NONLINEAR INTERFERENCE CANCELLATION USING NEURAL NETWORKS

Andrzej Cichocki, Sergiy A. Vorobyov, and Tomasz Rutkowski

Lab. for Open Information Systems, Riken, Brain Science Institute
2-1 Hirosawa, Saitama 351-0198 Wako-shi, Japan
`cia(svor,tomek)@open.brain.riken.go.jp`

Abstract— In this paper neural network approach for nonlinear noise and interference cancellation is developed . The Hyper Radial Basis Function (HRBF) neural network with associated Manhattan learning algorithm is proposed for noise cancellation under assumption that reference noise is available. Extended multichannel model with multi-references noise signal is also presented and fast learning algorithm is developed. Performance and validity of the learning algorithms are illustrated by simulation.

I. INTRODUCTION

In many applications, especially in biomedical signal processing the source signals are corrupted by various interferences and noises. Optimum noise cancellation usually requires nonlinear dynamic system for efficient noise cancellation [3], [4], [7]. Traditional linear Finite Impulse Response (FIR) adaptive filter may not achieve acceptable level of noise cancellation for many real-world problems since noise signals are related to available (measured) reference signals in a complex dynamic and nonlinear way [1], [7].

In this paper we propose to use HRBF neural network first proposed by Poggio and Girosi [9].

Most of the work about linear and nonlinear noise cancellation assume that additive noise can be modeled as Gaussian process [2] - [8]. In this paper we relax this assumption and we investigate several neural network models and associated adaptive learning algorithms for optimal noise cancellation.

II. PROBLEM FORMULATION

The basic problem is illustrated in Fig.1.

We observe corrupted by additive noise source signal

$$d(k) = s(k) + \nu(k), \tag{1}$$

where $s(k)$ is the unknown primary source signal and $\nu(k)$ is the undesired interference or noise signal. We assume that only noisy signal $d(k)$ and reference noise $\nu_R(k)$ are available. Moreover, we assume that between reference noise $\nu_R(k)$ and noise $\nu(k)$ exist unknown nonlinear dynamic relationship described by
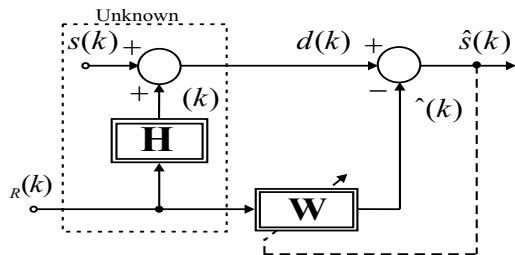

Figure 1: Interference cancelling principle with reference noise ν_R.

nonlinear filter **H**. Our objective is to design an appropriate neural network which estimates nonlinear filter **H** in such a way that we estimate optimally noise $\nu(k)$ and subtract it from signal $d(k)$ (see Fig.1). So, primary source signal can be estimated.

In many applications we have several observation of the same signal as

$$d_i(k) = g_i s(k) + \nu_i(k), \quad (i = 1, 2, ..., l), \tag{2}$$

where scaling coefficients g_i are unknown and $\nu_i(k)$ are additive noises. Our objective is to estimate original source signal $s(k)$ taking into account all available information.

III. HYPER RADIAL BASIS FUNCTION NETWORK FOR NONLINEAR INTERFERENCE CANCELLATION

A HRBF network have been introduced first by Poggio and Girosi [9]. The main idea is to consider the mapping to be approximated (for our problem, it is nonlinear filter **W**) as the sum of several radial basis functions, each one with its own prior, that is stabilizer. The corresponding regularization principle then yields a superposition of different Green's functions, in particular Gaussian functions, with different widths. We are interested here in the special case of Green's functions that allow us to consider HRBF network as a two-layer neural network in which the hidden layer perform an adaptive nonlinear transformation with adjustable weight parameters so that M dimensional input space $\overline{\nu}_{\mathbf{R}}(k) = [\nu_R(k), \nu_R(k-1), ..., \nu_R(k-M+1)]^\top$ is

Figure 2: Model of HRBF network for nonlinear interference cancellation.

