

SIGMA-DELTA RESOLUTION ENHANCEMENT FOR FAR-FIELD ACOUSTIC SOURCE SEPARATION

Amin Fazel and Shantanu Chakrabartty

Department of Electrical and Computer Engineering
Michigan State University
East Lansing, Michigan 48824-1226

ABSTRACT

Many source separation algorithms fail to deliver robust performance when applied to signals recorded using high-density microphone arrays where distance between sensor elements is much smaller than the wavelength of the signal of interest. This can be attributed to limited dynamic range (determined by analog-to-digital conversion) of the sensor which is insufficient to overcome the artifacts due to cross-channel redundancy, non-homogenous mixing and high-dimensionality of the signal space. In this paper we propose a novel framework that overcomes these limitations by integrating learning algorithms directly with analog-to-digital conversion. At the core of the proposed approach is a novel regularized min-max optimization approach that yields “delta-sigma” limit-cycles. An on-line adaptation modulates the limit-cycles to enhance resolution in the signal sub-spaces containing non-redundant information. Numerical experiments simulating far-field recording conditions demonstrate consistent improvements over a benchmark setup used for independent component analysis (ICA).

Index Terms— Sigma-delta modulation, independent component analysis, machine learning, microphone arrays

1. INTRODUCTION

One of several emerging areas where micro/nano-scale integration promises significant breakthroughs is in the field of acoustic sensing. It is envisioned that next generation of intelligent hearing devices will integrate hundreds of micro/nano-scale microphones [1], separate speech from noise, track conversations in cluttered environments and thus provide significant improvements in speech intelligibility for individuals with hearing impairments. Separation and localization of acoustic sources using micro/nano scale microphone arrays, however, poses a significant challenge due to fundamental limitations imposed by the physics of sound propagation [1]. The smaller the distance between the recording elements, the more difficult it is to measure localization and separation cues. In its classical setting, ICA [2] and other source separation approaches are formulated independent of the signal

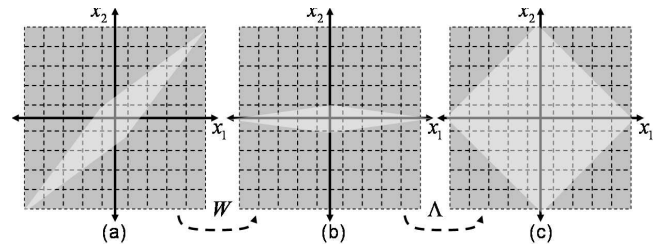


Fig. 1. Illustration of the proposed approach: (a) input signal distribution, (b) signal transformation and (c) resolution enhancement

measurement process (analog-to-digital process) and therefore do not consider the detrimental effects of finite resolution on the performance of the learning algorithm. However, in the case of micro/nano-scale microphone arrays, the mutual dependency of signal measurement and the learning algorithms can not be ignored due to the following reasons: (a) **Far-field effects:** Distance between recording elements on the array is much smaller than the distance of the sources to the sensor array. As a result, the mixing of signals at the sensors is near singular; (b) **Near-far effects:** A stronger source that is nearer to the sensor array can completely mask weak background sources; and (c) High-dimensionality of input analog signals due to high integration density of the microphones.

Recently, a least mean square (LMS) method [3] has been applied for resolving acute differences in analog acoustic signals. However, the approach is not scalable to larger arrays as it requires direct measurement of higher-order gradients. In this paper we present an analog-to-digital conversion algorithm that integrates learning directly with delta-sigma modulation. Traditionally, delta-sigma modulators have been the architecture of choice for any audio based processing as the topology is robust to analog imperfection and can easily resolve differences of more than 120 dB. (more than 16 bits). Our approach will be to formulate delta-sigma modulation within the framework of statistical learning such that the algorithm can optimally quantize non-redundant signal

