind ICA with piecewise linear model for barge-in-able robot audition 2277-2282. [Crossref]
413. Jing Sui, Tülay Adali, Godfrey D. Pearlson, Vincent P. Clark, Vince D. Calhoun. 2009. A method for accurate group difference detection by constraining the mixing coefficients in an ICA framework. *Human Brain Mapping* **30**:9, 2953-2970. [Crossref]
414. Michele Scarpiniti, Andrea Picaro, Raffaele Parisi, Aurelio Uncini. A partitioned frequency domain algorithm for convolutive blind source separation 1-6. [Crossref]
415. Zhenning Zhang, Huafei Sun, Fengwei Zhong. 2009. Natural gradient-projection algorithm for distribution control. *Optimal Control Applications and Methods* **30**:5, 495-504. [Crossref]
416. Tiemin Mei, Fuliang Yin, Jun Wang. 2009. Blind Source Separation Based on Cumulants With Time and Frequency Non-Properties. *IEEE Transactions on Audio, Speech, and Language Processing* **17**:6, 1099-1108. [Crossref]
417. Ken-ichi Tamura, Miho Komiya, Masato Inoue, Yoshiyuki Kabashima. 2009. Decoding Algorithm of Low-density Parity-check Codes based on Bowman-Levin Approximation. *New Generation Computing* **27**:4, 347-363. [Crossref]
418. Makoto Miyakoshi, Moyoko Tomiyasu, Epifanio Bagarinao, Shumei Murakami, Toshiharu Nakai. 2009. A Phantom Study On Component Segregation for MR Images Using ICA. *Academic Radiology* **16**:8, 1025-1028. [Crossref]
419. I U Shah, T S Durrani. Broadband Constraint Natural Gradient based Adaptive processing 193-196. [Crossref]
420. Haiting Tian, Jing Jin, Chunxi Zhang, Ningfang Song. 2009. Informax-Based Data Fusion for Sensor Network. *IEEE Sensors Journal* **9**:7, 820-827. [Crossref]

421. Mikael Kuusela, Tapani Raiko, Antti Honkela, Juha Karhunen. A gradient-based algorithm competitive with variational Bayesian EM for mixture of Gaussians 1688-1695. [Crossref]

422. Yin Huan-ping, Sun Zong-hai. An online learning algorithm of support vector regression based on natural gradient 5615-5618. [Crossref]

423. Frank Nielsen, Richard Nock. 2009. Sided and Symmetrized Bregman Centroids. *IEEE Transactions on Information Theory* **55**:6, 2882-2904. [Crossref]

424. Rafael Boloix-Tortosa, Franciso J. Simois-Tirado, Juan Jose Murillo-Fuentes. Blind adaptive channel estimation for OFDM systems 191-195. [Crossref]

425. A. Oka, L. Lampe. 2009. Incremental Distributed Identification of Markov Random Field Models in Wireless Sensor Networks. *IEEE Transactions on Signal Processing* **57**:6, 2396-2405. [Crossref]

426. Elizabeth Hoppe, Michael Roan. 2009. Non-linear, adaptive array processing for acoustic interference suppression. *The Journal of the Acoustical Society of America* **125**:6, 3835-3843. [Crossref]

427. Davide Bacciu, Ian H. Jarman, Terence A. Etchells, Paulo J.G. Lisboa. Patient stratification with competing risks by multivariate Fisher distance 213-220. [Crossref]

428. T. Ishibashi, H. Nakashima, K. Inoue, H. Gotanda. 2009. Blind Source Separation Under a Dynamic Acoustic Environment. *Proceedings of the ISCIE International Symposium on Stochastic Systems Theory and its Applications* **2009**:0, 318-323. [Crossref]

429. L.R. Vega, H. Rey, J. Benesty, S. Tressens. 2009. A Family of Robust Algorithms Exploiting Sparsity in Adaptive Filters. *IEEE Transactions on Audio, Speech, and Language Processing* **17**:4, 572-581. [Crossref]

430. Ercan E. Kuruoğlu, Ayşın Ertüzün, Deniz Gençağa. 2009. WITHDRAWN: Separation of non-stationary mixtures of cross-correlated processes using sequential Monte Carlo. *Digital Signal Processing* . [Crossref]

431. Jae-Min Kim, Hyun-Soo Kang. 2009. Efficient Coding Technique for 4X4 Intra Prediction Modes using the Statistical Distribution of Intra Modes of Adjacent Intra Blocks. *The Journal of the Korea Contents Association* **9**:4, 12-18. [Crossref]

432. Jin-Sung Yoon, Gye-Young Kim, Hyung-Il Choi. 2009. Development of an Adult Image Classifier using Skin Color. *The Journal of the Korea Contents Association* **9**:4, 1-11. [Crossref]

433. Sang-Hoon Oh. 2009. Comparisons of Linear Feature Extraction Methods. *The Journal of the Korea Contents Association* **9**:4, 121-130. [Crossref]

434. Ryu Takeda, Kazuhiro Nakadai, Toru Takahashi, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno. ICA-based efficient blind dereverberation and echo cancellation method for barge-in-able robot audition 3677-3680. [Crossref]

435. Jian-Qiang Liu, Da-Zheng Feng, Wei-Wei Zhang. 2009. Adaptive Improved Natural Gradient Algorithm for Blind Source Separation. *Neural Computation* **21**:3, 872-889. [Abstract] [Full Text] [PDF] [PDF Plus]

436. M. Yukawa. 2009. Krylov-Proportionate Adaptive Filtering Techniques Not Limited to Sparse Systems. *IEEE Transactions on Signal Processing* **57**:3, 927-943. [Crossref]

437. Yogesh P. Awate. Algorithms for variance reduction in a policy-gradient based actor-critic framework 130-136. [Crossref]

438. Sam McKennoch, Thomas Voegtlin, Linda Bushnell. 2009. Spike-Timing Error Backpropagation in Theta Neuron Networks. *Neural Computation* **21**:1, 9-45. [Abstract] [Full Text] [PDF] [PDF Plus]

439. Ling-Zhi Liao. Sparse Coding for Natural Images Based on Pearson-Type VII Density Functions 1429-1432. [Crossref]

440. Zhao Min, Li Weijun, Zhou Guoxu, Zhou Zhiheng. Blind Sources Separation Algorithm Based on Adaptive Givens Rotations 95-99. [Crossref]

441. Ling-Zhi Liao. Learning Sparse Representations Using a Parametric Cauchy Density 994-1002. [Crossref]

442. Felipe P. do Carmo, Joaquim T. de Assis, Vania Vieira Estrela, Alessandra M. Coelho. Blind signal separation and identification of mixtures of images 337-342. [Crossref]

443. Fadi Abu-Amara, Ikhlas Abdel-Qader. 2009. Hybrid Mammogram Classification Using Rough Set and Fuzzy Classifier. *International Journal of Biomedical Imaging* **2009**, 1-12. [Crossref]

444. Suxian Zhang, Hailin Liu, Jiechang Wen, Weili Chen. A New Algorithm Estimating the Mixing Matrix for the Sparse Component Analysis 25-29. [Crossref]

445. Manuele Bicego, Marco Cristani, Vittorio Murino, Elżbieta Pękalska, Robert P. W. Duin. Clustering-Based Construction of Hidden Markov Models for Generative Kernels 466-479. [Crossref]

446. Bin Xia. Multichannel Blind Deconvolution Using the Conjugate Gradient 612-620. [Crossref]

447. Jason A. Palmer, Ken Kreutz-Delgado, Scott Makeig. Probabilistic Formulation of Independent Vector Analysis Using Complex Gaussian Scale Mixtures 90-97. [Crossref]

448. Richard Nock, Frank Nielsen. Intrinsic Geometries in Learning 175-215. [Crossref]
449. Shun-ichi Amari. Information Geometry and Its Applications: Convex Function and Dually Flat Manifold 75-102. [Crossref]
450. Ou Shifeng, Gao Ying, Jin Gang, Zhang Xuehui. Variable Step Size Algorithm for Blind Source Separation Using a Combination of Two Adaptive Separation Systems 649-652. [Crossref]
451. Jingyu Liu, Godfrey Pearlson, Andreas Windemuth, Gualberto Ruano, Nora I. Perrone-Bizzozero, Vince Calhoun. 2009. Combining fMRI and SNP data to investigate connections between brain function and genetics using parallel ICA. *Human Brain Mapping* **30**:1, 241-255. [Crossref]
452. Yogesh P. Awate. Policy-Gradient Based Actor-Critic Algorithms 505-509. [Crossref]
453. Dahui Li, Ying Zheng, Jinku Han. Blind Separation Algorithm with Improved Learning Rule for Audio Signal 166-169. [Crossref]
454. Tsuyoshi Ueno, Shin-ichi Maeda, Motoaki Kawanabe, Shin Ishii. Optimal Online Learning Procedures for Model-Free Policy Evaluation 473-488. [Crossref]
455. Wai Yie Leong. Blind Source Extraction 1672-1677. [Crossref]
456. Tiemin Mei, Alfred Mertins. 2008. Convolutive Blind Source Separation Based on Disjointness Maximization of Subband Signals. *IEEE Signal Processing Letters* **15**, 725-728. [Crossref]
457. Dahui Li, Ming Diao, Xuefeng Dai. Blind Separation Algorithm for Audio Signal Based on Genetic Algorithm and Neural Network 436-440. [Crossref]
458. Tetsuro Morimura, Eiji Uchibe, Kenji Doya. 2008. Natural actor-critic with baseline adjustment for variance reduction. *Artificial Life and Robotics* **13**:1, 275-279. [Crossref]
459. Xueying Zhang, Zhenhua Zhi, Xiaomei Zhang. Variable step-size speech blind separation employing Laplacian normal mixture distribution model 785-788. [Crossref]
460. Wei Kong, Charles R. Vanderburg, Hiromi Gunshin, Jack T. Rogers, Xudong Huang. 2008. A review of independent component analysis application to microarray gene expression data. *BioTechniques* **45**:5, 501-520. [Crossref]
461. Wai Yie Leong, D.P. Mandic. 2008. Post-Nonlinear Blind Extraction in the Presence of Ill-Conditioned Mixing. *IEEE Transactions on Circuits and Systems I: Regular Papers* **55**:9, 2631-2638. [Crossref]
462. Jing Sui, Vince D. Calhoun. Exploration of the optimal group-discriminating features using CC-ICA 1410-1414. [Crossref]
463. Baeksuk Chu, Daehie Hong, Jooyoung Park, Jae-Hun Chung. 2008. Passive dynamic walker controller design employing an RLS-based natural actor–critic learning algorithm. *Engineering Applications of Artificial Intelligence* **21**:7, 1027-1034. [Crossref]
464. Jun-Mei Yang, Hideaki Sakai. An ICA-based adaptive filter algorithm for system identification using a state space approach 244-247. [Crossref]
465. Kazue Sega, Yohei Nakada, Takashi Matsumoto. Online Bayesian learning for dynamical classification problem using natural sequential prior 392-397. [Crossref]
466. Nizar Bouguila. A data-driven mixture kernel for count data classification using support vector machines 26-31. [Crossref]
467. Guang Deng, Wai-Yin Ng. 2008. An adaptive penalized maximum likelihood algorithm. *Signal Processing* **88**:9, 2323-2334. [Crossref]
468. Sam McKennoch, Thomas Voegtlin, Linda Bushnell. 2008. Spike-Timing Error Backpropagation in Theta Neuron Networks. *Neural Computation*, ahead of print080804143617793-37. [Crossref]
469. F. Cousseau, T. Ozeki, S.-i. Amari. 2008. Dynamics of Learning in Multilayer Perceptrons Near Singularities. *IEEE Transactions on Neural Networks* **19**:8, 1313-1328. [Crossref]
470. B. Baddeley. 2008. Reinforcement Learning in Continuous Time and Space: Interference and Not Ill Conditioning Is the Main Problem When Using Distributed Function Approximators. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **38**:4, 950-956. [Crossref]
471. Takeshi Amishima, Atsushi Okamura, Shinichi Morita, Tetsuo Kirimoto. Permutation scheme by tracking association for ICA separated signal blocks 3431-3435. [Crossref]
472. Tiemin Mei, Alfred Mertins, Fuliang Yin, Jiangtao Xi, Joe F. Chicharo. 2008. Blind source separation for convolutive mixtures based on the joint diagonalization of power spectral density matrices. *Signal Processing* **88**:8, 1990-2007. [Crossref]
473. Zhishun Wang, Bradley S. Peterson. 2008. Partner-matching for the automated identification of reproducible ICA components from fMRI datasets: Algorithm and validation. *Human Brain Mapping* **29**:8, 875-893. [Crossref]
474. Ana González, José R. Dorronsoro. 2008. Natural conjugate gradient training of multilayer perceptrons. *Neurocomputing* **71**:13-15, 2499-2506. [Crossref]

