AN ONLINE QUASI-NEWTON ALGORITHM FOR BLIND SIMO IDENTIFICATION

Emanuël A.P. Habets and Patrick A. Naylor

Department of Electrical and Electronic Engineering, Imperial College London, UK
Email: {e.habets, p.naylor}@imperial.ac.uk

ABSTRACT

In the last decade various time- and frequency-domain algorithms were derived to blindly identify acoustic systems. One of these algorithms is the multichannel Newton (MCN) algorithm, which is also the basis of the well known normalized multichannel frequency-domain least-mean-square (NMCFLMS) algorithm. A major drawback of the MCN is that it requires the computation and inversion of a Hessian matrix, which involves extensive computation making it unsuitable for online applications. In this paper, we therefore derive and investigate an efficient online multichannel quasi-Newton (MCQN) algorithm that updates the inverse of the Hessian by analyzing successive gradient vectors. The new MCQN is shown to exhibit similar performance to MCN but with much reduced complexity.

***Index Terms*—** blind system identification, Newton method, quasi-Newton method, multichannel signal processing.

1. INTRODUCTION

The idea of blind system identification (BSI) was first introduced to the communications community by Sato with the intention of designing efficient communication systems that did not require a training phase [1]. BSI can find a variety of speech applications such as speech dereverberation [2] and time delay estimation for sound source localization [3]. In these applications the source signal is unobservable and the impulse responses of the acoustic channels between the source and sensors are blindly estimated.

Within the acoustic signal processing community various batch [4] and adaptive [2, 5, 6] algorithms for BSI have been developed that are based on minimizing the cross-relation error between sensor pairs. These algorithms can accurately determine the impulse responses of a multichannel system using a finite number of samples when additive noise in the system outputs is weak and certain identifiability conditions are met. Adaptive algorithms are able to track the dynamic nature of acoustic impulse responses and are therefore far more suitable for real-time applications than batch algorithms. In [2], Huang et al. derived a multichannel least-mean-square

(MCLMS) algorithm and a multichannel Newton (MCN) algorithm. Their results clearly indicate the that performance of the MCN is superior to the MCLMS. Because the MCN algorithm requires the computation and inversion of a Hessian matrix that contains the second derivatives of the cost function, the computational complexity of the MCN algorithm is significantly larger than that of the MCLMS algorithm. The MCN algorithm was also used as a basis of the normalized multichannel frequency-domain least-mean-square (NMCFLMS) algorithm that was proposed in [5] by the same authors. In the frequency domain the required correlations and convolutions can be computed more efficiently than in the time domain, thereby reducing the computational complexity. In addition, the derivation of the NMCFLMS requires an approximation of the nondiagonal Hessian matrix. The NMCFLMS algorithm is one of the most computationally efficient algorithms but requires several seconds of data to converge.

In this paper we derive and investigate a new multichannel quasi-Newton (MCQN) algorithm for BSI in the time domain. In Quasi-Newton methods the Hessian matrix of second derivatives of the cost function to be minimized do not need to be computed at any stage, and the Hessian is updated by analyzing successive gradient vectors instead. Here we employ one of the most popular quasi-Newton methods known as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [7], by which we reduce the computation complexity compared to the MCN algorithm.

This paper is organized as follows. In Section 2 we formulate the BSI problem. In Section 3, we briefly review the MCN algorithm. We then derive the MCQN algorithm in Section 4. Finally, we compare the MCN and MCQN in Section 5.

