

Resolution Enhancement in $\Sigma\Delta$ Learners for Super-Resolution Source Separation

Amin Fazel, *Student Member, IEEE*, Amit Gore, and Shantanu Chakrabartty, *Member, IEEE*

Abstract—Many source separation algorithms fail to deliver robust performance when applied to signals recorded using high-density sensor arrays where the distance between sensor elements is much less than the wavelength of the signals. 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 large cross-channel redundancy, non-homogeneous mixing and high-dimensionality of the signal space. This paper proposes a novel framework that overcomes these limitations by integrating statistical learning directly with the signal measurement (analog-to-digital) process which enables high fidelity separation of linear instantaneous mixtures. At the core of the proposed approach is a min-max optimization of a regularized objective function that yields a sequence of quantized parameters which asymptotically tracks the statistics of the input signal. Experiments with synthetic and real recordings demonstrate significant and consistent performance improvements when the proposed approach is used as the analog-to-digital front-end to conventional source separation algorithms.

Index Terms—Source separation, Oversampling converters, $\Sigma\Delta$ modulation, Analog-to-information converters, High-density sensing, Super-resolution.

I. INTRODUCTION

ADVANCES in miniaturization are enabling integration of an ever increasing number of recording elements within a single sensor device [1]. Examples of such high-density sensors include microelectrode arrays used for recording neural signals [2], [3], [4] and microphone arrays used for next generation hearing prosthesis [5], [6], [7]. A key challenge in high-density sensing is to be able to image events (acoustic, optical or electrical) occurring in its environment with high fidelity (spatial and temporal). However, due to the dispersive nature of the surrounding media, each element of the sensor array records a mixture of signals generated by independent events in its environment. Recovery of the independent signals from the recorded mixtures lies within the domain of blind source separation (special instance of manifold learning) where algorithms like independent component analysis (ICA) have been employed [8], [9], [10], [11], [12], [13], [14]. However, several factors limit the performance of traditional source separation techniques when applied to analog signals acquired from high-density sensor arrays:

- **Far-field effects:** For miniature arrays, sources are usually located at distances much larger than the distance between recording elements. As a result, the mixing of signals at the recording elements is near singular. Separating independent sources from near ill-conditioned mixtures would require super-resolution signal processing to reliably identify the parameters of the separation manifold.
- **Near-far effects:** For miniature sensor arrays, a stronger source that is closer to the array can mask the signal produced by background sources. Separating the background sources in presence of the strong masker would again require super-resolution processing of the input signals.

DSP based source separation algorithms are typically implemented subsequent to a quantization operation (analog-to-digital conversion) and hence do not consider the detrimental effects of finite resolution due to the quantizer. In particular, for a high-density sensor array, a naive implementation of a quantizer that uniformly partitions each dimension (pulse coded modulation) of the input signal space could lead to a significant loss of information. To understand the effect of this degradation consider the framework of a conventional source separation algorithm as shown in Fig. 1(a). The “analog” signal $\mathbf{x} \in \mathcal{R}^M$ recorded at each of the sensor array is given by $\mathbf{x} = \mathbf{A}\mathbf{s}$, $\mathbf{A} \in \mathcal{R}^M \times \mathcal{R}^M$ being the mixing matrix and $\mathbf{s} \in \mathcal{R}^M$ being the independent sources of interest. This simplified linear model is applicable to both instantaneous mixing as well as to convolutive mixing formulation [16]. The recorded signals are first digitized and then processed by a digital signal processor (DSP) which implements the source separation algorithm (for example FastICA or SOBI [25]). For the sake of simplicity, assume that the algorithm is able to identify the correct unmixing matrix given by $\mathbf{W} = \mathbf{A}^{-1}$ which is then used to recover the source signals $\tilde{\mathbf{s}} \in \mathcal{R}^M$. The effect of quantization in this approach can be captured using a simple additive model as shown in Fig. 1(a)

$$\tilde{\mathbf{s}}_d = \mathbf{W}(\mathbf{x} + \mathbf{q}) = \mathbf{s} + \mathbf{A}^{-1}\mathbf{q} \quad (1)$$

where \mathbf{q} denotes the additive quantization error introduced during the digitization process. The reconstruction error between the recovered signal $\tilde{\mathbf{s}}_d$ and the source signal \mathbf{s} can then be expressed as

$$\|\tilde{\mathbf{s}}_d - \mathbf{s}\| = \|\mathbf{A}^{-1}\mathbf{q}\| \leq \|\mathbf{A}^{-1}\| \cdot \|\mathbf{q}\| \quad (2)$$

where $\|\cdot\|$ denotes a matrix and vector norm [23]. Equation (2) indicates that under ideal reconstruction conditions, the performance of conventional source separation algorithm is limited

Copyright (c) 2008 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

A. Fazel and S. Chakrabartty are with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, 48824 USA (e-mails: {fazel,shantanu}@egr.msu.edu).