475. U. Manmontri, P.A. Naylor. 2008. A Class of Frobenius Norm-Based Algorithms Using Penalty Term and Natural Gradient for Blind Signal Separation. *IEEE Transactions on Audio, Speech, and Language Processing* **16**:6, 1181-1193. [Crossref]

476. Tiemin Mei, Alfred Mertins, Fuliang Yin. On the generalization of blind source separation algorithms from instantaneous to convolutive mixtures 482-486. [Crossref]

477. Ryota Horie. 2008. An optimization framework of biological dynamical systems. *Journal of Theoretical Biology* **253**:1, 45-54. [Crossref]

478. Simone Fiori. 2008. Leap-frog-type learning algorithms over the Lie group of unitary matrices. *Neurocomputing* **71**:10-12, 2224-2244. [Crossref]

479. V.G. Reju, Soo Ngee Koh, Ing Yann Soon. 2008. Partial separation method for solving permutation problem in frequency domain blind source separation of speech signals. *Neurocomputing* **71**:10-12, 2098-2112. [Crossref]

480. Miguel Barao, Joao M. Lemos. An efficient Kullback-Leibler optimization algorithm for probabilistic control design 198-203. [Crossref]

481. Daan Wierstra, Tom Schaul, Jan Peters, Juergen Schmidhuber. Natural Evolution Strategies 3381-3387. [Crossref]

482. Wai Yie Leong, Wei Liu, Danilo P. Mandic. 2008. Blind source extraction: Standard approaches and extensions to noisy and post-nonlinear mixing. *Neurocomputing* **71**:10-12, 2344-2355. [Crossref]

483. Wanlong Li, Ju Liu, Jun Du, Shuzhong Bail. Solving permutation problem in frequency-domain blind source separation using microphone sub-arrays 67-72. [Crossref]

484. Anand Oka, Lutz Lampe. 2008. Energy Efficient Distributed Filtering With Wireless Sensor Networks. *IEEE Transactions on Signal Processing* **56**:5, 2062-2075. [Crossref]

485. Jani Even, Hiroshi Saruwatari, Kiyohiro Shikano. Frequency Domain Blind Signal Extraction: Application to Fast Estimation of Diffuse Background Noise 212-215. [Crossref]

486. Byungchan Kim, Byungduk Kang, Shinsuk Park, Sungchul Kang. Learning robot stiffness for contact tasks using the natural actor-critic 3832-3837. [Crossref]

487. Jan Peters, Stefan Schaal. 2008. Reinforcement learning of motor skills with policy gradients. *Neural Networks* **21**:4, 682-697. [Crossref]

488. Simone Fiori. 2008. A Study on Neural Learning on Manifold Foliations: The Case of the Lie Group SU(3). *Neural Computation* **20**:4, 1091-1117. [Abstract] [PDF] [PDF Plus]

489. SIMONE FIORI. 2008. LEARNING BY CRITERION OPTIMIZATION ON A UNITARY UNIMODULAR MATRIX GROUP. *International Journal of Neural Systems* **18**:02, 87-103. [Crossref]

490. Zhaoshui He, Shengli Xie, Liqing Zhang, Andrzej Cichocki. 2008. A Note on Lewicki-Sejnowski Gradient for Learning Overcomplete Representations. *Neural Computation* **20**:3, 636-643. [Abstract] [PDF] [PDF Plus]

491. Haikun Wei, Jun Zhang, Florent Cousseau, Tomoko Ozeki, Shun-ichi Amari. 2008. Dynamics of Learning Near Singularities in Layered Networks. *Neural Computation* **20**:3, 813-843. [Abstract] [PDF] [PDF Plus]

492. Ah Chung Tsoi, Liangsuo Ma. 2008. A Balanced Approach to Multichannel Blind Deconvolution. *IEEE Transactions on Circuits and Systems I: Regular Papers* **55**:2, 599-613. [Crossref]

493. Jacek P. Dmochowski, Zicheng Liu, Philip A. Chou. Blind source separation in a distributed microphone meeting environment for improved teleconferencing 89-92. [Crossref]

494. S. Squartini, A. Arcangeli, F. Piazza. 2008. Stability analysis of natural gradient learning rules in overdetermined ICA. *Signal Processing* **88**:3, 761-766. [Crossref]

495. Jonathan Le Roux, Hirokazu Kameoka, Nobutaka Ono, Shigeki Sagayama, Alain de Cheveigne. Modulation analysis of speech through orthogonal FIR filterbank optimization 4189-4192. [Crossref]

496. Michael R. Bastian, Jacob H. Gunther, Todd K. Moon. Sobolev Gradients and Neural Networks 2085-2088. [Crossref]

497. Jan Peters, Stefan Schaal. 2008. Natural Actor-Critic. *Neurocomputing* **71**:7-9, 1180-1190. [Crossref]

498. Nuo Zhang, Jianming Lu, Takashi Yahagi. 2008. A method of independent component analysis based on radial basis function networks using noise estimation. *Electronics and Communications in Japan* **91**:3, 45-52. [Crossref]

499. Jean R. Paul, Tanya Vladimirova. Blind equalization with recurrent neural networks using natural gradient 178-183. [Crossref]

500. Zhirong Yang, Jorma Laaksonen. 2008. Principal whitened gradient for information geometry. *Neural Networks* **21**:2-3, 232-240. [Crossref]

501. T.E. Abrudan, J. Eriksson, V. Koivunen. 2008. Steepest Descent Algorithms for Optimization Under Unitary Matrix Constraint. *IEEE Transactions on Signal Processing* **56**:3, 1134-1147. [Crossref]

502. Hualiang Li, T. Adali. 2008. A Class of Complex ICA Algorithms Based on the Kurtosis Cost Function. *IEEE Transactions on Neural Networks* **19**:3, 408-420. [Crossref]
503. Hirofumi Nakajima, Kazuhiro Nakadai, Yuji Hasegawa, Hiroshi Tsujino. Adaptive step-size parameter control for real-world blind source separation 149-152. [Crossref]
504. J. A. Palmer, S. Makeig, K. Kreutz-Delgado, B. D. Rao. Newton method for the ICA mixture model 1805-1808. [Crossref]
505. David N. Levin. 2008. Using state space differential geometry for nonlinear blind source separation. *Journal of Applied Physics* **103**:4, 044906. [Crossref]
506. Michel Narozny, Michel Barret, Dinh-Tuan Pham. 2008. ICA based algorithms for computing optimal 1-D linear block transforms in variable high-rate source coding. *Signal Processing* **88**:2, 268-283. [Crossref]
507. Hualiang Li, Tülay Adalı. 2008. Complex-Valued Adaptive Signal Processing Using Nonlinear Functions. *EURASIP Journal on Advances in Signal Processing* **2008**:1, 765615. [Crossref]
508. Amar Kachenoura, Laurent Albera, Lotfi Senhadji, Pierre Comon. 2008. Ica: a potential tool for bci systems. *IEEE Signal Processing Magazine* **25**:1, 57-68. [Crossref]
509. A. Al-Hinai, M. Ibnkahla. Neural Network Nonlinear MIMO Channel Identification and Receiver Design 835-839. [Crossref]
510. Rafik Zayani, Ridha Bouallegue, Daniel Roviras. 2008. Adaptive Predistortions Based on Neural Networks Associated with Levenberg-Marquardt Algorithm for Satellite Down Links. *EURASIP Journal on Wireless Communications and Networking* **2008**:1, 132729. [Crossref]
511. Sertan Girgin, Philippe Preux. Basis Expansion in Natural Actor Critic Methods 110-123. [Crossref]
512. M. Mohan Sondhi. Adaptive Echo Cancelation for Voice Signals 903-928. [Crossref]
513. Tuan A. Duong, Margaret A. Ryan, Vu A. Duong. 2007. Space Invariant Independent Component Analysis and ENose for Detection of Selective Chemicals in an Unknown Environment. *Journal of Advanced Computational Intelligence and Intelligent Informatics* **11**:10, 1197-1203. [Crossref]
514. Zhirong Yang, Jorma Laaksonen. 2007. Multiplicative updates for non-negative projections. *Neurocomputing* **71**:1-3, 363-373. [Crossref]
515. Bin Xia, Hong Xie. Blind Source Separation of Temporal Correlated Signals 549-555. [Crossref]
516. Guillaume Jouffroy. Design of Simple Limit Cycles with Recurrent Neural Networks for Oscillatory Control 50-55. [Crossref]
517. Anand Oka, Lutz Lampe. Model Identification for Wireless Sensor Networks 3013-3018. [Crossref]
518. Liangsuo Ma, Ah Chung Tsoi. 2007. A unified balanced approach to multichannel blind deconvolution. *Signal, Image and Video Processing* **1**:4, 369-384. [Crossref]
519. Florent Guenter, Aude G. Billard. Using reinforcement learning to adapt an imitation task 1022-1027. [Crossref]
520. Ngo Anh Vien, TaeChoong Chung. Natural Gradient Policy for Average Cost SMDP Problem 11-18. [Crossref]
521. Amirali Shayan Arani, Yi Zhu, Yi-Ning Cheng, Chung-Kuan Cheng, Shien-Fong Lin, Peng-Sheng Chen. Exploring Cardioneural Signals from Noninvasive ECG Measurement 1134-1138. [Crossref]
522. Amirali Shayan Arani, Yi Zhu, Wanping Zhang, Tzyy-Ping Jung, Jeng-Ren Duann, Scott Makeig, Chung-Kuan Cheng. Spatial Density Reduction in the Study of the ECG Signal using Independent Component Analysis 5497-5500. [Crossref]
523. Yutaka Nakamura, Takeshi Mori, Masa-aki Sato, Shin Ishii. 2007. Reinforcement learning for a biped robot based on a CPG-actor-critic method. *Neural Networks* **20**:6, 723-735. [Crossref]
524. Hua Cai, Junxi Sun, Shifeng Ou. Blind Speech Separation Employing Laplacian Normal Mixture Distribution Model 3185-3189. [Crossref]
525. Zhirong Yang, Jorma Laaksonen. Approximated Geodesic Updates with Principal Natural Gradients 1320-1325. [Crossref]
526. Masatoshi Funabashi, Kazuyuki Aihara. 2007. Modeling birdsong learning with a chaotic Elman network. *Artificial Life and Robotics* **11**:2, 162-166. [Crossref]
527. Ana González, José R. Dorronsoro. 2007. Natural learning in NLDA networks. *Neural Networks* **20**:5, 610-620. [Crossref]
528. Zhenhai Wang, C. H. Chen. Adaptive Blind Source Separation in Underwater Wireless Speech Communication 1-6. [Crossref]
529. Jani Even, Kenji Sugimoto. 2007. An ICA approach to semi-blind identification of strictly proper systems based on interactor polynomial matrix. *International Journal of Robust and Nonlinear Control* **17**:8, 752-768. [Crossref]
530. Takaaki Ishibashi, Shingo Tamatsuka, Masataka Sugahara, Katsuhiro Inoue, Hiromu Gotanda, Kousuke Kumamaru. 2007. Separation Performance of ICA Algorithms Dependent on the Source Number. *Proceedings of the ISCIE International Symposium on Stochastic Systems Theory and its Applications* **2007**:0, 94-99. [Crossref]