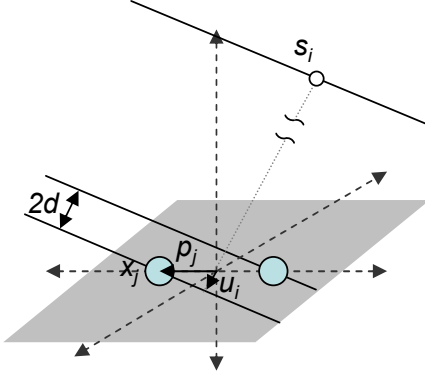


Fig. 2. Far-field recording on a miniature microphone array

sub-spaces. The core principle is depicted using Figure 1 which shows a typical distribution for a two-dimensional signal acquired through a high-density array. It can be seen in Figure 1(a) that the distribution is near singular and that the measurements x_1 and x_2 show high-degree of correlation. Our approach is to first determine a quantized representation of a transformation matrix \mathbf{W} , that will align the data distribution along the orthogonal axes, each representing an independent component (shown in Figure 1(b)). Based on this alignment, the scale of the quantization operation will be adjusted along each of the axes such that all the quantization levels (represented by ticks) span a significant region of the signal space (Figure 1(c)).

2. FAR-FIELD ACOUSTIC MODELING

In literature far-field acoustics have been extensively studied within the context of array processing and plenacoustic models [4]. We concisely describe a simplistic model that have been previously used for miniature microphone arrays. For audio signals (100-20,000 Hz), microphone arrays with inter-element distances less than $3.4cm$ (coherence length) can be approximated by far-field, where the acoustic wavefronts can be considered planar (see Figure 2). Also, for miniature microphone arrays the distance to the acoustic sources from the center of the array can be assumed to be larger than inter-element distance. We express the signal $x_j(\mathbf{p}_j, t)$ recorded at j^{th} microphone located at a 3-D position vector $\mathbf{p}_j = (x, y, z)$ as a superposition $i \in \{1, \dots, D\}$ independent sources $s_i(t)$ recorded at the reference microphone (located at the center of the array) [4]. This can be written as

$$x_j(\mathbf{p}_j, t) = \sum_i c_i(\mathbf{p}_j) s_i(t - \tau_i(\mathbf{p}_j)) \quad (1)$$

where $c_i(\mathbf{p}_j)$ and $\tau_i(\mathbf{p}_j)$ are the attenuation and delay, relative to the center of microphone array, for the source $s_i(t)$ at the position \mathbf{p}_j . Under far-field conditions it can be assumed that

$c_i(\mathbf{p}_j) \approx 1$ and $\tau_i(\mathbf{p}_j) \ll t$. Similar other treatments, equation (1) can be approximated using Taylor's series expansion as

$$x_j(\mathbf{p}_j, t) \approx \sum_i s_i(t) - \sum_i \tau_i(\mathbf{p}_j) \dot{s}_i(t). \quad (2)$$

The first right hand part of the equation (2) signifies a common-mode signal and the second part signifies an instantaneous mixture of the derivative of the source signals. Fortunately, for miniature arrays, the time delays can be expressed as linear terms as $\tau_i(\mathbf{p}_j) = \mathbf{u}_i^T \mathbf{p}_j / c$, where \mathbf{u}_i is the unit normal vector of the wavefront of source i . For distance of j^{th} microphone from reference position $|d|_{min} = \mathbf{u}_i^T \mathbf{p}_j = 1mm$, $c = 340m/s$ and signal frequency of $1000Hz$, any signal processor has to resolve signals less than $-70dB$ relative to the common-mode.

3. $\Sigma\Delta$ LEARNING ALGORITHM

In this section we describe a min-max optimization framework that unifies statistical learning with $\Sigma\Delta$ modulation. Given a random input vector $\mathbf{x} \in \mathcal{R}^M$ and an internal state vector $\mathbf{v} \in \mathcal{R}^M$, a $\Sigma\Delta$ learner determines a linear transformation matrix $\mathbf{W} \in \mathcal{R}^M \times \mathcal{R}^M$ according to the following optimization criterion:

$$\max_{\mathbf{W} \in \mathcal{C}} (\min_{\mathbf{v}} C(\mathbf{v}, \mathbf{W})) \quad (3)$$

where

$$C(\mathbf{v}, \mathbf{W}) = \Omega(\mathbf{v}) - \mathbf{v}^T E_{\mathbf{x}} \{ \mathbf{W}^T \mathbf{x} \}. \quad (4)$$