A. Gore is with General Electric Global Research, Niskayuna, NY 12309 USA.

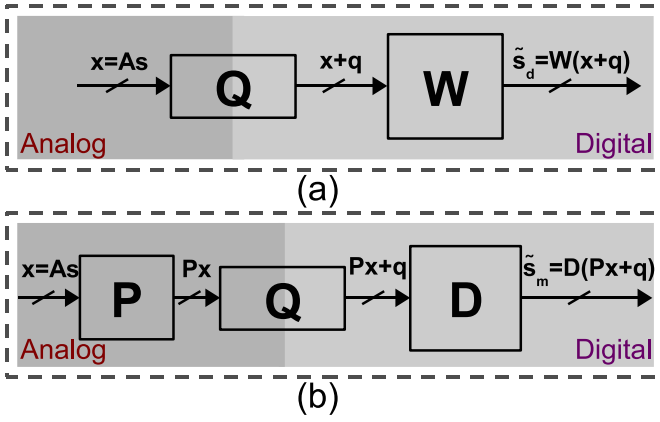


Fig. 1: System architecture where the source separation algorithm is applied (a) after quantization (b) after analog projection and quantization

by: (a) quantization error (accuracy of analog-to-digital conversion) and (b) the nature of the mixing operation determined by \mathbf{A} . For high-density sensors, the mixing typically tends to be ill-conditioned ($\|\mathbf{A}^{-1}\| \gg 1$), as a result the reconstruction error due to equation (2) could be large.

Now consider the framework shown in Fig.1(b) which is at the core of the proposed resolution enhancement approach. The signals recorded by the sensor array is first transformed by \mathbf{P} (in the analog domain) before being quantized. In this case, the reconstructed signal $\tilde{s}_m \in \mathcal{R}^M$ can be expressed as

$$\tilde{s}_m = \mathbf{D}(\mathbf{P}\mathbf{x} + \mathbf{q}). \quad (3)$$

For source separation $\mathbf{D}\mathbf{P} = \mathbf{A}^{-1}$, for which the reconstructed signal now can be expressed as

$$\tilde{s}_m = \mathbf{s} + \mathbf{A}^{-1}\mathbf{P}^{-1}\mathbf{q} \quad (4)$$

which leads to the reconstruction error

$$\|\tilde{s}_m - \mathbf{s}\| = \|\mathbf{A}^{-1}\mathbf{P}^{-1}\mathbf{q}\| \leq \|\mathbf{A}^{-1}\mathbf{P}^{-1}\| \cdot \|\mathbf{q}\|. \quad (5)$$

Thus, the reconstruction error can now be controlled by the choice of the transform \mathbf{P} and is not completely determined by the mixing transform \mathbf{A} . An interesting choice of the matrix \mathbf{P} is the one that satisfies

$$\|\mathbf{A}^{-1}\mathbf{P}^{-1}\| = 1 \quad (6)$$

which ensures that the input signals are normalized before processed by the DSP based source separation algorithm. Equation 5 then reduces to

$$\|\tilde{s}_m - \mathbf{s}\| \leq \|\mathbf{q}\| \quad (7)$$

and the expected performance improvement when employing the framework in Fig. 1(b) over the framework in Fig. 1(a) is given by

$$PI = -20 \log \|\mathbf{A}^{-1}\|. \quad (8)$$

Equation (8), thus, shows that the for near-singular mixing, $\|\mathbf{A}^{-1}\| \gg 1$, the performance improvement based on the resolution enhancement technique shown in Fig. 1(b) could be significant. However, the performance improvement is valid

only if the analog projection \mathbf{P} can be precisely and adaptively determined during the process of quantization (“analog-to-digital” conversion). This procedure is unlike traditional multi-channel “analog-to-digital” conversion where each signal channel is uniformly quantized the input signal without taking into consideration the spatial statistics of the input signal. Since the projection P is also quantized, the precision to which the condition (6) is satisfied is also important. In this regard, oversampling “analog-to-digital” converters like $\Sigma\Delta$ modulators are attractive since the topology is robust to analog imperfection and can easily achieve dynamic ranges greater than 120 dB (more than 16 bits or accuracy) [20]. In this paper, we extend our work on $\Sigma\Delta$ learning [22] to include theoretical and measurement results obtained for separating sources whose mixtures are near-singular. In particular, we show that the learning algorithm can efficiently and adaptively quantize non-redundant analog signal sub-spaces which leads to significant performance improvement for any DSP based source separation algorithm.