531. Takaaki Ishibashi, Masataka Sugahara, Shingo Tamatsuka, Katsuhiro Inoue, Hiromu Gotanda, Kousuke Kumamaru. 2007. Estimation of the Number of Unknown Source Signals and Its Application to EEG Analysis. *Proceedings of the ISCIE International Symposium on Stochastic Systems Theory and its Applications* **2007**:0, 100-105. [Crossref]
532. Fuliang Yin, Tiemin Mei, Jun Wang. 2007. Blind-Source Separation Based on Decorrelation and Nonstationarity. *IEEE Transactions on Circuits and Systems I: Regular Papers* **54**:5, 1150-1158. [Crossref]
533. Yen-Wei Chen, Xianhua Han, Shiro Oikawa, Akinori Fujita. Independent Component Analysis for Removing X-ray Scatter in X-ray Images 1-4. [Crossref]
534. Aiko Furukawa, Junji Kiyono. 2007. Separation of harmonic excitation responses from contaminated measurements based on ICA. *Engineering Structures* **29**:4, 591-608. [Crossref]
535. Martin Riedmiller, Jan Peters, Stefan Schaal. Evaluation of Policy Gradient Methods and Variants on the Cart-Pole Benchmark 254-261. [Crossref]
536. Evaldo Arajo de Oliveira. 2007. The Rosenblatt Bayesian Algorithm Learning in a Nonstationary Environment. *IEEE Transactions on Neural Networks* **18**:2, 584-588. [Crossref]
537. A. Kachenoura, L. Albera, L. Senhadji. 2007. Séparation aveugle de sources en ingénierie biomédicale. *IRBM* **28**:1, 20-34. [Crossref]
538. Heeyoul Choi, Seungjin Choi. 2007. A relative trust-region algorithm for independent component analysis. *Neurocomputing* **70**:7-9, 1502-1510. [Crossref]
539. N. Hironaga, A.A. Ioannides. 2007. Localization of individual area neuronal activity. *NeuroImage* **34**:4, 1519-1534. [Crossref]
540. Lei Li, Fei Yan, Fu-Zheng Zha, Ling-Ye Nie. A New Independent Component Analysis Algorithm Based on Extended-Natural Gradient 2416-2420. [Crossref]
541. Tadahiro Azetsu, Eiji Uchino, Noriaki Suetake. 2007. Blind Separation and Sound Localization by Using Frequency-domain ICA. *Soft Computing* **11**:2, 185-192. [Crossref]
542. Jun-Mei Yang, Hideaki Sakai. A New Adaptive Filter Algorithm for System Identification using Independent Component Analysis III-1341-III-1344. [Crossref]
543. Jingyu Liu, Vince Calhoun. PARALLEL INDEPENDENT COMPONENT ANALYSIS FOR MULTIMODAL ANALYSIS: APPLICATION TO FMRI AND EEG DATA 1028-1031. [Crossref]
544. Masashi Ohata, Kiyotoshi Matsuoka, Toshiharu Mukai. 2007. An adaptive blind separation method using para-Hermitian whitening filter for convolutively mixed signals. *Signal Processing* **87**:1, 33-50. [Crossref]
545. Stefano Squartini, Andrea Arcangeli, Francesco Piazza. 2007. Stability Analysis of Natural Gradient Learning Rules in Complete ICA: A Unifying Perspective. *IEEE Signal Processing Letters* **14**:1, 54-57. [Crossref]
546. Meenakshi Singh, Deepak Kumar Singh, Prem K. Kalra. Multilayer Generalized Mean Neuron model for Blind Source Separation 562-566. [Crossref]
547. Maha Elsabrouty. Fast Converging Blind Signal Separation Algorithm using the Bussgang Cost Function and the Natural Gradient 229-232. [Crossref]
548. Mark D. Plumbley. Geometry and Manifolds for Independent Component Analysis IV-1397-IV-1400. [Crossref]
549. Tilay Adall, Hualiang Li. A Practical Formulation for Computation of Complex Gradients and its Application to Maximum Likelihood ICA II-633-II-636. [Crossref]
550. Yasunori Nishimori, Shotaro Akaho, Samer Abdallah, Mark D. Plumbley. Flag Manifolds for Subspace ICA Problems IV-1417-IV-1420. [Crossref]
551. Seungjin Choi. 2006. Differential learning algorithms for decorrelation and independent component analysis. *Neural Networks* **19**:10, 1558-1567. [Crossref]
552. Kenji Nakayama, Haruo Katou, Akihiro Hirano. A Feedback Approach and Its Learning Algorithm for Over Complete Blind Source Separation 459-462. [Crossref]
553. Hiroki Matsumoto, Tatsuki Kashihara, Toshihiro Furukawa. A Proposal of Echo Canceller using Blind Signal Separation method 467-469. [Crossref]
554. Liangsuo Ma, Ah Chung Tsoi. 2006. Balanced parameterization of multichannel blind deconvolutive systems: A continuous time realization. *Neurocomputing* **70**:1-3, 206-218. [Crossref]
555. Hyung-Min Park, Sang-Hoon Oh, Soo-Young Lee. 2006. A modified infomax algorithm for blind signal separation. *Neurocomputing* **70**:1-3, 229-240. [Crossref]

556. Ryo Mukai, Hiroshi Sawada, Shoko Araki, Shoji Makino. 2006. Frequency-Domain Blind Source Separation of Many Speech Signals Using Near-Field and Far-Field Models. *EURASIP Journal on Advances in Signal Processing* **2006**:1. . [Crossref]
557. Bin Xia, Liqing Zhang. 2006. Blind Deconvolution in Nonminimum Phase Systems Using Cascade Structure. *EURASIP Journal on Advances in Signal Processing* **2007**:1. . [Crossref]
558. T. Mei, J. Xi, F. Yin, A. Mertins, J.F. Chicharo. 2006. Blind Source Separation Based on Time-Domain Optimization of a Frequency-Domain Independence Criterion. *IEEE Transactions on Audio, Speech and Language Processing* **14**:6, 2075-2085. [Crossref]
559. Sheng Wan, L.E. Banta. 2006. Parameter Incremental Learning Algorithm for Neural Networks. *IEEE Transactions on Neural Networks* **17**:6, 1424-1438. [Crossref]
560. H. Sawada, S. Araki, R. Mukai, S. Makino. 2006. Blind Extraction of Dominant Target Sources Using ICA and Time-Frequency Masking. *IEEE Transactions on Audio, Speech and Language Processing* **14**:6, 2165-2173. [Crossref]
561. Seung-Hyun Jin, Yong-Ju Kwon, Jin-Su Jeong, Suk Won Kwon, Dong-Hoon Shin. 2006. Increased information transmission during scientific hypothesis generation: Mutual information analysis of multichannel EEG. *International Journal of Psychophysiology* **62**:2, 337-344. [Crossref]
562. Keiji Miura, Masato Okada, Shun-ichi Amari. 2006. Estimating Spiking Irregularities Under Changing Environments. *Neural Computation* **18**:10, 2359-2386. [Abstract] [PDF] [PDF Plus]
563. Jimin Ye, Xianda Zhang, Xiaolong Zhu. 2006. Blind source separation with unknown and dynamically changing number of source signals. *Science in China Series F: Information Sciences* **49**:5, 627-638. [Crossref]
564. Jan Peters, Stefan Schaal. Policy Gradient Methods for Robotics 2219-2225. [Crossref]
565. Iren Valova, Natacha Gueorguieva, Georgi Georgiev. Blind Source Separation with Neural Networks: Demixing Sources From Mixtures with Different Parameters 1-11. [Crossref]
566. Jayanta Basak. 2006. Online Adaptive Decision Trees: Pattern Classification and Function Approximation. *Neural Computation* **18**:9, 2062-2101. [Abstract] [PDF] [PDF Plus]
567. T. Murakami, T. Tanaka, Y. Ishida. A Fast Algorithm for ICA Deduced from a Closed-Form Solution of Kurtosis Maximization 223-228. [Crossref]
568. Shin Ishii, Wako Yoshida. 2006. Part 4: Reinforcement learning: Machine learning and natural learning. *New Generation Computing* **24**:3, 325-350. [Crossref]
569. Anthony Lombard, Herbert Buchner, Walter Kellermann. Multidimensional Localization of Multiple Sound Sources Using Blind Adaptive MIMO System Identification 7-12. [Crossref]
570. Seung-hyon Nam. Frequency-Domain Normalized Multichannel Blind Deconvolution for Convolutive Speech Mixtures: Modifications and Properties 307-312. [Crossref]
571. Jacob Gunther, Todd Moon. A Natural Gradient Algorithm for Multichannel Blind Deconvolution: Frequency Domain Criteria and Time Domain Updates 60-65. [Crossref]
572. Wai Leong, Danilo Mandic. Towards Adaptive Blind Extraction of Post-Nonlinearly Mixed Signals 91-96. [Crossref]
573. Wai Yie Leong, Danilo P. Mandic. Blind Sequential Extraction of Post-Nonlinearly Mixed Sources using Kalman Filtering 137-140. [Crossref]
574. Sanggyun Kim, Chang Yoo. Underdetermined Blind Source Separation Based on Generalized Gaussian Distribution 103-108. [Crossref]
575. Arthur C. Tsai, Michelle Liou, Tzyy-Ping Jung, Julie A. Onton, Philip E. Cheng, Chien-Chih Huang, Jeng-Ren Duann, Scott Makeig. 2006. Mapping single-trial EEG records on the cortical surface through a spatiotemporal modality. *NeuroImage* **32**:1, 195-207. [Crossref]
576. Samareh Samadi, Massoud Babaie-Zadeh, Christian Jutten. 2006. Quasi-optimal EASI algorithm based on the Score Function Difference (SFD). *Neurocomputing* **69**:13-15, 1415-1424. [Crossref]
577. Makoto Terumitsu, Yukihiko Fujii, Kiyotaka Suzuki, Ingrid L. Kwee, Tsutomu Nakada. 2006. Human primary motor cortex shows hemispheric specialization for speech. *NeuroReport* **17**:11, 1091-1095. [Crossref]
578. Patrick A. Naylor, Jingjing Cui, Mike Brookes. 2006. Adaptive algorithms for sparse echo cancellation. *Signal Processing* **86**:6, 1182-1192. [Crossref]
579. Huan Tao, Jian-yun Zhang, Lin Yu. Adaptive Blind Source Seperation Using Temporal Predictability 280-283. [Crossref]
580. Qihong Huang, Hu Chen, Zhao Liu. Texture Classification Using Spectral Histogram Representations and SVMs 226-229. [Crossref]

581. T. Ishibashi, K. Inoue, H. Gotanda, K. Kumamaru. 2006. Properties of Acoustic Transfer Function in Real Environments and Their Application to ICA. *Proceedings of the ISCIE International Symposium on Stochastic Systems Theory and its Applications* **2006**:0, 246-251. [Crossref]

582. Shun-ichi Amari, Hyeyoung Park, Tomoko Ozeki. 2006. Singularities Affect Dynamics of Learning in Neuromanifolds. *Neural Computation* **18**:5, 1007-1065. [Abstract] [PDF] [PDF Plus]

583. Xiaolong Zhu, Xianda Zhang. 2006. A signal-adaptive algorithm for blind separation of sources with mixed kurtosis signs. *Journal of Electronics (China)* **23**:3, 399-403. [Crossref]

584. Xiao-Long Zhu, Xian-Da Zhang, Ji-Min Ye. 2006. A Generalized Contrast Function and Stability Analysis for Overdetermined Blind Separation of Instantaneous Mixtures. *Neural Computation* **18**:3, 709-728. [Abstract] [PDF] [PDF Plus]

585. Xiao-Long Zhu, Xian-Da Zhang, Zi-Zhe Ding, Ying Jia. 2006. Adaptive nonlinear PCA algorithms for blind source separation without prewhitening. *IEEE Transactions on Circuits and Systems I: Regular Papers* **53**:3, 745-753. [Crossref]

586. Hyeyong Park. 2006. Part 2: Multilayer perceptron and natural gradient learning. *New Generation Computing* **24**:1, 79-95. [Crossref]

587. W. Wan. 2006. Implementing Online Natural Gradient Learning: Problems and Solutions. *IEEE Transactions on Neural Networks* **17**:2, 317-329. [Crossref]

588. Xianhua Zeng, Siwei Luo. An Improved Natural Gradient ICA Algorithm Based on Lie Group Invariance . [Crossref]

589. Takaaki Ishibashi, Katsuhiro Inoue, Hiromu Gotanda, Kousuke Kumamaru. Studies on Estimation of the Sources Number in Blind Source Separation Problems 5169-5174. [Crossref]

590. Wai Yie Leong, John Homer, Danilo P. Mandic. 2006. An Implementation of Nonlinear Multiuser Detection in Rayleigh Fading Channel. *EURASIP Journal on Wireless Communications and Networking* **2006**, 1-9. [Crossref]

591. V. Nikulin. Learning with Mean-Variance Filtering, SVM and Gradient-based Optimization 2193-2200. [Crossref]

592. W.Y. Leong, J. Homer. 2006. Blind multiuser receiver for DS-CDMA wireless system. *IEE Proceedings - Communications* **153**:5, 733. [Crossref]