Figure 3: Multichannel nonlinear interference cancellation model.

mapped to the one-dimensional output space $\hat{\nu}(k)$: $R^M \to R$ as

$$
\begin{aligned}
\hat{\nu}(k) &= w_0 + \sum_{j=1}^{n} w_j \Phi_j(\rho_j) \\
&= w_0 + \mathbf{w}^\top \mathbf{\Phi}(\rho),
\end{aligned} \tag{3}
$$

where $\mathbf{w} = [w_1, w_2, ..., w_n]^\top \in R^n$ is the vector of weights, $\mathbf{\Phi}(\rho) = [\Phi_1(\rho_1), \Phi_2(\rho_2), ..., \Phi_n(\rho_n)]^\top$ is the vector consist of generalized Gaussian functions $\Phi_j(\rho_j) = \frac{1}{2}\exp(-\frac{1}{2}\rho_j^2)$ and $\rho_j^2 = \| \mathbf{Q}_j(\overline{\nu}_\mathbf{R}(k) - \mathbf{c}_j) \|^2 = (\overline{\nu}_\mathbf{R}(k) - \mathbf{c}_j)^\top \mathbf{Q}_j^\top \mathbf{Q}_j(\overline{\nu}_\mathbf{R}(k) - \mathbf{c}_j) = \sum_{\kappa=1}^{M} \left[\sum_{\varsigma=1}^{M} q_j^{\kappa\varsigma}(\nu_R(k-\varsigma+1) - \mathbf{c}_{j\varsigma}) \right]^2$, $j = 1, ..., n$, with adaptive centers $\mathbf{c}_j = [c_{j1}, ..., c_{jM}]^\top$ and $\mathbf{Q}_j$ are covariance matrices (see also Fig.2). Note that if $(M \times M)$ positive definite matrix $\mathbf{Q}_j^\top \mathbf{Q}_j$ reduces to a diagonal matrix $\mathrm{diag}(\sigma_{j1}^{-2}, ..., \sigma_{jM}^{-2})$ the HRBF network is simplified to the RBF neural network.

For identification of nonlinear filter $\mathbf{W}$ using HRBF network we must estimate all parameters: w_i, $i = 0, 1, ..., n$, $\mathbf{c}_j$ and $\mathbf{Q}_j$.

Unfortunately, for interference cancellation problem, there is no explicit training set of input-output examples for HRBF network learning. The objective becomes minimization of the power of the output signal $\hat{s}(k)$, which can be written as

$$
\begin{aligned}
\hat{s}(k) &= e(k) = d(k) - \hat{\nu}(k) \\
&= d(k) - w_0 - \mathbf{w}^\top \mathbf{\Phi}(\rho).
\end{aligned} \tag{4}
$$

Our objective now is to estimate all free parameters $\Theta = \{\mathbf{w}, \{\mathbf{c}_j\}, \{\mathbf{Q}_j\}\}$ of HRBF network using standard multi-step or one-step power function like following function

$$
J(\Theta) = \frac{1}{2}e^2(k) = \frac{1}{2}\hat{s}^2(k). \tag{5}
$$

To avoid stacking in local minima we apply Manhattan learning formula of the form [10]:

$$
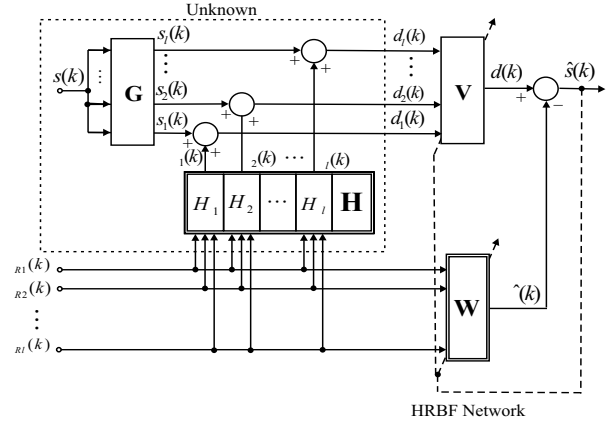\Theta(k) = \Theta(k-1) - \eta(k)\,\mathrm{sign}\frac{\partial J(\Theta)}{\partial \Theta}, \tag{6}
$$

where $\eta(k)$ is self-adaptive learning steps. In the simplest case $\eta(k)$ may be fixed.