$E_{\mathbf{x}} \{ \cdot \}$ denotes an expectation operator with respect to the random variable \mathbf{x} . \mathcal{C} denotes a constraint space on the transformation matrix \mathbf{W} . $\Omega(\cdot)$ is a piece-wise linear regularization functions that will be used for implementing quantization operators. This is illustrated in Figure 3 which shows examples of one-dimensional regularization functions $\Omega(\cdot)$. The piece-wise behavior of $\Omega(\cdot)$ will lead to discontinuous gradients $Q = \nabla \Omega$ (shown in Figure 3) which are equivalent to functions used for signal quantization. The minimization step in equation (3) will ensure that the state vector \mathbf{v} is correlated with the transformed input signal $\mathbf{W}\mathbf{x}$ (tracking step) and the maximization step in (3) will adapt the parameters \mathbf{W} such that it minimizes the correlation (de-correlation step). The formulation bears similarities with game-theoretic approaches where tracking and de-correlation have been formulated as conflicting objectives. The uniqueness of the proposed approach, compared to other optimization techniques to solve (3) is the use of bounded gradients to generate $\Sigma\Delta$ limit-cycles. This is illustrated in Figure 4 which illustrates the proposed optimization procedure using a two-dimensional contour. Provided the input \mathbf{x} and the norm of the linear transformation $\|\mathbf{W}\|_{\infty}$ are bounded and the regularization function Ω satisfies the Lipschitz condition, the optimal solution to (3) is well defined and is given by $\mathbf{v}^* = 0$ (see Figure 4). In the proposed approach, however, only the path to the final

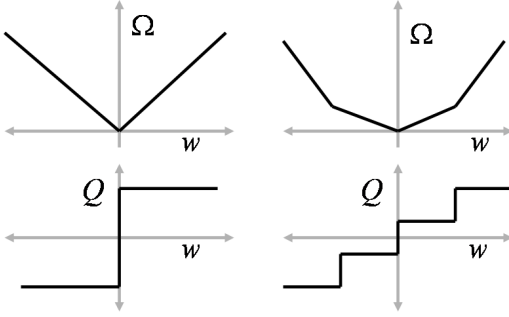


Fig. 3. One dimensional piece-wise linear regularization functions (left) two level (right) multi-level

solution \mathbf{w}^* and the limit-cycles about the solution (see Figure 4) will be of importance. The path and the limit-cycles will encode the topology of the optimization manifold which is defined by input vector \mathbf{x} and the transformation \mathbf{W} .

3.1. First-order $\Sigma\Delta$ Modulation

The link between optimization (3) and delta-sigma modulation is the minimization part in (3) where applying a stochastic gradient descent step yields

$$\mathbf{v}_n = \mathbf{v}_{n-1} + \mathbf{W}_n^T \mathbf{x}_n - \mathcal{D}_n \quad (5)$$

with n signifying the time steps and $\mathcal{D}_n = \nabla\Omega(\mathbf{v}_{n-1})$ being the quantized representation according to functions shown in Figure 3. Note that the formulation (5) does not require any learning rate parameters typically used in other neural network approaches. As the recursion (5) progresses bounded limit cycles are produced about the solution \mathbf{v}^* (see Figure 4). It can be shown that for $\|\mathbf{W}_n\|_\infty \leq 1$, $\|\mathbf{v}_n\|_\infty \leq 1$, which leads to $E_n\{\mathcal{D}_n\} \xrightarrow{n \rightarrow \infty} E_n\{\mathbf{W}_n \mathbf{x}_n\}$, where $E_n\{\cdot\}$ denotes an empirical expectation with respect to time indices n . Thus, recursion (5) produces a quantized sequence whose mean asymptotically encodes the transformed input at infinite resolution. It can also be shown that for a finite N iterations of (5) yields a quantized representation that is $\log_2(K)$ bits accurate.