593. K. Ikeda. On Geometric Structure of Quasi-Additive Learning Algorithms 1123-1127. [Crossref]

594. Kaiping Lin, Long Jin, Jianling Lin, Binglian Chen. The Relationship Between the Multi-Collinearity and the Generalization Capability of the Neural Network Forecast Model 56-60. [Crossref]

595. Nuo Zhang, Jianming Lu, Takashi Yahagi. 2006. A Method of Independent Component Analysis Based on Radial Basis Function Networks Using Noise Estimation. *IEEJ Transactions on Electronics, Information and Systems* **126**:6, 780-787. [Crossref]

596. Lingxiang Zhu, Liang Zou. Application of Cluster Analysis Neural Network in the Safety Evaluation for the Urban Underground Gas Pipeline 2845-2849. [Crossref]

597. A. Gonzalez, J.R. Dorronsoro. A Note on Conjugate Natural Gradient Training of Multilayer Perceptrons 887-891. [Crossref]

598. P. Gao, W.L. Woo, S.S. Dlay. 2006. Weierstrass approach to blind source separation of multiple nonlinearly mixed signals. *IEE Proceedings - Circuits, Devices and Systems* **153**:4, 332. [Crossref]

599. Lei Li, Yu Wang, Xing-hui Wang. Natural Gradient Algorithm Based on a Class of Activation Functions and its Applications in BSS 2985-2989. [Crossref]

600. P. Gao, W.L. Woo, S.S. Dlay. 2006. Non-linear independent component analysis using series reversion and Weierstrass network. *IEE Proceedings - Vision, Image, and Signal Processing* **153**:2, 115. [Crossref]

601. Lei Li. Flexible Kernel Independent Component Analysis Algorithm and its Local Stability on Feature Space 2990-2994. [Crossref]

602. X. Shen, H. Xu, F. Cong, J. Lei, K. Huang, Y. Zhang, G. Meng. 2006. Blind equalisation algorithm of FIR MIMO system in frequency domain. *IEE Proceedings - Vision, Image, and Signal Processing* **153**:5, 703. [Crossref]

603. T.A. Duong, De-Ling Liu, I. Kanik. Neural Network Prediction of Reduced Ion Mobility of Chemical Compound Based on Molecular Structure 1078-1084. [Crossref]

604. Fasong Wang, Hongwei Li, Rui Li. Data Mining with Independent Component Analysis 6043-6047. [Crossref]

605. J. Peters, S. Schaal. Reinforcement Learning for Parameterized Motor Primitives 73-80. [Crossref]

606. Heeyoul Choi, Seungjin Choi. Relative Gradient Learning for Independent Subspace Analysis 3919-3924. [Crossref]

607. Anbang Xu, Xin Jin, Ping Guo, R. Bie. KICA Feature Extraction in Application to FNN based Image Registration 3602-3608. [Crossref]

608. Akihiro Ichijo, Takehiro Hamada, Tetsuya Tabaru, Kazushi Nakano. A Combined Approach using Subspace and Beamforming Methods for Time-Frequency Domain Blind Source Separation 3280-3284. [Crossref]

609. Ling-Zhi Liao, Si-Wei Luo, Mei Tian, Lian-Wei Zhao. Learning Overcomplete Representations with a Generalized Gaussian Prior 432-441. [Crossref]

610. A. Eleuteri, R. Tagliaferri, L. Milano. 2005. A novel information geometric approach to variable selection in MLP networks. *Neural Networks* **18**:10, 1309-1318. [Crossref]

611. C.-T. Lin, W.-C. Cheng, S.-F. Liang. 2005. A 3-D Surface Reconstruction Approach Based on Postnonlinear ICA Model. *IEEE Transactions on Neural Networks* **16**:6, 1638-1650. [Crossref]

612. S. Fiori. 2005. Formulation and Integration of Learning Differential Equations on the Stiefel Manifold. *IEEE Transactions on Neural Networks* **16**:6, 1697-1701. [Crossref]

613. Harold Szu, Ivica Kopriva. 2005. Unsupervised learning with stochastic gradient. *Neurocomputing* **68**, 130-160. [Crossref]

614. Nizar Bouguila, Djemel Ziou. 2005. Using unsupervised learning of a finite Dirichlet mixture model to improve pattern recognition applications. *Pattern Recognition Letters* **26**:12, 1916-1925. [Crossref]

615. Jani Even, Eric Moisan. 2005. Blind source separation using order statistics. *Signal Processing* **85**:9, 1744-1758. [Crossref]

616. Mark D. Plumbley. 2005. Geometrical methods for non-negative ICA: Manifolds, Lie groups and toral subalgebras. *Neurocomputing* **67**, 161-197. [Crossref]

617. Yasunori Nishimori, Shotaro Akaho. 2005. Learning algorithms utilizing quasi-geodesic flows on the Stiefel manifold. *Neurocomputing* **67**, 106-135. [Crossref]

618. Xizhong Shen, Xizhi Shi. 2005. Online SOS-based multichannel blind equalization algorithm with noise. *Signal Processing* **85**:8, 1602-1610. [Crossref]

619. Anuj Srivastava, Xiuwen Liu. 2005. Tools for application-driven linear dimension reduction. *Neurocomputing* **67**, 136-160. [Crossref]

620. Hirokazu Asano, Hiroya Nakao. 2005. Independent Component Analysis of Spatiotemporal Chaos. *Journal of the Physical Society of Japan* **74**:6, 1661-1665. [Crossref]

621. W.L. Woo, S.S. Dlay. 2005. Neural network approach to blind signal separation of mono-nonlinearly mixed sources. *IEEE Transactions on Circuits and Systems I: Regular Papers* **52**:6, 1236-1247. [Crossref]

622. T. Ishibashi, K. Inoue, H. Gotanda, K. Kumamaru. 2005. A Solution of Permutation Problem Inherent in Frequency Domain ICA. *Proceedings of the ISCIE International Symposium on Stochastic Systems Theory and its Applications* **2005**:0, 259-264. [Crossref]

623. Massoud Babaie-Zadeh, Christian Jutten. 2005. A general approach for mutual information minimization and its application to blind source separation. *Signal Processing* **85**:5, 975-995. [Crossref]

624. Daniele Vigliano, Raffaele Parisi, Aurelio Uncini. 2005. An information theoretic approach to a novel nonlinear independent component analysis paradigm. *Signal Processing* **85**:5, 997-1028. [Crossref]

625. Shun-ichi Amari, Hiroyuki Nakahara. 2005. Difficulty of Singularity in Population Coding. *Neural Computation* **17**:4, 839-858. [Abstract] [PDF] [PDF Plus]

626. T. Tanaka. 2005. Generalized weighted rules for principal components tracking. *IEEE Transactions on Signal Processing* **53**:4, 1243-1253. [Crossref]

627. D. Erdogmus, O. Fontenla-Romero, J.C. Principe, A. Alonso-Betanzos, E. Castillo. 2005. Linear-Least-Squares Initialization of Multilayer Perceptrons Through Backpropagation of the Desired Response. *IEEE Transactions on Neural Networks* **16**:2, 325-337. [Crossref]

628. Huafu Chen, Dezhong Yao, Ling Zeng. 2005. A BFGS-ICA algorithm and application in localization of brain activities. *Neurocomputing* **64**, 513-519. [Crossref]

629. Kun Zhang, Lai-Wan Chan. 2005. Extended Gaussianization Method for Blind Separation of Post-Nonlinear Mixtures. *Neural Computation* **17**:2, 425-452. [Abstract] [PDF] [PDF Plus]

630. M. Inoue, M. Komiya, Y. Kabashima. An LDPCC decoding algorithm based on bowman-levin approximation -comparison with bp and CCCP- 444-448. [Crossref]

631. H. Matsumoto, T. Kashihara, T. Furukawa. A consideration on a stereo echo canceller using blind signal separation 301-304. [Crossref]

632. M. Tufail, M. Abe, M. Kawamata. Blind separation of mixed kurtosis signals using local exponential nonlinearities 39-42 Vol. 1. [Crossref]

633. Eiji Uchino, Noriaki Suetake, Morihiko Sakano. 2005. Blind deconvolution by using phase spectral constraints and natural gradient. *IEICE Electronics Express* **2**:9, 316-320. [Crossref]

634. M. Tufail, M. Abe, M. Kawamata. Truncated exponential nonlinearities for independent component analysis 381-384. [Crossref]

635. Jeng-Ren Duann, Tzyy-Ping Jung, S. Makeig, T.J. Sejnowski. Repeated decompositions reveal the stability of infomax decomposition of fMRI data 5324-5327. [Crossref]

636. Tadej Kosel, Igor Grabec, Franc Kosel. 2005. Intelligent location of two simultaneously active acoustic emission sources. *Aerospace Science and Technology* **9**:1, 45-53. [Crossref]

637. C. Xiang, S. Ding, T.H. Lee. 2005. Geometrical Interpretation and Architecture Selection of MLP. *IEEE Transactions on Neural Networks* **16**:1, 84-96. [Crossref]

638. MASATO INOUE, SHIN-ICHI NISHIMURA, GEN HORI, HIROYUKI NAKAHARA, MICHIKO SAITO, YOSHIHIRO YOSHIHARA, SHUN-ICHI AMARI. 2004. IMPROVED PARAMETER ESTIMATION FOR VARIANCE-STABILIZING TRANSFORMATION OF GENE-EXPRESSION MICROARRAY DATA. *Journal of Bioinformatics and Computational Biology* **02**:04, 669-679. [Crossref]

639. Elena Celledoni, Simone Fiori. 2004. Neural learning by geometric integration of reduced 'rigid-body' equations. *Journal of Computational and Applied Mathematics* **172**:2, 247-269. [Crossref]

640. HYUNG-MIN PARK, JONG-HWAN LEE, TAESU KIM, UN-MIN BAE, BYUNG TAEK KIM, KI-YOUNG PARK, CHANG-MIN KIM, SOO-YOUNG LEE. 2004. MODELING AUDITORY PATHWAY FOR INTELLIGENT INFORMATION ACQUISITION. *International Journal of Information Acquisition* **01**:04, 345-356. [Crossref]

641. N. Bouguila, D. Ziou, J. Vaillancourt. 2004. Unsupervised Learning of a Finite Mixture Model Based on the Dirichlet Distribution and Its Application. *IEEE Transactions on Image Processing* **13**:11, 1533-1543. [Crossref]

642. Masa-aki Sato, Taku Yoshioka, Shigeki Kajihara, Keisuke Toyama, Naokazu Goda, Kenji Doya, Mitsuo Kawato. 2004. Hierarchical Bayesian estimation for MEG inverse problem. *NeuroImage* **23**:3, 806-826. [Crossref]

643. A.J. Caamano, R. Boloix-Tortosa, J. Ramos, J.J. Murillo-Fuentes. 2004. Hybrid Higher-Order Statistics Learning in Multiuser Detection. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)* **34**:4, 417-424. [Crossref]

644. Jaakko Peltonen, Arto Klami, Samuel Kaski. 2004. Improved learning of Riemannian metrics for exploratory analysis. *Neural Networks* **17**:8-9, 1087-1100. [Crossref]

645. D.T. Pham. 2004. Fast Algorithms for Mutual Information Based Independent Component Analysis. *IEEE Transactions on Signal Processing* **52**:10, 2690-2700. [Crossref]

646. David Sodoyer, Laurent Girin, Christian Jutten, Jean-Luc Schwartz. 2004. Developing an audio-visual speech source separation algorithm. *Speech Communication* **44**:1-4, 113-125. [Crossref]

647. SIMONE FIORI. 2004. RELATIVE UNCERTAINTY LEARNING THEORY: AN ESSAY. *International Journal of Neural Systems* **14**:05, 293-311. [Crossref]

648. Jayanta Basak. 2004. Online Adaptive Decision Trees. *Neural Computation* **16**:9, 1959-1981. [Abstract] [PDF] [PDF Plus]

649. Shiro Ikeda, Toshiyuki Tanaka, Shun-ichi Amari. 2004. Stochastic Reasoning, Free Energy, and Information Geometry. *Neural Computation* **16**:9, 1779-1810. [Abstract] [PDF] [PDF Plus]