2. PROBLEM FORMULATION

For an M-channel single-input-multiple-output system the mth impulse response with L coefficients can be denoted as

$$\mathbf{h}_m = [h_{m,0} \ h_{m,1} \ \ldots \ h_{m,L-1}]^T, \qquad (1)$$

for $m = 1, 2, \ldots, M$, and the mth sensor signal can be expressed as

$$x_m(n) = \sum_{j=0}^{L-1} h_{m,j} \, s(n-j) + b_m(n), \qquad (2)$$

where $s(n)$ is the source signal and $b_m(n)$ is the additive noise. The additive noise is assumed to be zero-mean and uncorrelated with the source signal. In vector form, (2) can be written

$$\mathbf{x}_m(n) = \mathbf{H}_m \mathbf{s}(n) + \mathbf{b}_m(n), \qquad (3)$$

where $\mathbf{s}(n) = [s(n)\ s(n-1)\ \dots\ s(n-2L+2)]^T$, $\mathbf{x}_m(n) = [x_m(n)\ x_m(n-1)\ \dots\ x_m(n-L+1)]^T$, $\mathbf{b}_m(n) = [b_m(n)\ b_m(n-1)\ \dots\ b_m(n-L+1)]^T$, and $\mathbf{H}_m$ is the $L \times (2L-1)$ convolution matrix for the mth channel such that

$$\mathbf{H}_m = \begin{bmatrix} h_{m,0} & \cdots & h_{m,L-1} & \cdots & \cdots & 0 \\ 0 & h_{m,0} & \cdots & h_{m,L-1} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & h_{m,0} & \cdots & h_{m,L-1} \end{bmatrix} \qquad (4)$$

The problem of BSI is to find $\mathbf{h} = [\mathbf{h}_1^T\ \mathbf{h}_2^T\ \dots\ \mathbf{h}_M^T]^T$ using only $\mathbf{x}(n) = [\mathbf{x}_1^T(n)\ \mathbf{x}_2^T(n)\ \dots\ \mathbf{x}_M^T(n)]^T$ up to a nonzero scale factor across all channels. This scale factor is irrelevant in most of acoustic signal processing applications.

According to [8], two conditions are necessary and sufficient for blind identifiability of a single-input-multiple-output (SIMO) system using BSI algorithms:

1. The channel transfer functions do not share any common zeros, i.e., the polynomials formed by $\mathbf{h}_m$ $(1 \leq m \leq M)$ are co-prime;

2. The autocorrelation matrix of the input signal $\mathbf{R}_{ss} = \mathcal{E}\{\mathbf{s}(n)\mathbf{s}^T(n)\}$ is of full rank, where $\mathcal{E}\{\cdot\}$ denotes the mathematical expectation, such that SIMO system can be fully excited.

Minimization of the cross-relation [8] based on second order statistics of the observed signals has served as the basis of many BSI algorithms. In the absence of noise,

$$\mathbf{x}_m^T(n)\mathbf{h}_l = \mathbf{x}_l^T(n)\mathbf{h}_m, \quad m, l = 1, 2, \dots, M,\ m \neq l, \quad (5)$$

where $\mathbf{h}_l$ and $\mathbf{h}_m$ denote the lth and mth acoustic impulse responses, respectively. The *a priori* error at discrete time n, which is associated with the lth and mth acoustic impulse responses, is given by

$$e_{ml}(n) = \mathbf{x}_m^T(n)\widehat{\mathbf{h}}_l(n-1) - \mathbf{x}_l^T(n)\widehat{\mathbf{h}}_m(n-1). \quad (6)$$

Subsequently, we can define the total *a priori* squared-error as

$$\xi(n) = \sum_{m=1}^{M-1} \sum_{l=m+1}^{M} e_{ml}^2(n), \qquad (7)$$

where we excluded the cases where $e_{mm} = 0$ for $m \in \{1, 2, \dots, M\}$ and count the pairs $e_{ml} = -e_{lm}$ only once. The cost function to be minimized is defined as

$$J(n) = \frac{\xi(n)}{\|\widehat{\mathbf{h}}(n-1)\|_2^2}. \qquad (8)$$

3. REVIEW MULTICHANNEL NEWTON

In general, there is a tradeoff between the excess mean-square-error, the rate of convergence, and the ability of an adaptive algorithm to track changes in the unknown system. A good balance of these competing design objectives results from the unit-norm-constrained multichannel Newton algorithm [2], in which estimated impulse responses are updated by:

$$\widehat{\mathbf{h}}(n) = \frac{\widehat{\mathbf{h}}(n-1) + \rho\,\mathbf{p}(n)}{\|\widehat{\mathbf{h}}(n-1) + \rho\,\mathbf{p}(n)\|_2}, \qquad (9)$$

where ρ $(0 < \rho < 1)$ denotes the step-size that is commonly chosen close to 1, and

$$\mathbf{p}(n) = -\mathcal{E}\left\{\nabla^2 J(n)\right\}^{-1} \nabla J(n), \qquad (10)$$

where $\nabla^2 J(n)$ and $\nabla J(n)$ are respectively the Hessian matrix and the gradient vector of $J(n)$ with respect to $\widehat{\mathbf{h}}(n-1)$. Existence of $\mathbf{p}(n)$ requires that the Hessian can be approximated by a quadratic approximation of the Taylor series and is nonsingular. With the unit-norm constraint $\|\widehat{\mathbf{h}}(n-1)\|_2 = 1$, the gradient vector and an approximation of the Hessian matrix are given by [2]

$$\nabla J(n) = \widetilde{\mathbf{R}}(n)\widehat{\mathbf{h}}(n-1) - \xi(n)\widehat{\mathbf{h}}(n-1) \qquad (11)$$

and

$$\mathbf{V}(n) \triangleq \mathcal{E}\{\nabla^2 J(n)\} \qquad (12)$$
$$\approx 2\widehat{\mathbf{R}}(n) - 4\widehat{\mathbf{h}}(n-1)\widehat{\mathbf{h}}^T(n-1)\widehat{\mathbf{R}}(n) - 4\widehat{\mathbf{R}}(n)\widehat{\mathbf{h}}(n-1)\widehat{\mathbf{h}}^T(n-1), \qquad (13)$$

where $\widetilde{\mathbf{R}}(n)$ is an instantaneous estimate of

$$\mathbf{R} = \begin{bmatrix} \sum_{m\neq 1} \mathbf{R}_{x_m x_m} & -\mathbf{R}_{x_2 x_1} & \cdots & -\mathbf{R}_{x_M x_1} \\ -\mathbf{R}_{x_1 x_2} & \sum_{m\neq 2} \mathbf{R}_{x_m x_m} & \cdots & -\mathbf{R}_{x_M x_2} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{R}_{x_1 x_M} & -\mathbf{R}_{x_2 x_M} & \cdots & \sum_{m\neq M} \mathbf{R}_{x_m x_m} \end{bmatrix} \qquad (14)$$

where $\mathbf{R} = \mathcal{E}\{\mathbf{x}(n)\mathbf{x}^T(n)\}$, $\mathbf{R}_{x_m x_l} = \mathcal{E}\{\mathbf{x}_m(n)\mathbf{x}_l^T(n)\}$, and $\widehat{\mathbf{R}}(n)$ is given by

$$\widehat{\mathbf{R}}(n) = \eta\,\widehat{\mathbf{R}}(n-1) + \widetilde{\mathbf{R}}(n), \qquad (15)$$

where η $(0 < \eta < 1)$ is an exponential forgetting factor.

In practice the Hessian matrix might contain very small eigenvalues. In order to ensure that the inverse of the Hessian exists the regularization term $\lambda\mathbf{I}$ $(0 \leq \lambda \ll 1)$ can be added to (12), where $\mathbf{I}$ denotes an identity matrix of size $ML \times ML$.

One of the main drawbacks of the MCN algorithm is the computational complexity required to compute and invert the Hessian matrix.

4. ONLINE MULTICHANNEL QUASI-NEWTON

Quasi-Newton methods were developed to reduce the computational complexity of the Newton method. In quasi-Newton algorithms the Hessian is updated by analyzing successive gradient vectors. Here we propose a quasi-Newton algorithm for BSI that is based on the BFGS optimization method [7].