Gradient components for all free parameters are

$$
\begin{aligned}
\frac{\partial J(\Theta)}{\partial w_0} &= -(d(k) - \hat{\nu}(k)), \\
\nabla_\mathbf{w} J(\Theta) &= -\mathbf{\Phi}(\rho)(d(k) - \hat{\nu}(k)), \\
\frac{\partial J(\Theta)}{\partial \mathbf{Q}_j} &= \delta_j \mathbf{Q}_j(\overline{\nu}_\mathbf{R}(k) - \mathbf{c}_j)(\overline{\nu}_\mathbf{R}(k) - \mathbf{c}_j)^\top, \\
\nabla_{\mathbf{c}_j} J(\Theta) &= -\delta_j \mathbf{Q}_j^\top \mathbf{Q}_j(\overline{\nu}_\mathbf{R}(k) - \mathbf{c}_j), \tag{7}
\end{aligned}
$$

where $\delta_j = \frac{1}{2}w_j \exp(-\frac{1}{2}(\overline{\nu}_\mathbf{R}(k) - \mathbf{c}_j)^\top \mathbf{Q}_j^\top \mathbf{Q}_j(\overline{\nu}_\mathbf{R}(k) - \mathbf{c}_j))(d(k) - \hat{\nu}(k)) = w_j \Phi_j(\rho_j)(d(k) - \hat{\nu}(k))$.

Note that HRBF network is extension of the RBF network and it can be proved that it will work good ever if the sequences $\nu(k)$ and $\nu_R(k)$ have non-Gaussian distribution (for example, Laplacian or Cauchy distribution).

IV. EXTENDED MULTICHANNEL MODEL WITH MULTI-REFERENCES NOISE SIGNAL

In this section we consider more sophisticated model shown in Fig.3. In this model we assume that we observe l noisy signals $d_i(k)$ (see Eq. (2)). We also observe l reference noises that are uncorrelated with a source signal $s(k)$ but are related in nonlinear way, represented symbolically by filters $\mathbf{H}_i$, to noises $\nu_i(k)$.

We use the algorithm proposed in last section for estimation of HRBF network $\mathbf{W}$ weights only with one difference: the vector $\overline{\nu}_\mathbf{R}(k)$ now is $\overline{\nu}_\mathbf{R}(k) = [\nu_{R1}(k), \nu_{R1}(k-1), ..., \nu_{R1}(k-M_1+1), \nu_{R2}(k), \nu_{R2}(k-1), ..., \nu_{R2}(k-M_2+1), ..., \nu_{Rn}(k), \nu_{Rl}(k-1), ..., \nu_{Rl}(k-M_l+1)]^\top$.

The original source signal is reconstructed by linear network

$$
\hat{s}(k) = \sum_{i=1}^{l} v_i d_i(k) - \hat{\nu}(k) = \mathbf{v}^\top \mathbf{d}(k) - \hat{\nu}(k) \tag{8}
$$

with constraint

$$\mathbf{v}^\top \mathbf{e}_l = 1, \qquad (9)$$

because we should take into account that one-dimension corrupted signal $d(k)$ must be unbiased. Here $\mathbf{v} = [v_1, v_2, ..., v_l]^\top$ is $(l \times 1)$ weight vector of linear network $\mathbf{V}$, $\mathbf{d}(k) = [d_1(k), d_2(k), ..., d_l(k)]^\top$ and $\mathbf{e}_l = [1, 1, ..., 1]^\top$ is $(l \times 1)$ vectors.