650. H. Sawada, R. Mukai, S. Araki, S. Makino. 2004. A Robust and Precise Method for Solving the Permutation Problem of Frequency-Domain Blind Source Separation. *IEEE Transactions on Speech and Audio Processing* **12**:5, 530-538. [Crossref]

651. Ji-Min Ye, Xiao-Long Zhu, Xian-Da Zhang. 2004. Adaptive Blind Separation with an Unknown Number of Sources. *Neural Computation* **16**:8, 1641-1660. [Abstract] [PDF] [PDF Plus]

652. M. Welling, R.S. Zemel, G.E. Hinton. 2004. Probabilistic Sequential Independent Components Analysis. *IEEE Transactions on Neural Networks* **15**:4, 838-849. [Crossref]

653. S.A. Cruces-Alvarez, A. Cichocki, S. Amari. 2004. From Blind Signal Extraction to Blind Instantaneous Signal Separation: Criteria, Algorithms, and Stability. *IEEE Transactions on Neural Networks* **15**:4, 859-873. [Crossref]

654. Huafu Chen, Dezhong Yao. 2004. Discussion on the choice of separated components in fMRI data analysis by spatial independent component analysis. *Magnetic Resonance Imaging* **22**:6, 827-833. [Crossref]

655. Scott Makeig, Arnaud Delorme, Marissa Westerfield, Tzyy-Ping Jung, Jeanne Townsend, Eric Courchesne, Terrence J Sejnowski. 2004. Electroencephalographic Brain Dynamics Following Manually Responded Visual Targets. *PLoS Biology* **2**:6, e176. [Crossref]

656. David Philipona, Olivier J.-M.D. Coenen. 2004. Model of granular layer encoding in the cerebellum. *Neurocomputing* **58-60**, 575-580. [Crossref]

657. N. Xu, X. Gao, B. Hong, X. Miao, S. Gao, F. Yang. 2004. BCI Competition 2003—Data Set IIb: Enhancing P300 Wave Detection Using ICA-Based Subspace Projections for BCI Applications. *IEEE Transactions on Biomedical Engineering* **51**:6, 1067-1072. [Crossref]

658. Scott Makeig, Stefan Debener, Julie Onton, Arnaud Delorme. 2004. Mining event-related brain dynamics. *Trends in Cognitive Sciences* **8**:5, 204-210. [Crossref]

659. Xiuwen Liu, A. Srivastava, K. Gallivan. 2004. Optimal linear representations of images for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**:5, 662-666. [Crossref]

660. Masato Inoue, Hyeyoung Park, Masato Okada. 2004. Dynamics of the adaptive natural gradient descent method for soft committee machines. *Physical Review E* **69**:5. . [Crossref]

661. L. Zhang, A. Cichocki, S. Amari. 2004. Multichannel Blind Deconvolution of Nonminimum-Phase Systems Using Filter Decomposition. *IEEE Transactions on Signal Processing* **52**:5, 1430-1442. [Crossref]

662. A. Ossadtchi, S. Baillet, J.C. Mosher, D. Thyerlei, W. Sutherling, R.M. Leahy. 2004. Automated interictal spike detection and source localization in magnetoencephalography using independent components analysis and spatio-temporal clustering. *Clinical Neurophysiology* **115**:3, 508-522. [Crossref]

663. L. Zhang, A. Cichocki, S. Amari. 2004. Self-Adaptive Blind Source Separation Based on Activation Functions Adaptation. *IEEE Transactions on Neural Networks* **15**:2, 233-244. [Crossref]

664. Weixiang Liu, Nanning Zheng, Xi Li. 2004. Relative gradient speeding up additive updates for nonnegative matrix factorization. *Neurocomputing* **57**, 493-499. [Crossref]

665. Hyeyoung Park, Noboru Murata, Shun-ichi Amari. 2004. Improving Generalization Performance of Natural Gradient Learning Using Optimized Regularization by NIC. *Neural Computation* **16**:2, 355-382. [Abstract] [PDF] [PDF Plus]

666. A. Dabrowski, D. Cetnarowicz, T. Marciniak. Analysis of speech separation for ASR systems 345-350. [Crossref]

667. Fabian J. Theis, Elmar W. Lang, Carlos G. Puntonet. 2004. A geometric algorithm for overcomplete linear ICA. *Neurocomputing* **56**, 381-398. [Crossref]

668. J. Jouffroy, J.-J.E. Slotine. Methodological remarks on contraction theory 2537-2543 Vol.3. [Crossref]

669. Simone Fiori, Pietro Burrascano. 2004. One-unit 'rigid-bodies' learning rule for principal/independent component analysis with application to ECT-NDE signal processing. *Neurocomputing* **56**, 233-255. [Crossref]

670. Jian Cheng, Qingshan Liu, Hanqing Lu. Texture classification using kernel independent component analysis 620-623 Vol.1. [Crossref]

671. Huafu Chen, Dezhong Yao. 2004. A composite ICA algorithm and the application in localization of brain activities. *Neurocomputing* **56**, 429-434. [Crossref]

672. SIMONE FIORI, ROBERTO ROSSI. 2003. STIEFEL-MANIFOLD LEARNING BY IMPROVED RIGID-BODY THEORY APPLIED TO ICA. *International Journal of Neural Systems* **13**:05, 273-290. [Crossref]

673. Bao-Liang Lu, Koji Ito. 2003. Converting general nonlinear programming problems into separable programming problems with feedforward neural networks. *Neural Networks* **16**:7, 1059-1074. [Crossref]

674. Alessandro Bissacco, Payam Saisan, Stefano Soatto. 2003. Modeling Human Gaits with Subtleties. *IFAC Proceedings Volumes* **36**:16, 1375-1380. [Crossref]

675. Shun-ichi Amari, Tomoko Ozeki, Hyeyoung Park. 2003. Learning and inference in hierarchical models with singularities. *Systems and Computers in Japan* **34**:7, 34-42. [Crossref]

676. Simone Fiori. 2003. Singular Value Decomposition Learning on Double Stiefel Manifold. *International Journal of Neural Systems* **13**:03, 155-170. [Crossref]

677. Eiji Mizutani, James W. Demmel. 2003. On structure-exploiting trust-region regularized nonlinear least squares algorithms for neural-network learning. *Neural Networks* **16**:5-6, 745-753. [Crossref]

678. Shun-Tian Lou, Xian-Da Zhang. 2003. Fuzzy-based learning rate determination for blind source separation. *IEEE Transactions on Fuzzy Systems* **11**:3, 375-383. [Crossref]

679. Jianting Cao, N. Murata, S.-i. Amari, A. Cichocki, T. Takeda. 2003. A robust approach to independent component analysis of signals with high-level noise measurements. *IEEE Transactions on Neural Networks* **14**:3, 631-645. [Crossref]

680. M.D. Plumbley. 2003. Algorithms for nonnegative independent component analysis. *IEEE Transactions on Neural Networks* **14**:3, 534-543. [Crossref]

681. Masato Inoue, Hyeyoung Park, Masato Okada. 2003. On-Line Learning Theory of Soft Committee Machines with Correlated Hidden Units –Steepest Gradient Descent and Natural Gradient Descent–. *Journal of the Physical Society of Japan* **72**:4, 805-810. [Crossref]

682. Simone Fiori. 2003. Overview of independent component analysis technique with an application to synthetic aperture radar (SAR) imagery processing. *Neural Networks* **16**:3-4, 453-467. [Crossref]

683. Tadej Kosel, Igor Grabec, Franc Kosel. 2003. Intelligent location of two simultaneously active acoustic emission sources: Part II. *Aircraft Engineering and Aerospace Technology* **75**:2, 137-142. [Crossref]

684. A. Bortoletti, C. Di Fiore, S. Fanelli, P. Zellini. 2003. A new class of quasi-newtonian methods for optimal learning in mlp-networks. *IEEE Transactions on Neural Networks* **14**:2, 263-273. [Crossref]

685. S. Araki, R. Mukai, S. Makino, T. Nishikawa, H. Saruwatari. 2003. The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech. *IEEE Transactions on Speech and Audio Processing* **11**:2, 109-116. [Crossref]

686. Yoshio Onozaka, Masahiro Nakagawa. 2003. Back propagation learning with periodic chaos neurons. *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)* **86**:3, 11-19. [Crossref]

687. H.S. Sahambi, K. Khorasani. 2003. A neural-network appearance-based 3-D object recognition using independent component analysis. *IEEE Transactions on Neural Networks* **14**:1, 138-149. [Crossref]

688. Kiyotaka SUZUKI, Hitoshi MATSUZAWA, Hironaka IGARASHI, Masaki WATANABE, Naoki NAKAYAMA, Ingrid L. KWEE, Tsutomu NAKADA. 2003. All-phase MR Angiography Using Independent Component Analysis of Dynamic Contrast Enhanced MRI Time Series: .PHI.-MRA. *Magnetic Resonance in Medical Sciences* **2**:1, 23-27. [Crossref]

689. Scott C. Douglas, Xiaoan Sun. 2003. Convolutive blind separation of speech mixtures using the natural gradient. *Speech Communication* **39**:1-2, 65-78. [Crossref]

690. Peixun Luo, K. Y. Michael Wong. 2003. Dynamical and stationary properties of on-line learning from finite training sets. *Physical Review E* **67**:1. . [Crossref]

691. Liqing Zhang, Bin Xia. Temporal independent component analysis for separating event-related potentials 1362-1365 Vol.2. [Crossref]

692. Jorge Igual, Luis Vergara, Andrés Camacho, Ramón Miralles. 2003. Independent component analysis with prior information about the mixing matrix. *Neurocomputing* **50**, 419-438. [Crossref]

693. Wei Lu, Jagath C. Rajapakse. 2003. Eliminating indeterminacy in ICA. *Neurocomputing* **50**, 271-290. [Crossref]

694. M. Ibnkahla, J. Yuan. A neural network MLSE receiver based on natural gradient descent: application to satellite communications 33-36 vol.1. [Crossref]

695. Nihat Ay. 2002. Locality of Global Stochastic Interaction in Directed Acyclic Networks. *Neural Computation* **14**:12, 2959-2980. [Abstract] [PDF] [PDF Plus]

696. Juan J. Murillo-Fuentes, Francisco J. González-Serrano. 2002. Median equivariant adaptive separation via independence: application to communications. *Neurocomputing* **49**:1-4, 389-409. [Crossref]

697. Wakako Hashimoto. 2002. Separation of independent components from data mixed by several mixing matrices. *Signal Processing* **82**:12, 1949-1961. [Crossref]

698. Jianting Cao, Noboru Murata, Shun-ichi Amari, Andrzej Cichocki, Tsunehiro Takeda. 2002. Independent component analysis for unaveraged single-trial MEG data decomposition and single-dipole source localization. *Neurocomputing* **49**:1-4, 255-277. [Crossref]

699. Heinz Mathis, Marcel Joho. 2002. Blind signal separation in noisy environments using a three-step quantizer. *Neurocomputing* **49**:1-4, 61-78. [Crossref]

700. Shotaro Akaho. 2002. Conditionally independent component analysis for supervised feature extraction. *Neurocomputing* **49**:1-4, 139-150. [Crossref]

701. Thomas P. von Hoff, Allen G. Lindgren. 2002. Adaptive step-size control in blind source separation. *Neurocomputing* **49**:1-4, 119-138. [Crossref]

702. Seungjin Choi, Heonseok Hong, Hervé Glotin, Frédéric Berthommier. 2002. Multichannel signal separation for cocktail party speech recognition: a dynamic recurrent network. *Neurocomputing* **49**:1-4, 299-314. [Crossref]

703. Xiao-Long Zhu, Xian-Da Zhang. 2002. Adaptive RLS algorithm for blind source separation using a natural gradient. *IEEE Signal Processing Letters* **9**:12, 432-435. [Crossref]

704. Sergio Cruces, Luis Castedo, Andrzej Cichocki. 2002. Robust blind source separation algorithms using cumulants. *Neurocomputing* **49**:1-4, 87-118. [Crossref]

705. M. Ibnkahla. 2002. Natural gradient learning neural networks for adaptive inversion of Hammerstein systems. *IEEE Signal Processing Letters* **9**:10, 315-317. [Crossref]

706. Simone Fiori. 2002. Blind deconvolution by simple adaptive activation function neuron. *Neurocomputing* **48**:1-4, 763-778. [Crossref]

707. Katsuyuki Hagiwara. 2002. Regularization learning, early stopping and biased estimator. *Neurocomputing* **48**:1-4, 937-955. [Crossref]