The paper is organized as follows: Section II presents a $\Sigma\Delta$ learning framework that estimates the matrix \mathbf{P} using an online stochastic gradient algorithm. The algorithm processes samples a continuous time analog signal and produces a sequence of quantized representations that encodes the matrix \mathbf{P} and the transformed input signal. We present some of the theoretical properties of the proposed approach and demonstrate the benefits of the approach using representative examples. Section III and IV presents experimental results obtained when the proposed sigma-delta learner interfaces with a conventional ICA module and is used for separating speech samples recorded using a miniature microphone array. Section V concludes with some remarks about the future work in this area. Before we present the $\Sigma\Delta$ learning algorithm in detail we summarize some of the mathematical notations that will be used in this paper.

\mathbf{A} (bold capital letters)	A Matrix and its elements will be denoted by $a_{ij}, i = 1, 2, \dots; j = 1, 2, \dots$
x (normal lowercase letters)	A Scalar variable.
\mathbf{x} (bold lowercase letters)	A Vector and its elements will be denoted by $x_i, i = 1, 2, \dots$
$\mathbf{x}[n]$	A sequence of vectors where $n = 1, 2, \dots$ denotes a discrete time index.
$\ \mathbf{x}\ _p$	The L_p norm of a vector and is given by $\ \mathbf{x}\ _p = (\sum_i x_i ^p)^{1/p}$.
\mathbf{A}^T	Transpose of \mathbf{A} .
$\mathbf{x}[n]$	Discrete time sequences.
$\mathcal{E}_{\mathbf{x}}\{\cdot\}$	Statistical expectation and is given by $\mathcal{E}_{\mathbf{x}}\{\cdot\} = \int_{\mathbf{x}} (\cdot) p_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}$, where $p_{\mathbf{x}}(\mathbf{x})$ denotes the probability density function of \mathbf{x} .
$\mathcal{E}_n\{\cdot\}$	Empirical expectation and is defined as a temporal average of a discrete time sequence: $\mathcal{E}_n\{\mathbf{d}_n\} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mathbf{d}[n]$.

II. OPTIMIZATION FRAMEWORK FOR $\Sigma\Delta$ LEARNING

The underlying principle behind the proposed resolution enhancement technique is illustrated using Fig. 2 which shows a two dimensional signal distribution along with the respective signal quantization levels (depicted using rectangular tick

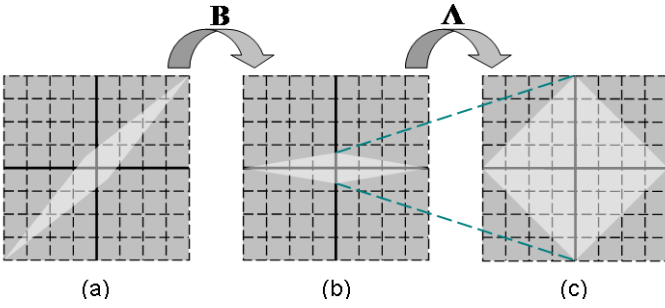


Fig. 2: Illustration of the two-dimensional signal distribution for: (a) the input signals ; (b) signals obtained after transformation \mathbf{B} and (c) signals obtained after resolution enhancement

marks). In this example, the signal distribution has been chosen to cover only a small region of the quantization space which would be the case for near-singular mixing. Thus, in a traditional implementation where each dimension is quantized independently of the other there would be a significant information loss due to quantization. Our approach towards estimating \mathbf{P} while performing signal quantization will be to decompose $\mathbf{P} \in \mathcal{R}^M \times \mathcal{R}^M$ as a product two simple matrices $\mathbf{\Lambda} \in \mathcal{R}^M \times \mathcal{R}^M$ and $\mathbf{B} \in \mathcal{R}^M \times \mathcal{R}^M$ such that $\mathbf{P} = \mathbf{\Lambda}\mathbf{B}$. The transformation matrix \mathbf{B} will first “approximately” align the data distribution along the orthogonal axes, each axis representing an independent (orthogonal) component (shown in Fig. 2(b)). Based on this alignment, the signal distribution will be scaled according to a diagonal matrix $\mathbf{\Lambda}$ such that the quantization levels now span a significant region of the signal space (Fig. 2(c)). Our objective will be to compute these transforms \mathbf{B} and $\mathbf{\Lambda}$ recursively while performing signal quantization. Even though the proposed procedure bears similarity to recursive techniques reported in many online “whitening” algorithms [15], [16], [13], the key difference for the proposed approach is that the adaptation and estimation of the projection matrix P is coupled with the quantization process. Thus, unlike traditional online “whitening” techniques, in the proposed approach any imperfections or errors in the quantization process can be corrected through the adaptation of P . This approach can therefore be visualized as a “smart” analog-to-digital converter as shown in Fig. 3 which not only produces quantized (digitized) representation of the input signal \mathbf{d} but also quantized (digitized) representations of the transforms \mathbf{B} and $\mathbf{\Lambda}$. In our formulation, the estimation algorithm for P (\mathbf{B}