Minimization of cost function

$$J = \frac{1}{2}\sum_{k=1}^{N} \hat{s}^2(k) = \frac{1}{2}\sum_{k=1}^{N}(\mathbf{v}^\top \mathbf{d}(k) - \hat{\nu}(k))^2, \qquad (10)$$

under constraints (9) lead to the Lagrangian

$$
\begin{aligned}
L \ = \ & \frac{1}{2}\sum_{k=1}^{N}(\mathbf{v}^\top \mathbf{d}(k) - \hat{\nu}(k))^2 \\
& + \ \lambda(\mathbf{v}^\top \mathbf{e}_l - 1).
\end{aligned}
\qquad (11)
$$

Hence, Kuhn-Tucker equations can be expressed as

$$
\begin{aligned}
\nabla_{\mathbf{v}(k)} L \ = \ & \sum_{k=1}^{N} \mathbf{d}^\top(k)\mathbf{v}\mathbf{d}(k) \\
& - \sum_{k=1}^{N} \mathbf{d}(k)\hat{\nu}(k) - \lambda \mathbf{e}_l = 0, \\
\frac{\partial L}{\partial \lambda} \ = \ & \mathbf{v}^\top \mathbf{e}_l - 1 = 0,
\end{aligned}
\qquad (12)
$$

where λ is nonnegative Lagrange multiplier, N is the width of window.

Introducing into consideration new definitions for errors covariance matrix $\mathbf{P}^{-1} = \left(\sum_{k=1}^{N} \mathbf{d}(k)\mathbf{d}^\top(k)\right)^{-1}$ and least mean square estimation $\tilde{\mathbf{v}} = \left(\sum_{k=1}^{N} \mathbf{d}(k)\mathbf{d}^\top(k)\right)^{-1}\left(\sum_{k=1}^{N} \mathbf{d}(k)\hat{\nu}(k)\right)$ the solution of the system (12) can be given by:

$$
\begin{aligned}
\lambda \ = \ & 2\frac{\mathbf{e}_l^\top \tilde{\mathbf{v}} - 1}{\mathbf{e}_l^\top \mathbf{P}^{-1}\mathbf{e}_l}, \\
\mathbf{v} \ = \ & \tilde{\mathbf{v}} - \mathbf{P}^{-1}\frac{\mathbf{e}_l^\top \tilde{\mathbf{v}} - 1}{\mathbf{e}_l^\top \mathbf{P}^{-1}\mathbf{e}_l}\mathbf{e}_l.
\end{aligned}
\qquad (13)
$$

Finally using Sherman-Morrison lemma about recursive matrix inversion we obtain following learning algorithm for on-line tuning of vector $\mathbf{v}(k)$ of linear network $\mathbf{V}$:

$$
\begin{aligned}
\mathbf{P}^{-1}(k) \ = \ & \mathbf{P}^{-1}(k-1) \\
& - \frac{\mathbf{P}^{-1}(k-1)\mathbf{d}(k)\mathbf{d}^\top(k)\mathbf{P}^{-1}(k-1)}{1 + \mathbf{d}^\top(k)\mathbf{P}^{-1}(k-1)\mathbf{d}(k)}, \\
\tilde{\mathbf{v}}(k) \ = \ & \tilde{\mathbf{v}}(k-1) \\
& + \frac{\mathbf{P}^{-1}(k-1)\mathbf{d}(k)\hat{s}(k)}{1 + \mathbf{d}^\top(k)\mathbf{P}^{-1}(k-1)\mathbf{d}(k)}, \\
\mathbf{v}(k) \ = \ & \tilde{\mathbf{v}}(k) - \mathbf{P}^{-1}(k)\frac{\mathbf{e}_l^\top \tilde{\mathbf{v}}(k) - 1}{\mathbf{e}_l^\top \mathbf{P}^{-1}(k)\mathbf{e}_l}\mathbf{e}_l.
\end{aligned}
\qquad (14)
$$

It is obvious that we can achieve more precise estimate of $\hat{s}(k)$ and better convergence using multi-references noise signal network than scalar reference noise signal network.