708. Todd K. Moon, Jacob H. Gunther. 2002. Contravariant adaptation on structured matrix spaces. *Signal Processing* **82**:10, 1389-1410. [Crossref]

709. Jordi Sole, Christian Jutten, Anisse Taleb. 2002. Parametric approach to blind deconvolution of nonlinear channels. *Neurocomputing* **48**:1-4, 339-355. [Crossref]

710. H. Abdulkader, F. Langlet, D. Roviras, F. Castanie. 2002. Natural gradient algorithm for neural networks applied to non-linear high power amplifiers. *International Journal of Adaptive Control and Signal Processing* **16**:8, 557-576. [Crossref]

711. I. Christoyianni, A. Koutras, E. Dermatas, G. Kokkinakis. 2002. Computer aided diagnosis of breast cancer in digitized mammograms. *Computerized Medical Imaging and Graphics* **26**:5, 309-319. [Crossref]

712. Si Wu, K Michael Wong. Neural Networks . [Crossref]

713. Katsuyuki Hagiwara. 2002. On the Problem in Model Selection of Neural Network Regression in Overrealizable Scenario. *Neural Computation* **14**:8, 1979-2002. [Abstract] [PDF] [PDF Plus]

714. Minami Mihoko, Shinto Eguchi. 2002. Robust Blind Source Separation by Beta Divergence. *Neural Computation* **14**:8, 1859-1886. [Abstract] [PDF] [PDF Plus]

715. R.K. Martin, W.A. Sethares, R.C. Williamson, C.R. Johnson. 2002. Exploiting sparsity in adaptive filters. *IEEE Transactions on Signal Processing* **50**:8, 1883-1894. [Crossref]

716. A. Taleb. 2002. A generic framework for blind source separation in structured nonlinear models. *IEEE Transactions on Signal Processing* **50**:8, 1819-1830. [Crossref]

717. Nicol N. Schraudolph. 2002. Fast Curvature Matrix-Vector Products for Second-Order Gradient Descent. *Neural Computation* **14**:7, 1723-1738. [Abstract] [PDF] [PDF Plus]

718. Fabian J. Theis, Ch. Bauer, Elmar W. Lang. 2002. Comparison of maximum entropy and minimal mutual information in a nonlinear setting. *Signal Processing* **82**:7, 971-980. [Crossref]

719. Noboru Murata, Motoaki Kawanabe, Andreas Ziehe, Klaus-Robert Müller, Shun-ichi Amari. 2002. On-line learning in changing environments with applications in supervised and unsupervised learning. *Neural Networks* **15**:4-6, 743-760. [Crossref]

720. SIMONE FIORI. 2002. UNSUPERVISED NEURAL LEARNING ON LIE GROUP. *International Journal of Neural Systems* **12**:03n04, 219-246. [Crossref]

721. S. Fiori. 2002. A theory for learning based on rigid bodies dynamics. *IEEE Transactions on Neural Networks* **13**:3, 521-531. [Crossref]

722. Mohamed Ibnkahla, Benoit Pochon. Natural gradient learning neural networks for modeling and identification of nonlinear systems with memory I-1057-I-1060. [Crossref]

723. Tadej Kosel, Igor Grabec, Franc Kosel. 2002. Time-delay estimation of acoustic emission signals using ICA. *Ultrasonics* **40**:1-8, 303-306. [Crossref]

724. Steven L. Gay, Scott C. Douglas. Normalized natural gradient adaptive filtering for sparse and non-sparse systems II-1405-II-1408. [Crossref]

725. Hiroshi Sawada, Ryo Mukai, Shoko Araki, Shoji Makino. Polar coordinate based nonlinear function for frequency-domain blind source separation I-1001-I-1004. [Crossref]

726. Juha Karvanen, Visa Koivunen. 2002. Blind separation methods based on Pearson system and its extensions. *Signal Processing* **82**:4, 663-673. [Crossref]

727. Tadej Kosel, Igor Grabec. 2002. Location of two simultaneously active continuous acoustic emission sources on an aluminum beam. *Aircraft Engineering and Aerospace Technology* **74**:1, 4-8. [Crossref]

728. Simone Fiori. 2002. Hybrid independent component analysis by adaptive LUT activation function neurons. *Neural Networks* **15**:1, 85-94. [Crossref]

729. Y. Matsuyama, R. Kawamura. Supervised map ICA: applications to brain functional MRI 2259-2263 vol.5. [Crossref]

730. Seungjin Choi, Andrzej Cichocki, Shunichi Amari. 2002. Equivariant nonstationary source separation. *Neural Networks* **15**:1, 121-130. [Crossref]

731. Yogesh Singh, C. S. Rai. 2002. An independent component analysis technique for blind source separation. *Journal of Interdisciplinary Mathematics* **5**:3, 231-241. [Crossref]

732. Kiyotaka Suzuki, Tohru Kiryu, Tsutomu Nakada. 2002. Fast and precise independent component analysis for high field fMRI time series tailored using prior information on spatiotemporal structure. *Human Brain Mapping* **15**:1, 54-66. [Crossref]

733. W.L. Woo, S. Sali. 2002. General multilayer perceptron demixer scheme for nonlinear blind signal separation. *IEE Proceedings - Vision, Image, and Signal Processing* **149**:5, 253. [Crossref]

734. G. GELLE, M. COLAS, C. SERVIERE. 2001. BLIND SOURCE SEPARATION: A TOOL FOR ROTATING MACHINE MONITORING BY VIBRATIONS ANALYSIS?. *Journal of Sound and Vibration* **248**:5, 865-885. [Crossref]

735. L.-Q. Zhang, S. Amari, A. Cichocki. 2001. Semiparametric model and superefficiency in blind deconvolution. *Signal Processing* **81**:12, 2535-2553. [Crossref]

736. Tianping Chen, Shun-ichi Amari. 2001. Unified stabilization approach to principal and minor components extraction algorithms. *Neural Networks* **14**:10, 1377-1387. [Crossref]

737. Mark Zlochin, Yoram Baram. 2001. Manifold Stochastic Dynamics for Bayesian Learning. *Neural Computation* **13**:11, 2549-2572. [Abstract] [PDF] [PDF Plus]

738. Shotaro Akaho, Shinji Umeyama. 2001. Multimodal independent component analysis?A method of feature extraction from multiple information sources. *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)* **84**:11, 21-28. [Crossref]

739. SIMONE FIORI. 2001. PROBABILITY DENSITY FUNCTION LEARNING BY UNSUPERVISED NEURONS. *International Journal of Neural Systems* **11**:05, 399-417. [Crossref]

740. Nikolaos Ampazis, Stavros J. Perantonis, John G. Taylor. 2001. A dynamical model for the analysis and acceleration of learning in feedforward networks. *Neural Networks* **14**:8, 1075-1088. [Crossref]

741. Simone Fiori. 2001. A contribution to (neuromorphic) blind deconvolution by flexible approximated Bayesian estimation. *Signal Processing* **81**:10, 2131-2153. [Crossref]

742. Harold Szu, Ivica Kopriva. 2001. Artificial neural networks for noisy image super-resolution. *Optics Communications* **198**:1-3, 71-81. [Crossref]

743. Dinh-Tuan Pham, J.-F. Cardoso. 2001. Blind separation of instantaneous mixtures of nonstationary sources. *IEEE Transactions on Signal Processing* **49**:9, 1837-1848. [Crossref]

744. Masa-aki Sato. 2001. Online Model Selection Based on the Variational Bayes. *Neural Computation* **13**:7, 1649-1681. [Abstract] [PDF] [PDF Plus]

745. Simone Fiori. 2001. A Theory for Learning by Weight Flow on Stiefel-Grassman Manifold. *Neural Computation* **13**:7, 1625-1647. [Abstract] [PDF] [PDF Plus]

746. T.-P. Jung, S. Makeig, M.J. McKeown, A.J. Bell, T.-W. Lee, T.J. Sejnowski. 2001. Imaging brain dynamics using independent component analysis. *Proceedings of the IEEE* **89**:7, 1107-1122. [Crossref]

747. S.-I. Amari. 2001. Information geometry on hierarchy of probability distributions. *IEEE Transactions on Information Theory* **47**:5, 1701-1711. [Crossref]

748. S. Kaski, J. Sinkkonen, J. Peltonen. 2001. Bankruptcy analysis with self-organizing maps in learning metrics. *IEEE Transactions on Neural Networks* **12**:4, 936-947. [Crossref]

749. Thomas Wachtler, Te-Won Lee, Terrence J. Sejnowski. 2001. Chromatic structure of natural scenes. *Journal of the Optical Society of America A* **18**:1, 65. [Crossref]

750. M. Sugiyama, H. Ogawa. 2001. Incremental projection learning for optimal generalization. *Neural Networks* **14**:1, 53-66. [Crossref]

751. M. Sugiyama, H. Ogawa. 2001. Properties of incremental projection learning. *Neural Networks* **14**:1, 67-78. [Crossref]

752. K. Fukumizu. Chapter 17 Geometry of neural networks: Natural gradient for learning 731-769. [Crossref]

753. T.K. Moon, J. Gunther. Contravariant adaptation on structured parameter spaces 936-940 vol.2. [Crossref]

754. L. Castedo-Ribas, A. Cichocki, S. Cruces-Alvarez. 2000. An iterative inversion approach to blind source separation. *IEEE Transactions on Neural Networks* **11**:6, 1423-1437. [Crossref]

755. Te-Won Lee, M.S. Lewicki, T.J. Sejnowski. 2000. ICA mixture models for unsupervised classification of non-Gaussian classes and automatic context switching in blind signal separation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**:10, 1078-1089. [Crossref]

756. Shun-ichi Amari. 2000. Estimating Functions of Independent Component Analysis for Temporally Correlated Signals. *Neural Computation* **12**:9, 2083-2107. [Abstract] [PDF] [PDF Plus]

757. Thomas P. von Hoff, Allen G. Lindgren, August N. Kaelin. 2000. Transpose properties in the stability and performance of the classic adaptive algorithms for blind source separation and deconvolution. *Signal Processing* **80**:9, 1807-1822. [Crossref]

758. H Park, S.-I Amari, K Fukumizu. 2000. Adaptive natural gradient learning algorithms for various stochastic models. *Neural Networks* **13**:7, 755-764. [Crossref]

759. Filipe Aires, Alain Chédin, Jean-Pierre Nadal. 2000. Independent component analysis of multivariate time series: Application to the tropical SST variability. *Journal of Geophysical Research: Atmospheres* **105**:D13, 17437-17455. [Crossref]

760. Mohamed Ibnkahla. 2000. Applications of neural networks to digital communications – a survey. *Signal Processing* **80**:7, 1185-1215. [Crossref]

761. Tsutomu Nakada, Kiyotaka Suzuki, Yukihiko Fujii, Hitoshi Matsuzawa, Ingrid L Kwee. 2000. Independent component-cross correlation-sequential epoch (ICS) analysis of high field fMRI time series: direct visualization of dual representation of the primary motor cortex in human. *Neuroscience Research* **37**:3, 237-244. [Crossref]

762. A. Navia-Vázquez, A. R. Figueiras-Vidal. 2000. Efficient Block Training of Multilayer Perceptrons. *Neural Computation* **12**:6, 1429-1447. [Abstract] [PDF] [PDF Plus]

763. Shun-ichi Amari, Tian-Ping Chen, Andrzej Cichocki. 2000. Nonholonomic Orthogonal Learning Algorithms for Blind Source Separation. *Neural Computation* **12**:6, 1463-1484. [Abstract] [PDF] [PDF Plus]

764. Shun-ichi Amari, Hyeyoung Park, Kenji Fukumizu. 2000. Adaptive Method of Realizing Natural Gradient Learning for Multilayer Perceptrons. *Neural Computation* **12**:6, 1399-1409. [Abstract] [PDF] [PDF Plus]

765. Te-Won Lee, M. Girolami, A.J. Bell, T.J. Sejnowski. 2000. A unifying information-theoretic framework for independent component analysis. *Computers & Mathematics with Applications* **39**:11, 1-21. [Crossref]

766. Tom Heskes. 2000. On "Natural" Learning and Pruning in Multilayered Perceptrons. *Neural Computation* **12**:4, 881-901. [Abstract] [PDF] [PDF Plus]