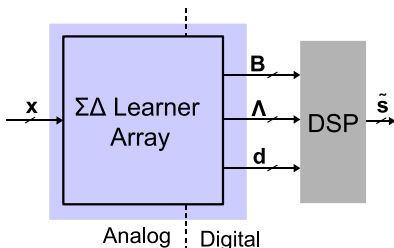


Fig. 3: Architecture of the proposed sigma-delta learning applied to a source separation problem

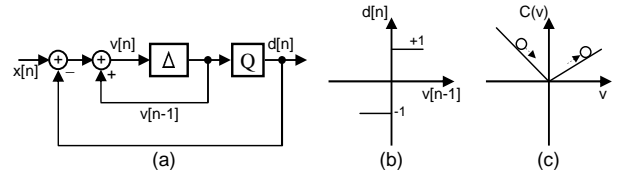


Fig. 4: (a) System architecture of a first order $\Sigma\Delta$ modulator; (b) input-output response of single bit quantizer; and (c) illustration of “limit-cycle” oscillations about the minima of the cost function $C(v)$

and $\mathbf{\Lambda}$) has been integrated within a $\Sigma\Delta$ modulation algorithm and hence the name “ $\Sigma\Delta$ learning. The choice of $\Sigma\Delta$ modulation is due to its robustness to hardware level artifacts (mismatch and non-linearity) which makes the modulation amenable for implementing high-resolution analog-to-digital converters [20]. Before we present a generalized formulation for $\Sigma\Delta$ learning we first we first present an optimization framework that can model the dynamics of first-order $\Sigma\Delta$ modulation.

A. Stochastic Gradient Descent and $\Sigma\Delta$ Modulators

First, we will use a one dimensional example to illustrate how a $\Sigma\Delta$ modulator can be modeled as an equivalent stochastic gradient descent based optimization problem. Consider an architecture of a well known first-order $\Sigma\Delta$ modulator [20] as shown in Fig. 4(a) which consists of a single discrete-time integrator in a feedback loop. The loop also consists of a quantizer Q which produces a sequence of digitized representation $d[n]$, where $n = 1, 2, \dots$ denotes a discrete time-index. Let $x[n] \in \mathcal{R}$ be the sampled analog input to the modulator and without any loss of generality let $d[n] \in \{+1, -1\}$ be the output of a single-bit quantizer given by $d[n] = \text{sgn}(v[n-1])$ (see Fig. 4(b)) where $v_n \in \mathcal{R}$ is the internal state variable or the output of the integrator as shown in Fig. 4(a). Then, the $\Sigma\Delta$ modulator in Fig. 4(a) implements the following recursion:

$$v[n] = v[n-1] + x[n] - d[n] \quad (9)$$

It can be seen from equation (9) that if $v[n]$ is bounded for all n , then

$$\frac{1}{N} \sum_{n=1}^N d[n] \xrightarrow{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N x[n]. \quad (10)$$

This implies that $\Sigma\Delta$ algorithm given by equation (9) produces a binary sequence $d[n]$ whose temporal average asymptotically converges to the temporal average of the input analog signal. This statistical dynamics is at the core of most $\Sigma\Delta$ modulators. However, from the perspective of statistical learning, the $\Sigma\Delta$ recursion in equation (9) can be viewed as a stochastic gradient step of the following optimization problem:

$$\min_v C(v) = \min_v [|v| - v\mathcal{E}_x(x)] \quad (11)$$

where $\mathcal{E}_x(\cdot)$ is the ensemble expectation of the random variable x . The optimization function $C(v)$ is shown in Fig. 4(c) for the case $|\mathcal{E}_x(x)| < 1$. The minima under this condition is $\min_v C(v) = 0$ which is achieved for $v = 0$ and thus does not contain any information about the statistical property of