3.2. $\Sigma\Delta$ de-correlation

The maximization step (de-correlation) in equation (3) yields updates for matrix \mathbf{W} according to:

$$\mathbf{W}_n = \mathbf{W}_{n-1} - 2^{-P} \mathcal{D}_n \psi(\mathbf{x}_n)^T; \mathbf{W}_n \in \mathcal{C} \quad (6)$$

where $\psi: \mathcal{R}^M \rightarrow \mathcal{R}^M$ function dependent on the transformation \mathbf{W} . For instance, $\psi(\cdot)$ could be chosen to be a quantized function which yields a completely digital update for (6). P in equation (6) is an update parameter which determines the resolution of the parameter matrix \mathbf{W} . In this paper, we have

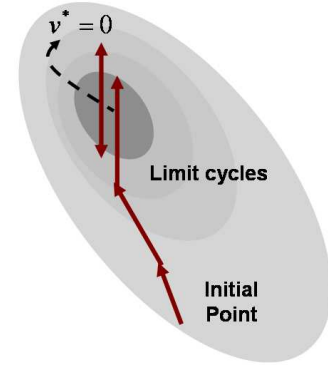


Fig. 4. Limit cycle behavior using bounded gradients

chosen $\psi(\mathbf{x}_n) = \mathcal{D}_n$ and the constraint space \mathcal{C} has been chosen to restrict \mathbf{W} to be a lower triangular matrix with all diagonal elements to be unity. The choice of this constraint guarantees convergence of the updates (5) and (6). It can be seen from the equation (6) that if $\|\mathbf{W}\|_\infty$ is bounded, recursion (6) will asymptotically lead to $E_n\{\mathcal{D}_n \mathcal{D}_n^T\} \rightarrow 0$ for $\mathbf{W}_\infty \in \mathcal{C}$. Thus, the proposed $\Sigma\Delta$ learning algorithm produces a quantized sequence that are mutually orthogonal.

3.3. $\Sigma\Delta$ Resolution Enhancement

One of the advantages of integrating signal de-correlation and dimensionality reduction with the analog-to-digital conversion is the ability to enhance the overall resolution of the system by “zooming” into the transformed signal space containing low energy (for example dimension x_2 in Figure 1(b)). This feature is essential for normalizing the signal power of independent sources, especially when one of the sources is masked by another dominant source or common-mode interference. The “zoom” mechanism can be incorporated by introducing a diagonal matrix $\Lambda \in \mathcal{R}^M \times \mathcal{R}^M$ into the cost function (4) as

$$\mathcal{C}(\mathbf{v}, \mathbf{W}, \Lambda) = \Omega(\Lambda^T \mathbf{v}) - \mathbf{v}^T E_x\{\mathbf{W}^T \mathbf{x}\}. \quad (7)$$

where the optimization (3) is also performed with respect to the parameter matrix Λ . The stochastic gradient step equivalent to recursion (5) is given by

$$\mathbf{v}_n = \mathbf{v}_{n-1} + (\mathbf{W}_{n-1}^T \mathbf{x}_n - \Lambda_n^T \mathcal{D}_n) \quad (8)$$

The asymptotic behavior of update (8) for equation (7) can be expressed as $E_n\{\mathcal{D}_n\} \xrightarrow{n \rightarrow \infty} \Lambda^{-1} E_n\{\mathbf{W}_n^T \mathbf{x}_n\}$. Thus reducing the magnitude of diagonal matrix Λ will result in an equivalent amplification of the transformed signal. The parameter Λ is determined based on the following element-wise update

$$\Lambda_i = \max |(\mathbf{W}_n \mathbf{x}_n)_i|; n > N_0 \quad (9)$$

which ensures that the updates (8) and (9) are always bounded.

4. RESULTS FROM NUMERICAL SIMULATIONS

For our experiments we simulated a recording conditions of a miniature array consisting of 4 omni-directional microphones. Three of the microphones were placed along a triangle with distance being 1mm, whereas the fourth microphone was placed at the centroid. The set up is similar to the conditions that have been reported in [5] where the simulation have been shown to closely approximate real-life anechoic conditions. To simulate the microphones' gain mismatch, the experiments were performed assuming up to 5% mismatch in the gain of the microphones.

For all experiments three independent speech signals were chosen as far-field sources. The outputs of the microphone array were first presented to the proposed $\Sigma\Delta$ learner, subsequent to which, only three of the outputs are used as inputs to the FastICA algorithm [2]. A benchmark used for comparative study consisted of $\Sigma\Delta$ converters which directly quantized the mixtures recorded at the microphones. The performance of the algorithms are quantified by signal-to-distortion ratio (SDR), signal-to-interference ratio (SIR) and signal-to-artifact ratio (SAR) [6] which takes into account degradation due to noise and cross-channel leakage.