767. K. Fukumizu, S. Amari. 2000. Local minima and plateaus in hierarchical structures of multilayer perceptrons. *Neural Networks* **13**:3, 317-327. [Crossref]

768. Tianping Chen, Shun-ichi Amari. 2000. Dynamic behavior of the robust decorrelation process. *Neurocomputing* **30**:1-4, 143-151. [Crossref]

769. S. Fiori. Stiefel-Grassman flow (SGF) learning: further results 343-348 vol.3. [Crossref]

770. K. Hagiwara, K. Kuno. Regularization learning and early stopping in linear networks 511-516 vol.4. [Crossref]

771. Yang Chen, Zhenya He. A block-adaptive blind separation algorithm for post-nonlinear mixture of sub- and super-Gaussian signals WEB4/10-WEB4/17 vol.2. [Crossref]

772. Y. Mtsuyama, T. Nimoto, N. Katsumata, Y. Suzuki, S. Furukawa. /spl alpha/-EM algorithm and /spl alpha/-ICA learning based upon extended logarithmic information measures 351-356 vol.3. [Crossref]

773. J.J. Murillo-Fuentes, F.J. González-Serrano. 2000. Improving stability in blind source separation with stochastic median gradient. *Electronics Letters* **36**:19, 1662. [Crossref]

774. Dominik Endres, Peter Riegler. 1999. Learning dynamics on different timescales. *Journal of Physics A: Mathematical and General* **32**:49, 8655-8663. [Crossref]

775. Shun-ichi Amari. 1999. Natural Gradient Learning for Over- and Under-Complete Bases in ICA. *Neural Computation* **11**:8, 1875-1883. [Abstract] [PDF] [PDF Plus]

776. L.-Q. Zhang, A. Cichocki, S. Amari. 1999. Natural gradient algorithm for blind separation of overdetermined mixture with additive noise. *IEEE Signal Processing Letters* **6**:11, 293-295. [Crossref]

777. A. Taleb, C. Jutten. 1999. Source separation in post-nonlinear mixtures. *IEEE Transactions on Signal Processing* **47**:10, 2807-2820. [Crossref]

778. J. Basak, S. Amari. 1999. Blind separation of uniformly distributed signals: a general approach. *IEEE Transactions on Neural Networks* **10**:5, 1173-1185. [Crossref]

779. H.H. Yang. 1999. Serial updating rule for blind separation derived from the method of scoring. *IEEE Transactions on Signal Processing* **47**:8, 2279-2285. [Crossref]

780. Bruno Apolloni, Egidio Battistini, Diego de Falco. 1999. Higher-order Boltzmann machines and entropy bounds. *Journal of Physics A: Mathematical and General* **32**:30, 5529-5538. [Crossref]

781. Silvia Scarpetta, Magnus Rattray, David Saad. 1999. Matrix momentum for practical natural gradient learning. *Journal of Physics A: Mathematical and General* **32**:22, 4047-4059. [Crossref]

782. Piërre van de Laar, Tom Heskes. 1999. Pruning Using Parameter and Neuronal Metrics. *Neural Computation* **11**:4, 977-993. [Abstract] [PDF] [PDF Plus]
783. Jayanta Basak, Shun-ichi Amari. 1999. Blind Separation of a Mixture of Uniformly Distributed Source Signals: A Novel Approach. *Neural Computation* **11**:4, 1011-1034. [Abstract] [PDF] [PDF Plus]
784. S.C. Douglas. 1999. Equivariant adaptive selective transmission. *IEEE Transactions on Signal Processing* **47**:5, 1223-1231. [Crossref]
785. Filipe Aires, Alain Chédin, Jean-Pierre Nadal. 1999. Analyse de séries temporelles géophysiques et théorie de l'information: L'analyse en composantes indépendantes. *Comptes Rendus de l'Académie des Sciences - Series IIA - Earth and Planetary Science* **328**:9, 569-575. [Crossref]
786. Peter Dayan. 1999. Recurrent Sampling Models for the Helmholtz Machine. *Neural Computation* **11**:3, 653-677. [Abstract] [PDF] [PDF Plus]
787. Magnus Rattray, David Saad. 1999. Analysis of natural gradient descent for multilayer neural networks. *Physical Review E* **59**:4, 4523-4532. [Crossref]
788. Te-Won Lee, M.S. Lewicki, M. Girolami, T.J. Sejnowski. 1999. Blind source separation of more sources than mixtures using overcomplete representations. *IEEE Signal Processing Letters* **6**:4, 87-90. [Crossref]
789. S. Amari. 1999. Superefficiency in blind source separation. *IEEE Transactions on Signal Processing* **47**:4, 936-944. [Crossref]
790. Te-Won Lee, Mark Girolami, Terrence J. Sejnowski. 1999. Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Subgaussian and Supergaussian Sources. *Neural Computation* **11**:2, 417-441. [Abstract] [PDF] [PDF Plus]
791. Andrzej Cichocki, Juha Karhunen, Wlodzimierz Kasprzak, Ricardo Vigário. 1999. Neural networks for blind separation with unknown number of sources. *Neurocomputing* **24**:1-3, 55-93. [Crossref]
792. S. Choi, A. Cichocki, S. Amari. Two spatio-temporal decorrelation learning algorithms and their application to multichannel blind deconvolution 1085-1088 vol.2. [Crossref]
793. Magnus Rattray, David Saad, Shun-ichi Amari. 1998. Natural Gradient Descent for On-Line Learning. *Physical Review Letters* **81**:24, 5461-5464. [Crossref]
794. Howard Hua Yang, Shun-ichi Amari. 1998. Complexity Issues in Natural Gradient Descent Method for Training Multilayer Perceptrons. *Neural Computation* **10**:8, 2137-2157. [Abstract] [PDF] [PDF Plus]
795. Mark Girolami. 1998. An Alternative Perspective on Adaptive Independent Component Analysis Algorithms. *Neural Computation* **10**:8, 2103-2114. [Abstract] [PDF] [PDF Plus]
796. Magnus Rattray, David Saad. 1998. Analysis of on-line training with optimal learning rates. *Physical Review E* **58**:5, 6379-6391. [Crossref]
797. S. Amari, A. Cichocki. 1998. Adaptive blind signal processing-neural network approaches. *Proceedings of the IEEE* **86**:10, 2026-2048. [Crossref]
798. J.-F. Cardoso. 1998. Blind signal separation: statistical principles. *Proceedings of the IEEE* **86**:10, 2009-2025. [Crossref]
799. Siegfried Bös. 1998. Matrix-update for accelerated on-line learning in multilayer neural networks. *Journal of Physics A: Mathematical and General* **31**:22, L413-L417. [Crossref]
800. Tianping Chen, Shun Ichi Amari, Qin Lin. 1998. A unified algorithm for principal and minor components extraction. *Neural Networks* **11**:3, 385-390. [Crossref]
801. Kiyotoshi MATSUOKA. 1998. Blind Signal Separation. *Journal of Japan Society for Fuzzy Theory and Systems* **10**:3, 394-400. [Crossref]
802. Richard Hahnloser. 1998. Learning algorithms based on linearization. *Network: Computation in Neural Systems* **9**:3, 363-380. [Crossref]
803. Shun-Ichi Amari, J.-F. Cardoso. 1997. Blind source separation-semiparametric statistical approach. *IEEE Transactions on Signal Processing* **45**:11, 2692-2700. [Crossref]
804. Andrew D. Back, Andreas S. Weigend. 1997. A First Application of Independent Component Analysis to Extracting Structure from Stock Returns. *International Journal of Neural Systems* **08**:04, 473-484. [Crossref]
805. A. Taleb, C. Jutten, S. Olympieff. Source separation in post nonlinear mixtures: an entropy-based algorithm 2089-2092. [Crossref]
806. Eiji Mizutani, J.W. Demmel. On sparsity-exploiting memory-efficient trust-region regularized nonlinear least squares algorithms for neural-network learning 242-247. [Crossref]
807. S. Amari. Information geometry of statistical inference - an overview 86-89. [Crossref]
808. A.D. Back, A.S. Weigend. What drives stock returns?-an independent component analysis 141-156. [Crossref]

809. M. Ohata, T. Mukai, K. Matsuoka. Blind separation with Gaussian mixture model for convolutively mixed sources V-13-V-16. [Crossref]
810. Y. Matsuyama, S. Imahara, N. Katsumata. Optimization transfer for computational learning: a hierarchy from f-ICA and alpha-EM to their offsprings 1883-1888. [Crossref]
811. Liqing Zhang, Libo Ma, Wenlu Yang, Bin Xia. Sparse Representation and Synaptic Adaptation of the Visual Sensory System 1961-1964. [Crossref]
812. B. Lu, B.L. Evans. Channel equalization by feedforward neural networks 587-590. [Crossref]
813. A. Trouve, Yong Yu. "Learning the kernel" through examples: an application to shape classification 121-124. [Crossref]
814. Hsiao-Chun Wu, J.C. Principe, J.G. Harris, Jui-Kuo Juan. Loss function for blind source separation-minimum entropy criterion and its generalized anti-Hebbian rules 910-915. [Crossref]
815. H. Sawada, S. Araki, R. Mukai, S. Makino. Blind extraction of a dominant source from mixtures of many sources using ICA and time-frequency masking 5882-5885. [Crossref]
816. A.C. Tsoi, L.S. Ma. Blind deconvolution of dynamical systems using a balanced parameterized state space approach IV-309-12. [Crossref]
817. N. Suetake, Y. Nakamura, T. Yamakawa. Maximum entropy ICA constrained by individual entropy maximization employing self-organizing maps 1038-1042. [Crossref]
818. H. Imamura, J. Okello, H. Ochi. Blind source separation of PSK and amplitude modulated signals 343-346. [Crossref]
819. H. Imamura, H. Ochi. A high capacity MIMO system based on blind deconvolution 617-621. [Crossref]
820. Yang Chen, Zhenya He. Post-nonlinear source separation: hard switching versus soft learning 403-406. [Crossref]
821. H. Sawada, S. Araki, R. Mukai, S. Makino. Blind source separation with different sensor spacing and filter length for each frequency range 465-474. [Crossref]
822. H. Sawada, S. Araki, R. Mukai, S. Makino. Blind Extraction of a Dominant Source Signal from Mixtures of Many Sources 61-64. [Crossref]
823. W.Y. Leong, J. Homer. Blind Multiuser Receiver in Rayleigh Fading Channel 155-161. [Crossref]
824. G.W. Mabey, J. Gunther, T. Bose. A Euclidean direction based algorithm for blind source separation using a natural gradient V-561-4. [Crossref]
825. M.S. Pedersen, U. Kjems, J. Larsen. On the difference between updating the mixing matrix and updating the separation matrix 297-300. [Crossref]
826. Seungjin Choi, A. Cichocki, S. Amari. Natural gradient learning for second-order nonstationary source separation 654-658. [Crossref]
827. Qi Lv, Xian-Da Zhang, Ying Jia. Kalman Filtering Algorithm for Blind Source Separation 257-260. [Crossref]
828. E. Principi, S. Squartini, F. Piazza. An ICA based approach for Blind Deconvolution of Three-dimensional Signals 5714-5717. [Crossref]
829. T. Shimotomai, T. Omori. A model of word meaning inference development in child 1236-1240. [Crossref]
830. N. Kato, H. Kashimura, H. Ikeda, M. Shimizu. An analysis-synthesis loop model using kernel method 253-262. [Crossref]
831. Fang Liu, Fan Jia. Immunity clone strategy based ICA 296-301. [Crossref]
832. K. Matsuoka, M. Ohata, T. Tokunari. A kurtosis-based blind separation of sources using the Cayley transform 369-374. [Crossref]
833. U. Manmontri, P.A. Naylor. Blind Signal Separation Using a Criterion Based on Principle of Minimal Disturbance V-829-V-832. [Crossref]
834. E. Uchino, T. Azetsu, M. Murata. Independent component analysis by using radial basis function network 494-497. [Crossref]
835. Seungjin Choi, A. Cichocki, Liqing Zhang, S. Amari. Approximate maximum likelihood source separation using the natural gradient 235-238. [Crossref]
836. R. Aichner, S. Araki, S. Makino, T. Nishikawa, H. Saruwatari. Time domain blind source separation of non-stationary convolved signals by utilizing geometric beamforming 445-454. [Crossref]
837. S. Kaski, J. Sinkkonen, J. Peltonen. Learning metrics for self-organizing maps 914-919. [Crossref]
838. S. Hochreiter, K. Obermayer. Optimal gradient-based learning using importance weights 114-119. [Crossref]
839. J. Murillo-Fuentes, F. Gonzalez-Serrano, J. Ramos, A.J. Caamano. Natural gradient based blind multiuser detection 1186-1189. [Crossref]