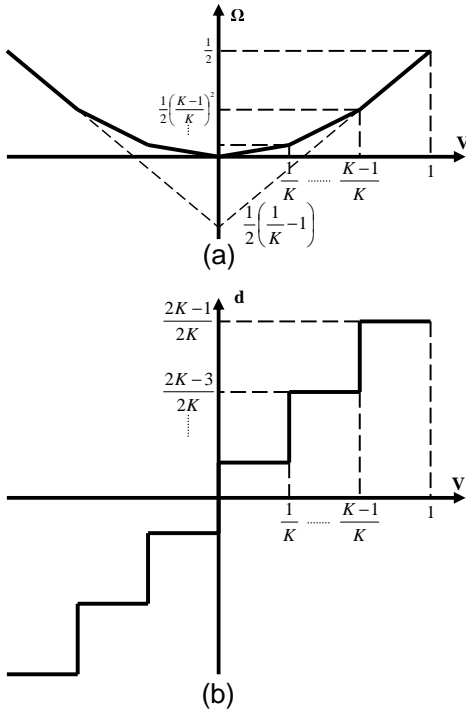


Fig. 5: One dimensional piece-wise linear regularization functions and the multi-bit quantization function as its gradient

x . The recursion (9) ensures that $v[n]$ approaches the minima and then exhibits oscillations about the minima (shown in Fig. 4(c)). Note that unlike conventional stochastic gradient based optimization techniques [21], recursion (9) does not require any learning rate parameters. This is unique to the proposed optimization framework where the stochastic gradient descent is used to generate limit-cycles (oscillations) about the minima of the cost function $C(v)$. The only requirement in such a framework is that the assumption that the input signal is bounded which ensures that the limit-cycles are bounded. Under this condition, the statistics of the limit-cycles can asymptotically encode the statistics of the input signal with infinite precision, as shown by equation (10). In the later sections, we will exploit this asymptotic property to precisely estimate the transform \mathbf{P} which can then be used for resolving the acute spatial cues in high-density sensor arrays. Another unique aspect of the proposed optimization framework for modeling $\Sigma\Delta$ modulators is that the cost function $C(v)$ links “analog-to-digital” conversion through the regularizer $|v|$ whose derivative leads to a single-bit quantizer (sgn function). The second term in $C(v)$ ensures that the statistics of the quantized stream $d[n]$ matches the statistics of the input analog signal $x[n]$.

We now extend this optimization framework to a multi-dimensional $\Sigma\Delta$ modulator which uses a multi-bit quantizer and incorporates transformations \mathbf{B} and Λ . Consider the following minimization problem

$$\min_{\mathbf{v}} \mathcal{C}(\mathbf{v}) \quad (12)$$

where the cost function $\mathcal{C}(\mathbf{v})$ is given by

$$\mathcal{C}(\mathbf{v}) = \Omega(\lambda^{-1}\mathbf{v}) - \mathbf{v}^T \mathcal{E}_x\{\mathbf{B}\mathbf{x}\}. \quad (13)$$

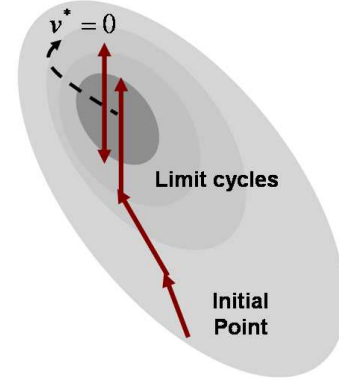


Fig. 6: Limit cycle behavior using bounded gradients

$\mathbf{x} \in \mathcal{R}^M$ is now an M dimensional analog input vector and $\mathbf{v} \in \mathcal{R}^M$ is an internal state vector. For the first part of this formulation, λ will be assumed to be a constant scalar and the transform B will be assumed to be a constant matrix. $\Omega(\cdot)$ denotes a piece-wise linear regularization function that is used for implementing quantization operators. An example of a regularization function $\Omega(\cdot)$ is shown in Fig. 5 for 1-dimensional input vector \mathbf{v} . Due to the piece-wise nature of the function $\Omega(\cdot)$ its gradient $\mathbf{d} = \nabla\Omega$ (shown in Fig. 5) is equivalent to scalar quantization operators. Without loss of generality, it will be assumed that the range of the quantization operator is limited between $[-1, 1]$. Therefore, for a $2K$ step quantization function the corresponding regularization function $\Omega(\cdot)$ is given by

$$\Omega(\mathbf{v}) = \sum_{j=1}^M \left| \frac{i}{2K} v_j \right|; \quad |v_j| \in [i-1, i] \quad (14)$$

for $i = 1, \dots, 2K$.

To reiterate, the uniqueness of the proposed approach, compared to other optimization techniques to solve (12) is the use of bounded gradients to generate $\Sigma\Delta$ limit-cycles. This is illustrated in Fig. 6 showing the proposed optimization procedure using a two-dimensional contour. Provided the input \mathbf{x} and the norm of the linear transformation $\|\mathbf{B}\|_\infty$ are bounded and the regularization function Ω satisfies the Lipschitz condition, the optimal solution to (12) is well defined and bounded from below which is shown in the next lemma:

Lemma 1: For the bounded matrix $\|\mathbf{B}\|_\infty \leq \lambda^{-1}$, bounded vector $\|\mathbf{x}\|_\infty \leq 1$, \mathcal{C} as defined in equation (13) is convex and is bounded by below according to $\mathcal{C}^* = \min_{\mathbf{v}} \mathcal{C}(\mathbf{v}) > \frac{1}{2}(\frac{1}{K} - 1)$.