The experimental results are summarized by Table 1 and Figure 5, where the SDR, SIR, and SAR are computed for each of the sources (S1-S3) for different values of over-sampling ratio (OSR) N . As the results show, the proposed $\Sigma\Delta$ learner (denoted by "with") outperforms its benchmark (denoted by "without").

5. CONCLUSION

In this paper, we have proposed a novel framework that integrates machine learning with analog-to-digital conversion. One of the applications of this integration is the ability to resolve acute differences in signals recorded using a miniature microphone array where classical approach of digitization followed by source separation fails to produce robust results. Using numerical simulations we have shown that the framework demonstrates consistent improvements in performance over a benchmark system when applied for independent component analysis.

6. REFERENCES

[1] R.N. Miles and R.R. Hoy, "The development of a biologically-inspired directional microphone for hearing aids," *Aud. and Neuro-Otology*: 11 (2), pp. 86-94, 2006.

[2] A. Hyvärinen, "Survey on independent component, analysis," *Neural Computing Surveys*, vol. 2, pp. 94-128, 1999.

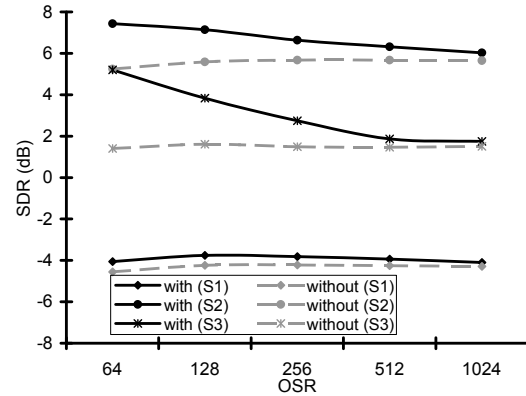


Fig. 5. Comparative study for different over-sampling ratio

Table 1. Comparative study of the proposed $\Sigma\Delta$ learner and classical $\Sigma\Delta$ converter

OSR		with learning			without learning		
		SDR	SIR	SAR	SDR	SIR	SAR
64	S1	-4.06	17.26	-3.94	-4.55	17.20	-4.44
	S2	7.43	22.90	7.58	5.24	16.09	5.72
	S3	5.20	36.70	5.20	1.40	26.05	1.43
128	S1	-3.76	14.96	-3.56	-4.24	16.85	-4.11
	S2	7.14	21.95	7.31	5.59	18.70	5.87
	S3	3.83	26.89	3.86	1.61	28.14	1.63
256	S1	-3.82	19.46	-3.75	-4.22	16.84	-4.10
	S2	6.64	18.33	7.00	5.67	16.98	6.09
	S3	2.75	30.31	2.76	1.49	40.29	1.49
512	S1	-3.94	17.14	-3.82	-4.26	16.78	-4.13
	S2	6.32	17.33	6.75	5.67	17.05	6.08
	S3	1.87	36.91	1.87	1.46	39.06	1.46
1024	S1	-4.10	17.03	-3.98	-4.29	16.76	-4.17
	S2	6.02	16.98	6.47	5.65	18.16	5.97
	S3	1.75	38.68	1.75	1.50	34.02	1.51

[3] A. Celik, M. Stanacevic, and G. Cauwenberghs, "Gradient flow independent component analysis in micropower VLSI," *Adv. NIPS*, 2006.

[4] M.N. Do, "Toward sound-based synthesis: the far-field case," *ICASSP*, Canada, 2004.

[5] G. Chabriel, J. Barrere, "An Instantaneous Formulation of Mixtures for Blind Separation of Propagating Waves," *IEEE Trans. Signal Processing*: 54 (1), pp. 49-58, 2006.

[6] E. Vincent, R. Gribonval, and C. Fvotte, "Performance measurement in Blind Audio Source Separation," *IEEE Trans. Audio, Speech and Language Processing*: 14 (4), pp. 1462-1469, Jul. 2006.