V. SIMULATION AND COMPARISON RESULTS

In this section, we shortly present some simulation results for the adaptive nonlinear interference canceler using HRBF network and comparison results for standard RBF network $\hat{\nu}(k) = w_0 + \sum_{i=1}^{M} w_i \exp\left(-\frac{\|\bar{\nu}_\mathbf{R}(k) - \mathbf{c}_i\|^2}{\sigma_i^2}\right)$, (here σ_i are width parameters) with Stochastic Gradient (SG) learning algorithm [3] and HRBF network, proposed in this paper. Performances of the nonlinear interference canceller based on HRBF and RBF networks were measured by the normalized mean square error

$$
\begin{aligned}
NMSE &= \frac{E\{(\hat{s}(k) - s(k))^2\}}{E\{\nu^2(k)\}} \\
&= \frac{\sum_{k=1}^{K}(\hat{s}(k) - d(k))^2}{\sum_{k=1}^{K}\nu^2(k)},
\end{aligned}
\qquad (15)
$$

where K is the number of iterations during learning period.

The initial weights w_j, $(j = 0, 1, 2, ..., n)$ of the HRBF and RBF networks were all initialized to small random values. The initial values of centers for RBF network were determined by K-means clustering on the first 100 samples of the input data and initial width parameters σ_i were set to the average of the M nearest-neighbor distances among the initialized centers. For HRBF network initial values of centers and components of matrices $\mathbf{Q}_j$, $j = 1, 2, ..., M$ were chosen to small random values.

If we would choose initial condition for HRBF network as follow: for centers as K-mean clusters that had obtained using 100 samples of the input data and for $\mathbf{Q}_j$: $q_{ii}^j = \sigma_j$, $q_{ik}^j = 0$ for all $i \neq k$, then RBF and HRBF would be quite equivalent before learning. So RBF network had a little bit more "convenient" initial conditions in our simulation than HRBF network.

Due to limit of space we present here only one illustrative example. The signal $s(k)$ and the reference noise were given by $s(k) = 0.7\cos(2\pi 0.1k) + 0.3\sin(2\pi 0.25k)$, $\nu_R(k) = \sin(2\pi 0.06k + \chi(k))$, where $\chi(k)$ is the random phase generated according Laplace pdf with $\sigma_\chi = 0.003$. Interference noise signal was $\nu(k) = \cos(2\pi 0.06k) - 0.5\sin^2(2\pi 0.06(k-1)) + (\cos(2\pi 0.06k) - 0.5\sin^2(2\pi 0.06(k-1)))^2$.

The interference canceller has to estimate the deterministic interference $\nu(k)$ using samples of the reference input $\nu_R(k)$ with a random non-Gaussian phase noise. The relationship between $\nu_R(k)$ and $\nu(k)$ is highly nonlinear and also the phase of reference signal
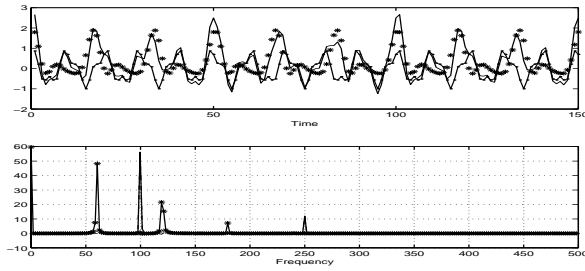
Figure 4: Waveforms and magnitude spectra for signals $d(k)$ -, $s(k)$ - . - and $\nu(k)$ * .
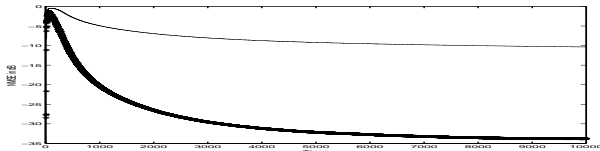

Figure 5: Normalized mean square error for RBF - and HRBF * networks.

randomly fluctuate. Fig. 4 shows the time waveforms and spectra of the corrupted signal $d(k) = s(k) + \nu(k)$, the useful and interference signals $s(k)$ and $\nu(k)$.