Proof: We will use topological property of norms [23] which states that for two integers p, q satisfying $\frac{1}{p} + \frac{1}{q} = 1$, the following relationship is valid for vectors \mathbf{v} and \mathbf{u}

$$|\mathbf{v}^T \mathbf{u}| \leq \|\mathbf{v}\|_p \|\mathbf{u}\|_q \quad (15)$$

Setting $\mathbf{u} = \mathcal{E}_x\{\mathbf{B}\mathbf{x}\}$ and applying equation (15) the following inequality is obtained:

$$\|\mathbf{v}\|_1 \|\mathbf{u}\|_\infty \geq |\mathbf{v}^T \mathbf{u}| \geq \mathbf{v}^T \mathbf{u} \geq \mathbf{v}^T \mathcal{E}_x\{\mathbf{B}\mathbf{x}\} \quad (16)$$

It can be easily verified that $\Omega(\mathbf{v}) \geq \|\mathbf{v}\|_1 - \frac{1}{2}(\frac{1}{K} - 1)$ which is shown graphically in Fig. 5 for a one dimensional case and hence can be extended element-wise to the multi-dimensional case.

Using the definition of the matrix norm and the given constraints, it can be easily seen $\|\mathbf{B}\|_\infty \geq \|\mathbf{B}\mathbf{x}\|_\infty \geq \|\mathbf{u}\|_\infty$. Thus, $\|\mathbf{u}\|_\infty \leq \lambda^{-1}$. Therefore, the inequality (16) leads to

$$\Omega(\lambda^{-1}\mathbf{v}) - \mathbf{v}^T \mathcal{E}_x\{\mathbf{B}\mathbf{x}\} \geq \lambda^{-1}\|\mathbf{v}\|_1 - \frac{1}{2}(\frac{1}{K} - 1) - \mathbf{v}^T \mathcal{E}_x\{\mathbf{B}\mathbf{x}\} \geq 0. \quad (17)$$

which proves that the cost function $\mathcal{C}(\mathbf{v})$ is bounded from below by \mathcal{C}^* .

□

However, for $\Sigma\Delta$ learning the trajectory toward the minima of the cost function (13) is of importance. A stochastic gradient minimization corresponding to the optimization problem (13) leads to

$$\mathbf{v}[n] = \mathbf{v}[n-1] + \mathbf{B}\mathbf{x}[n] - \lambda^{-1}\mathbf{d}[n] \quad (18)$$

with n signifying the time steps and $\mathbf{d}[n] = \nabla\Omega(\mathbf{v}[n-1])$ being the quantized representation according to functions shown in Fig. 5. Note also that formulation (18) does not require any learning rate parameters. As the recursion (18) progresses, bounded limit cycles are produced about the solution \mathbf{v}^* (see Fig. 6).

The following two lemma exploits the property of the first-order modulator to show that the auxiliary state variable $\mathbf{v}[n]$ defined by (18) is uniformly bounded if the input random vector \mathbf{x} and matrix \mathbf{B} are uniformly bounded.

Lemma 2: For any bounded input vector sequence satisfying $\|\mathbf{x}[n]\|_\infty \leq 1$ and the transformation matrix \mathbf{B} satisfying $\|\mathbf{B}\|_\infty \leq \lambda^{-1}$, the internal state vector $\mathbf{v}[n]$ defined by equation (18) is always bounded, i.e., $\|\mathbf{v}[n]\|_\infty \leq 2\lambda^{-1}$ for $n = 1, 2, \dots$

Proof: We will apply mathematical induction to prove this lemma. Without any loss of generality we will assume $\|\mathbf{v}[0]\|_\infty \leq 2\lambda^{-1}$. Suppose $\|\mathbf{v}[n-1]\|_\infty \leq 2\lambda^{-1}$, it then follows that $\|\mathbf{v}[n-1] - \nabla\Omega(\mathbf{v}[n-1])\|_\infty = \|\mathbf{v}[n-1] - \mathbf{d}[n]\|_\infty \leq \lambda^{-1}$. Because \mathbf{x} and \mathbf{B} are bounded and using equation (18), the following relationship holds

$$\begin{aligned} \|\mathbf{v}[n]\|_\infty &= \|\mathbf{v}[n-1] - \lambda^{-1}\mathbf{d}[n] + \mathbf{B}\mathbf{x}[n]\|_\infty \\ &\leq \|\mathbf{v}[n-1] - \lambda^{-1}\mathbf{d}[n]\|_\infty + \|\mathbf{B}\|_\infty \\ &\leq 2\lambda^{-1} \end{aligned} \quad (19)$$