The BFGS algorithm requires the difference vectors

$$\Delta \widehat{\mathbf{h}}(n) = \widehat{\mathbf{h}}(n) - \widehat{\mathbf{h}}(n-1) \tag{16}$$

and

$$\mathbf{g}_{\text{BFGS}}(n) = \nabla J(n+1) - \nabla J(n). \tag{17}$$

It is important to note that the cost function J of the on-line algorithm is time-variant, while it is time-invariant in the original BFGS method that is developed for batch processing. Therefore, we replace $\nabla J(n+1)$ by the gradient of the cost function $\tilde{J}(n) = \tilde{\xi}(n)/\|\mathbf{h}(n)\|_2^2$ that is based on the total *a posteriori* (rather than the *a priori*) squared-error given by

$$\tilde{\xi}(n) = \sum_{m=1}^{M-1} \sum_{l=m+1}^{M} \left[\mathbf{x}_m^T(n)\widehat{\mathbf{h}}_l(n) - \mathbf{x}_l^T(n)\widehat{\mathbf{h}}_m(n) \right]^2. \tag{18}$$

By replacing $\nabla J(n+1)$ in (17) by $\nabla \tilde{J}(n)$ and by using (11) and (16) we obtain

$$\mathbf{g}(n) = \widehat{\mathbf{R}}(n)\Delta\widehat{\mathbf{h}}(n) + \tilde{\xi}(n)\widehat{\mathbf{h}}(n) - \xi(n)\widehat{\mathbf{h}}(n-1). \tag{19}$$

To avoid gradient noise entering into the recursion, we used the expected value rather than the instantaneous value of the cross-correlation matrix $\mathbf{R}$ in (19).

As shown in [7], the Hessian can be updated recursively using (20). Rather than updating the Hessian it is computationally more efficient to update the inverse of the Hessian directly. The update for the inverse of the Hessian can be found by applying Woodbury matrix identity [9] to (20), which yields (21), where $\nu(n) = \Delta\widehat{\mathbf{h}}^T(n)\mathbf{g}(n)$.

We have found, through our experiments, that adding a regularization term $\lambda\Delta\widehat{\mathbf{h}}(n)$ $(0 \leq \lambda \ll 1)$ to $\mathbf{g}(n)$ increases the robustness of the MCQN algorithm in a similar way the term $\lambda\mathbf{I}$ increases the robustness of the MCN algorithm.

5. PERFORMANCE EVALUATION

In this Section, we evaluate the performance of the developed MCQN algorithm by simulations. Similar to other studies [2,

5,6], we used the normalized projection misalignment (NPM) in dB as the performance measure, which is given by

$$\text{NPM}(n) = 20\log_{10}\left(\left\| \widehat{\mathbf{h}}(n) - \frac{\widehat{\mathbf{h}}^T(n)\mathbf{h}}{\mathbf{h}^T\mathbf{h}} \right\|_2 \Big/ \left\| \widehat{\mathbf{h}}(n) \right\|_2 \right). \tag{22}$$

For the first experiment we used three sensors ($M = 3$), the source signal is an uncorrelated binary phase-shift-keying (BPSK) sequence of 8000 samples, and the additive noise is i.i.d. zero-mean Gaussian. We have used two types of channel both of length $L = 64$: i) with random coefficients drawn from a Gaussian distribution with zero-mean and unit variance, and ii) with random coefficients drawn from the same distribution but weighted by $\exp(-\alpha n)$ ($\alpha \geq 0$), such that the envelope of each channel approximates that found in an acoustic channel. We determined the average NPM over the last 50 iterations and over 50 Monte-Carlo trials for various signal-to-noise ratios (SNRs) that are given by
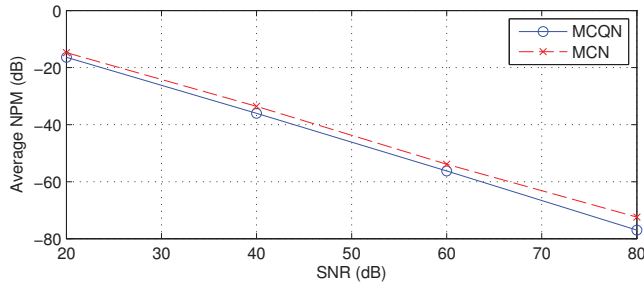
$$\text{SNR} = 10\log_{10}\left(\frac{\sigma_s^2\|\mathbf{h}\|_2^2}{M\sigma_b^2} \right), \tag{23}$$