840. Li Xiaojun, Zhang Xianda, Bao Zheng. Estimating DOA using independent component analysis 1389-1391. [Crossref]
841. Xiao-fei Shi, Ren-jie Liu, Xiao-ming Liu, Li Li. A Blind Source Separation Algorithm Based on a Unifying Model 696-699. [Crossref]
842. Fasong Wang, Hongwez Li, Rui Li, Yuantong Shen. Novel algorithm for independent component analysis with flexible core functions 132-135. [Crossref]
843. Y. Hayashi. Enhancement of a gradient descent or ascent with a function weight for the minimum α-divergence 98-103. [Crossref]
844. J.-F. Cardoso. Learning in manifolds: the case of source separation 136-139. [Crossref]
845. J. Karvanen, J. Eriksson, V. Koivunen. Maximum likelihood estimation of ICA model for wide class of source distributions 445-454. [Crossref]
846. Xizhong Shen, Dachao Hu, Guang Meng, Tao Wang. SOS-based Blind Equalization Algorithm of FIR MIMO System in Frequency Domain 263-267. [Crossref]
847. Peixun Luo, K.Y.M. Wong. Online learning with recycled examples: a cavity analysis 1341-1345. [Crossref]
848. S.C. Douglas, S. Amari, S.-Y. Kung. Gradient adaptation under unit-norm constraints 144-147. [Crossref]
849. M. Elsabrouty, M. Bouchard, T. Aboulnasr. A new On-Line Negentropy-based Algorithm for Blind Source Separation 756-759. [Crossref]
850. R. Manjunath, K.S. Gurumurth. Information geometry of differentially fed artificial neural networks 1521-1525. [Crossref]
851. Yun-Hui Liu, Si-Wei Luo, Ai-Jun Li, Hua Huang, Jin-Wei Wen. Information geometry on extendable hierarchical large scale neural network model 1380-1384. [Crossref]
852. H. Yan, H. Chen, Y. Xia, Y. Lai, D. Zhou. Independent component analysis for human epileptic spikes extraction 93-95. [Crossref]
853. H. Szu, I. Kopriva. Comparison of Lagrange constrained neural network with traditional ICA methods 466-471. [Crossref]
854. W.Y. Leong, J. Homer. Implementing ICA in blind multiuser detection 947-952. [Crossref]
855. S. Choi, A. Cichocki. A hybrid learning approach to blind deconvolution of MIMO systems 292-295. [Crossref]
856. J. Joseph, K.V.S. Hari. Adaptive estimation of parameters using partial information of desired outputs 1204-1208. [Crossref]
857. Xiao-fei Shi, Ren-jie Liu, Yao-liang Huang. A kurtosis-dependent parameterized blind source separation algorithm and stability analysis 487-490. [Crossref]
858. W.L. Woo, S. Sali. Neural network schemes for blind separation of sources from nonlinear mixtures 1227-1234. [Crossref]
859. S. Kaski. Learning metrics for exploratory data analysis 53-62. [Crossref]
860. H. Imamura, H. Ochi. A high capacity MIMO system based on blind deconvolution 2042-2046. [Crossref]
861. Bin-Chul Ihm, Dong-Jo Park. Acceleration of learning speed in neural networks by reducing weight oscillations 1729-1732. [Crossref]
862. A. Mansour, M. Kawamoto, N. Ohnishi. Blind separation for instantaneous mixture of speech signals: algorithms and performances 26-32. [Crossref]
863. K. Waheed, F.M. Salem. Blind source recovery for non-minimum phase surroundings V-33-V-36. [Crossref]
864. Tiemin Mei, Fuliang Yin, Jiangtao Xi, A. Mertins, J.F. Chicharo. A half-frequency domain approach for convolutive blind source separation based on Kullback-Leibler divergence 25-28. [Crossref]
865. Seungjin Choi. Differential learning and random walk model II-721-4. [Crossref]
866. K. Waheed, F.M. Salarn. State space blind source recovery of non-minimum phase environments II-422-II-425. [Crossref]
867. Yuan Liu, W. Mikhael. Blind source separation for frequency dependent channels 725-728. [Crossref]
868. W.L. Woo, S. Sali. A new demixer scheme for blind source separation using general neural network model 379-381. [Crossref]
869. M.G. Miller, N.E. Matsakis, P.A. Viola. Learning from one example through shared densities on transforms 464-471. [Crossref]
870. R. Horie, E. Aiyoshi. Variable metric gradient projection method and replicator equation 515-520. [Crossref]
871. S. Squartini, F. Piazza, A. Shawker. New Riemannian metrics for improvement of convergence speed in ICA based learning algorithms 3603-3606. [Crossref]
872. H. Abdulkader, D. Roviras, F. Castanie. Neural network applications to satellite nonlinear channel processing 145-146. [Crossref]
873. M. Toussaint. On model selection and the disability of neural networks to decompose tasks 245-250. [Crossref]
874. L.-Q. Zhang, A. Cichocki, S. Amari. Kalman filter and state-space approach to blind deconvolution 425-434. [Crossref]

875. M.R. Bastian, J.H. Gunther, T.K. Moon. An Improvement to the Natural Gradient Learning Algorithm for Multilayer Perceptrons 313-316. [Crossref]
876. S.C. Douglas, H. Sawada, S. Makino. Natural gradient blind deconvolution and equalization using causal FIR filters 197-201. [Crossref]
877. S. Bermejo, G. Bedoya, V. Parisi, J. Cabestany. An on-line water monitoring system using a smart ISFET array 2797-2802. [Crossref]
878. Seungjin Choi, Andrzej Cichocki, Shunichi Amari. Flexible independent component analysis 83-92. [Crossref]
879. H. Sawada, R. Mukai, S. Araki, S. Makino. A robust approach to the permutation problem of frequency-domain blind source separation V-381-4. [Crossref]
880. S. Amari, T. Ozeki, H. Park. Information geometry of adaptive systems 12-17. [Crossref]
881. Wei Lu, J.C. Rajapakse. Unique ICA solution by eliminating indeterminacy 388-393. [Crossref]
882. R. Manjunath, K.S. Gurumurthy. Artificial neural networks as building blocks of mixed signal FPGA 375-378. [Crossref]
883. D. Erdogmus, O. Fontenla-Romero, J.C. Principe, A. Alonso-Betanzos, E. Castillo, R. Jenssen. Accurate initialization of neural network weights by backpropagation of the desired response 2005-2010. [Crossref]
884. Jianfeng Chen, Xiaolong Zhu, Xianda Zhang. A multiuser detector based on blind signal separation 1809-1812. [Crossref]
885. J. Miguez, L. Castedo. Maximum likelihood blind multiuser detection 309-313. [Crossref]
886. Te-Won Lee, A. Ziehe, R. Orglmeister, T. Sejnowski. Combining time-delayed decorrelation and ICA: towards solving the cocktail party problem 1249-1252. [Crossref]
887. F. O'Regan, C. Heneghan. Algorithmic analysis and implementation of a novel natural gradient adaptive filter for sparse systems II-553-6. [Crossref]
888. H.R. Eghbalnia, A. Assadi, A. Bahrami. A probabilistic model of coherent spatiotemporal dynamics in neuronal data 381-386. [Crossref]
889. S.C. Douglas, S.-Y. Kung. Kuicnet Algorithms for Blind Deconvolution 7-12. [Crossref]
890. Y. Luo, J.A. Chambers. A modified underdetermined blind source separation algorithm using competitive learning 966-969. [Crossref]
891. L.-Q. Zhang, A. Cichocki, S. Amari. Geometrical structures of FIR manifold and their application to multichannel blind deconvolution 303-312. [Crossref]
892. Hyeyoung Park, Hyunjin Lee. Reconsideration to pruning and regularization for complexity optimization in neural networks 1649-1653. [Crossref]
893. M. Ohata, T. Tokunari, K. Matsuoka. An online algorithm for blind source separation with Gaussian mixture model 375-378. [Crossref]
894. N. Duffy. Using multiplicative algorithms to build cascade correlation networks 1861-1865. [Crossref]
895. E.G. Miller, C. Chefd'hotel. Practical non-parametric density estimation on a transformation group for vision II-114-21. [Crossref]
896. R.E. Mahoney, R.C. Williamson. Riemannian structure of some new gradient descent learning algorithms 197-202. [Crossref]
897. Meng Jicheng, Yang Wanlin. Nearest neighbor classifier based on riemannian metric in radar target recognition 851-853. [Crossref]
898. A. Celani, S. Bartoloni, A. Uncini, F. Piazza. A multirate approach to multichannel blind deconvolution I-677-I-680. [Crossref]
899. J.J. Murillo-Fuentes, M. Sanchez-Fernandez, F.J. Gonzalez-Serrano. Adaptive blind multiuser detection in asynchronous CDMA 1844-1848. [Crossref]
900. S.C. Douglas. Equivariant algorithms for selective transmission 1133-1136. [Crossref]
901. A. Taleb. Source separation in structured nonlinear models 3513-3516. [Crossref]
902. A. Cichocki, B. Orsier, A. Back, S.-I. Amari. On-line adaptive algorithms in non-stationary environments using a modified conjugate gradient approach 316-325. [Crossref]
903. J. Miquez, L. Castedo. Maximum likelihood blind source separation in Gaussian noise 343-352. [Crossref]
904. L.-Q. Zhang, A. Cichocki, S. Amari. Multichannel blind deconvolution of non-minimum phase systems using information backpropagation 210-216. [Crossref]
905. F. Castanie, D. Roviras. Neural networks in space communications 3-7. [Crossref]
906. M. Tufail, M. Abe, M. Kawamata. Blind Separation of Statistically Independent Signals with Mixed Sub-Gaussian and Super-Gaussian Probability Distributions 3027-3030. [Crossref]

907. L. Zhang, S. Amari, A. Cichocki. Estimating function approach to multichannel blind deconvolution 587-590. [Crossref]
908. L. Zhang, A. Cichocki. Blind deconvolution/equalization using state-space models 123-131. [Crossref]
909. J.J. Murillo-Fuentes, M. Sanchez-Fernandez, A. Caamano-Fernandea, F.J. Gonzalez-Serrano. Adaptive blind joint source-phase separation in digital communications 930-934. [Crossref]
910. Zhongfeng Li, Qiuhua Lin. FPGA Implementation of Infomax BSS Algorithm with Fixed-Point Number Representation 889-892. [Crossref]
911. Wen Jinwei, Luo Siwei, Huang Hua. Information geometry on ensemble HME model 679-682. [Crossref]
912. A. Cichocki, L. Zhang. Adaptive multichannel blind deconvolution using state-space models 296-299. [Crossref]
913. Seungjin Choi, O Young Lee. Nonstationary source separation 670-673. [Crossref]
914. S. Squartini, F. Piazza, F.J. Theis. New Riemannian metrics for speeding-up the convergence of over- and underdetermined ICA 4. [Crossref]
915. M. Ohata, T. Mukai, K. Matsuoka. Blind separation of convolutive mixtures of speech signals using linear combination model 33-36. [Crossref]
916. U. Manmontri, P.A. Naylor. Blind Identification Using Second-Order Statistics: a Nonstationarity and Nonwhiteness Approach 305-308. [Crossref]
917. Heeyoul Choi, Seungjin Choi. Relative Trust-Region Learning for ICA 261-264. [Crossref]
918. S.C. Douglas, A. Cichocki, S.-I. Amari. Multichannel blind separation and deconvolution of sources with arbitrary distributions 436-445. [Crossref]
919. S. Amari, S.C. Douglas, A. Cichocki, H.H. Yang. Multichannel blind deconvolution and equalization using the natural gradient 101-104. [Crossref]
920. Gonzalo Safont, Addisson Salazar, Alberto Rodriguez, Luis Vergara. An Experimental Sensitivity Analysis of Gaussian and Non-Gaussian Based Methods for Dynamic Modeling in EEG Signal Processing 4028-4041. [Crossref]
921. Manjunath Ramachandra, Pandit Pattabhirama. Analysis of the High-Speed Network Performance through a Prediction Feedback Based Model 162-178. [Crossref]