□

Lemma 3: For any bounded input vector $\|\mathbf{x}\|_\infty \leq 1$ and bounded transformation matrix \mathbf{B} , $\mathbf{d}[n]$ asymptotically satisfies estimates $\mathcal{E}_n\{\mathbf{d}[n]\} \xrightarrow{n \rightarrow \infty} \lambda \mathcal{E}_n\{\mathbf{B}\mathbf{x}[n]\}$.

Proof: Following N update steps the recursion given by equation (18) yields

$$\mathbf{B}\mathbf{x}[n] - \lambda^{-1}\mathcal{E}_n\{\mathbf{d}[n]\} = \frac{1}{N}(\mathbf{v}[N] - \mathbf{v}[0]) \quad (20)$$

which using the bounded property of random vector \mathbf{v} asymptotically leads to

$$\mathcal{E}_n\{\mathbf{d}[n]\} \xrightarrow{n \rightarrow \infty} \lambda \mathcal{E}_n\{\mathbf{B}\mathbf{x}[n]\} \quad (21)$$

□

Thus, according to Lemma 3, recursion (18) produces a quantized sequence whose mean asymptotically encodes the scaled transformed input at infinite resolution. It can also be shown that for a finite I iterations of (18) yields a quantized representation that is $\log_2(I)$ bits accurate.

B. $\Sigma\Delta$ Learning

In this section, we will extend the optimization framework to include on-line estimation of the transform \mathbf{B} . Here we again assume that λ is constant. Given an M dimensional random input vector $\mathbf{x} \in \mathcal{R}^M$ and an internal state vector \mathbf{v} , the $\Sigma\Delta$ learning algorithm estimates parameters of a linear transformation matrix $\mathbf{B} \in \mathcal{R}^M \times \mathcal{R}^M$ according to the following optimization function

$$\max_{\mathbf{B} \in \mathcal{C}} (\min_{\mathbf{v}} \mathcal{C}(\mathbf{v}, \mathbf{B})) \quad (22)$$

where

$$\mathcal{C}(\mathbf{v}, \mathbf{B}) = \Omega(\lambda^{-1}\mathbf{v}) - \mathbf{v}^T \mathcal{E}_x\{\mathbf{B}\mathbf{x}\}. \quad (23)$$

\mathcal{C} denotes a constraint space on the transformation matrix \mathbf{B} . The minimization step in equation (22) will ensure that the state vector \mathbf{v} is correlated with the transformed input signal $\mathbf{B}\mathbf{x}$ (tracking step) and the maximization step in (22) will adapt the matrix \mathbf{B} such that it minimizes the correlation (de-correlation step).

The stochastic gradient descent step corresponding to the minimization yields the recursion

$$\mathbf{v}[n] = \mathbf{v}[n-1] + \mathbf{B}[n]\mathbf{x}[n] - \lambda^{-1}\mathbf{d}[n]. \quad (24)$$

where $\mathbf{B}[n]$ denotes the transform matrix obtained at time instant n . The transform \mathbf{B} is then updated according to a stochastic gradient ascent step given by

$$\mathbf{B}[n] = \mathbf{B}[n-1] - 2^{-P}\mathbf{v}[n-1]\mathbf{x}[n]^T; \quad \mathbf{B}[n] \in \mathcal{C}. \quad (25)$$

P in equation (25) is a parameter which determines the resolution of updates the parameter matrix \mathbf{B} . If we assume that locally the matrix \mathbf{B}^* behaves as a positive definite matrix, equation (25) can be rewritten as

$$\begin{aligned} \mathbf{B}[n] &= \mathbf{B}[n-1] - 2^{-P}\mathbf{v}[n-1](\mathbf{B}[n]\mathbf{x}[n])^T \\ &= \mathbf{B}[n-1] - 2^{-P}\mathbf{d}[n]\mathbf{d}[n]^T \end{aligned} \quad (26)$$

where we have replaced the transformed input $\mathbf{B}[n]\mathbf{x}[n]$ by its asymptotic quantized representation $\mathbf{d}[n]$. Similarly we have replaced $\mathbf{v}[n-1]$ by its quantized representation $\mathbf{d}[n]$. The update can be generalized further by incorporating non-linear quantization function $\phi(\cdot)$ as

$$\mathbf{B}[n] = \mathbf{B}[n-1] - 2^{-P}\phi(\mathbf{d}[n])\mathbf{d}[n]^T \quad (27)$$