where σ_s^2 and σ_b^2 denote the variance of the source signal $s(n)$ and additive noise $b_m(n)$ that is assumed to be channel independent. The step-size was $\rho = 0.8$, the decay-rate $\alpha = 0.05$, $\eta = 0.98$ and $\lambda = 10^{-5}$. As shown in Fig. 1, the proposed MCQN algorithm performs similar or slightly better than the MCN algorithm for all SNRs, which might be attributed to the fact that the MCN algorithm uses an approximation of the Hessian matrix [given by (12)] while the MCQN recursively estimates the exact Hessian matrix. Although not shown here, we have noticed that the initial convergence of the MCQN algorithm is slightly slower (i.e., in the order of $50 - 200$ ms) compared to the MCN algorithm.
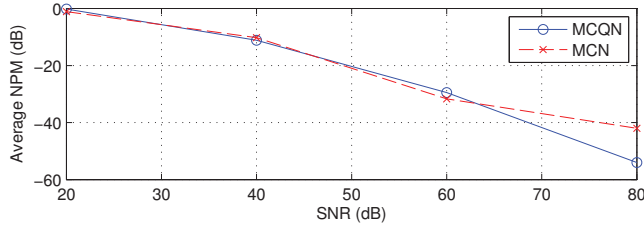
For the second experiment we used six sensors ($M = 6$), and a speech source signal of 10 seconds (sampling frequency is 8 kHz). The impulse responses are generated using the well known source-image method and are truncated to a length of $L = 256$. In Fig. 2 we show part of the source signal, the impulse responses, and the NPM as a function of time for the MCQN and NMCFLMS algorithm [5]. With an SNR of 60 dB, $\rho_{\text{MNQN}} = 0.8$, $\rho_{\text{NMCFLMS}} = 0.5$, $\eta = 0.95$, and $\lambda = 10^{-5}$ the MCQN algorithm was able to converge to approximately -40 dB within 500 milliseconds of speech data. The NMCFLMS converges to -15 dB in 10 seconds. Similar to the findings in [5], the MCN algorithm did not converge.

$$\mathbf{V}(n+1) = \mathbf{V}(n) + \frac{\mathbf{g}(n)\mathbf{g}^T(n)}{\mathbf{g}^T(n)\Delta\widehat{\mathbf{h}}(n)} - \frac{\mathbf{V}(n)\,\Delta\widehat{\mathbf{h}}(n)\Delta\widehat{\mathbf{h}}^T(n)\mathbf{V}^T(n)}{\Delta\widehat{\mathbf{h}}^T(n)\mathbf{V}(n)\Delta\widehat{\mathbf{h}}(n)} \tag{20}$$

$$\mathbf{V}^{-1}(n+1) = \left(I - \frac{\mathbf{g}(n)\Delta\widehat{\mathbf{h}}^T(n)}{\nu(n)} \right)^T \mathbf{V}^{-1}(n) \left(I - \frac{\mathbf{g}(n)\Delta\widehat{\mathbf{h}}^T(n)}{\nu(n)} \right) + \frac{\Delta\widehat{\mathbf{h}}(n)\Delta\widehat{\mathbf{h}}^T(n)}{\nu(n)} \tag{21}$$

(a) Random channels.