Due to size limitation we show the simulation results for only one configuration of networks, with $M = 5$ delays and $n = 20$ hidden units both for RBF and HRBF networks. Figs.5, 6, 7 show respectively: Fig.5 - normalized mean square error for RBF and HRBF networks; Fig.6 - the output spectra of RBF and HRBF networks spectra for $K = 10000$; and Fig.7 - the original reference signal $\nu(k)$ and its estimation $\hat{\nu}(k)$ using RBF and HRBF networks on the last 200 learning iterations.

We can see that HRBF network is able to learn more rapidly and allow to achieve smaller value of normalized mean square error than the standard RBF network. These results are similar for other configuration of RBF and HRBF networks and are general. Hence, these results suggest that the nonlinear canceller based on the HRBF network performed much better than the canceller based on the standard RBF network. However, it should be noted that HRBF network is sensitive to initial values of self-adaptive learning steps for learning algorithm. Hence, generally speaking, the procedure for adjustment of learning steps can be added to the algorithm.
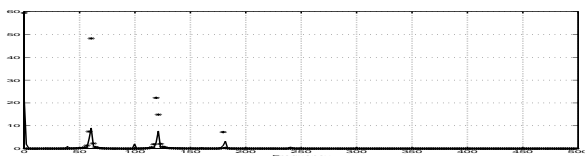

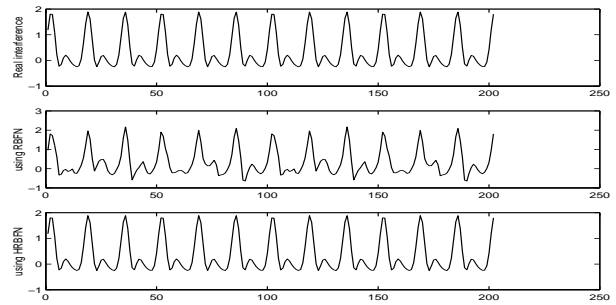Figure 6: Output spectra of RBF - and HRBF * networks.

Figure 7: Original reference signal $\nu(k)$ and its estimations $\hat{\nu}(k)$ using RBF and HRBF networks.

References

[1] A. Cichocki, W. Kasprzak, S. Amari, "Adaptive approach to blind source cancellation of additive and convolutional noise", *ICSP'96, Int. Conf. on Signal Processing*, Beijing, China, pp.412-415, 1996.

[2] B. Widrow, S.D. Stearn, *"Adaptive Signal Processing"*, Prentice-Hall, S.P.Series, 1985.

[3] I. Cha and S.A. Kassam, "Interference cancellation using radial basis function networks", *Signal Processing*, vol.47, pp.247-268, 1995.

[4] I. Cha and S.A. Kassam, "Channel equalization using adaptive complex radial basis function networks", *IEEE Journal on Selected Areas in Communications*, vol.13, No.1, pp.122-131, 1995.

[5] Y. Lu, N. Sundararajan and P. Saratchandran, "Performance evaluation of a sequential minimal radial basis function (RBF) neural network learning algorithm", *IEEE Trans. Neural Networks*, vol.9, No.2, pp.308-318, 1998.

[6] E. Bataillou and H. Rix, "A new method for noise cancellation", *Proc. of Conf. on Signal Processing VII: Theories and Applications*, pp.1046-1049, 1994.

[7] W.G. Knecht, M.E. Schenkel and G.S. Moschytz "Nonlinear filters for noise reduction", *Proc. of Conf. on Signal Processing VII: Theories and Applications*, pp.1500-1503, 1994.

[8] C. Serviere, D. Baudois, A. Silvent, "Noise cancelling with signal and reference correlated, using third order moments", (in book *Higher Order Statistics*), 1992, pp.279-282.

[9] T. Poggio and F. Girosi, "Network for approximation and learning",*Proc. of the IEEE*, vol.78, pp.1481-1497, 1990.

[10] A. Cichocki, and R. Unbehauen, *"Neural Networks for Optimization and Signal Processing"*, J. Wiley, 1994.