(b) Random channels with an exponentially decaying envelope.

Fig. 1. Average NPM for different SNRs and two types of channels after processing 8000 samples.

6. CONCLUSIONS

In this paper, an online quasi-Newton algorithm for blind system identification was developed. Because the developed MCQN algorithm computes the inverse of the Hessian recursively its computational complexity is lower than that of the MCN algorithm. The performance of the MCQN algorithm is comparable to that of the MCN algorithm but offers online processing capability.

7. REFERENCES

[1] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation," *IEEE Trans. Commun.*, vol. 6, pp. 679–682, June 1975.

[2] Y. Huang and J. Benesty, "Adaptive multi-channel least mean square and newton algorithms for blind channel identification," *Signal Processing*, vol. 82, pp. 1127–1138, Aug. 2002.

[3] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," *Journal Acoust. Soc. of America*, vol. 107, no. 1, pp. 384–391, 2000.

[4] S. Gannot and M. Moonen, "Subspace methods for multimicrophone speech dereverberation," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 11, pp. 1074–1090, 2003.

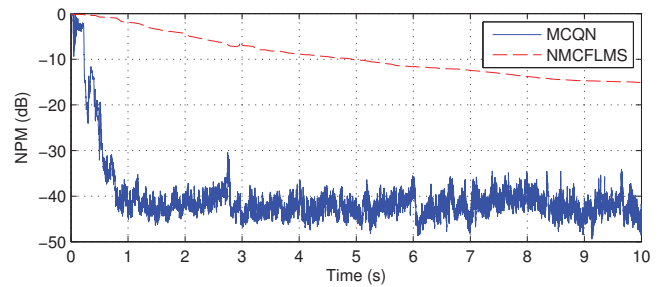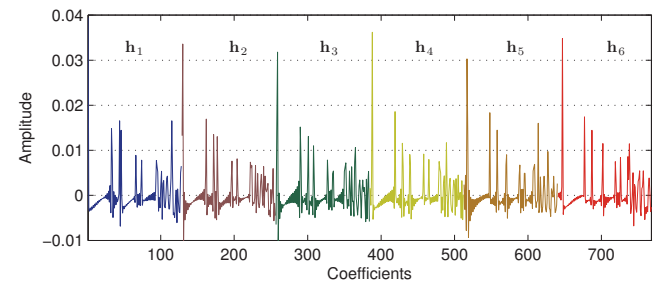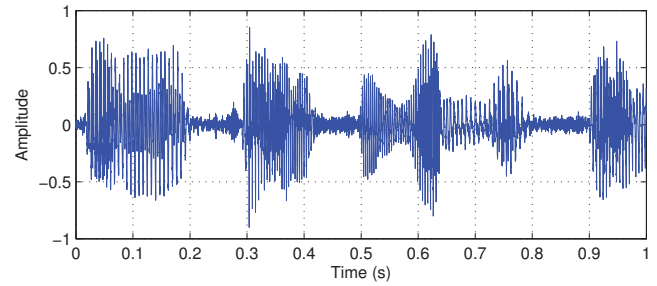[5] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification,"

Fig. 2. The MCQN and NMCFLMS algorithms are used to process a speech signal s(n) [top] convolved with acoustic impulse responses **h** [middle]. The resulting NPM as a function of time is depicted at the bottom.

IEEE Trans. Signal Process., vol. 51, no. 1, pp. 11–24, Jan. 2003.

[6] Y. Huang, J. Benesty, and J. Chen, "A blind channel identification-based two-stage approach to separation and dereverberation of speech signals in a reverberant environment," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 882–895, Sept. 2005.

[7] K. W. Brodlie, "An assessment of two approaches to variable metric methods," *Mathematical Programming*, vol. 12, pp. 344–355, 1977.

[8] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995.

[9] G. H. Golub and C. F. van Loan, *Matrix Computations*, MD: John Hopkins University Press, Balimore, third edition, 1996.