where $\phi : \mathcal{R}^M \rightarrow \mathcal{R}^M$ are functions dependent on the transformation \mathbf{B} . Here, we have chosen $\phi(\cdot) = \tanh(\cdot)$ and the constraint space \mathcal{C} has been chosen to restrict \mathbf{B} to be a lower triangular matrix with all diagonal elements to be unity. One of the ways to ensure that $\mathbf{B}[n] \in \mathcal{C} \quad \forall n$ is to apply the updates only to lower triangular elements $b_{ij}; i > j$. The

choice of this constraint guarantees convergence of the $\Sigma\Delta$ learning by ensuring \mathbf{B} is bounded.

Lemma 4: If the transform matrix \mathbf{B} is bounded then the quantized sequences $d_i[n]$ and $d_j[n]$ with $i \neq j$ are uncorrelated with respect to each other.

Proof: Using equation (27) the following relationships are obtained:

$$\begin{aligned} -2^{-P} \mathbf{d}[n] \phi(\mathbf{d}[n])^T &= \mathbf{B}[n] - \mathbf{B}[n-1] \\ -2^{-P} \mathcal{E}_n \{ \mathbf{d}[n] \phi(\mathbf{d}[n])^T \} &= \lim_{N \rightarrow \infty} \frac{\mathbf{B}[N]}{N} \\ \mathcal{E}_n \{ d_i[n] \phi(d_j[n]) \} &= 0 \quad \forall i \neq j \end{aligned} \quad (28)$$

Since this relationship holds for a generic form of $\phi(\cdot)$, the sequences $d_i[n]$ are (non-linearly) uncorrelated with respect to each other. \square

Equation (28) also provides a mechanism of reconstructing the input signal using the transformed output $\mathbf{d}[n]$ and the converged estimate of the transformation matrix $\mathbf{B}[n] \xrightarrow{n \rightarrow \infty} \mathbf{B}_\infty$. The input signal can be reconstructed using

$$\hat{\mathbf{x}} = \mathbf{B}_\infty^{-1} \lambda^{-1} \mathcal{E}_n \{ \mathbf{d}[n] \}. \quad (29)$$

The use of lower-triangular transforms for \mathbf{B} greatly simplifies the computation of the inverse \mathbf{B}_∞^{-1} through use of back-substitution techniques. Also, due to its lower-triangular form, the inverse of \mathbf{B}_∞ always exists and is well defined.

C. Resolution Enhancement

Once the transform \mathbf{B} has been determined such that the output of the $\Sigma\Delta$ learner is “de-correlated”, we can apply resolution enhancement by “zooming” into the transformed signal space that does not cover the quantization regions (see Fig. 2(b)). This can be achieved using another diagonal matrix Λ^{-1} which scales each axes as shown in the illustration 2(c). The $\Sigma\Delta$ cost function can be appropriately transformed to include the diagonal matrix $\Lambda \in \mathcal{R}^M \times \mathcal{R}^M$ as

$$C(\mathbf{v}, \mathbf{B}, \Lambda) = \Omega(\Lambda^{-1} \mathbf{v}) - \mathbf{v}^T \mathcal{E}_x \{ \mathbf{B} \mathbf{x} \}. \quad (30)$$

where the optimization (23) is also performed with respect to the parameter matrix Λ such that the constraint $\|B\|_\infty < \|\Lambda^{-1}\|_\infty$ is satisfied. This constraint is to ensure that $C(\mathbf{v}, \mathbf{B}, \Lambda)$ is always bounded from below. The stochastic gradient step equivalent to recursion (24) is given by

$$\mathbf{v}[n] = \mathbf{v}[n-1] + (\mathbf{B}[n-1] \mathbf{x}[n] - \Lambda^{-1}[n] \mathbf{d}[n]) \quad (31)$$

The asymptotic behavior of update (31) for equation (30) can be expressed as $\mathcal{E}_n \{ \mathbf{d}[n] \} \xrightarrow{n \rightarrow \infty} \Lambda \mathcal{E}_n \{ \mathbf{B}_n \mathbf{x}_n \}$. Thus, reducing the magnitude of diagonal elements of Λ will result in an equivalent amplification of the transformed signal. To satisfy the constraint on the transform \mathbf{B} and Λ , a suitable update for the diagonal matrix Λ and its elements λ_i are

$$\lambda_i = \max |(\mathbf{B}[n] \mathbf{x}[n])_i|; N_1 > n > N_0 \quad (32)$$

where N_0 is the number of iterations required for the matrix \mathbf{B} to stabilize and N_1 is the maximum observation period used to determine the parameters λ_i .

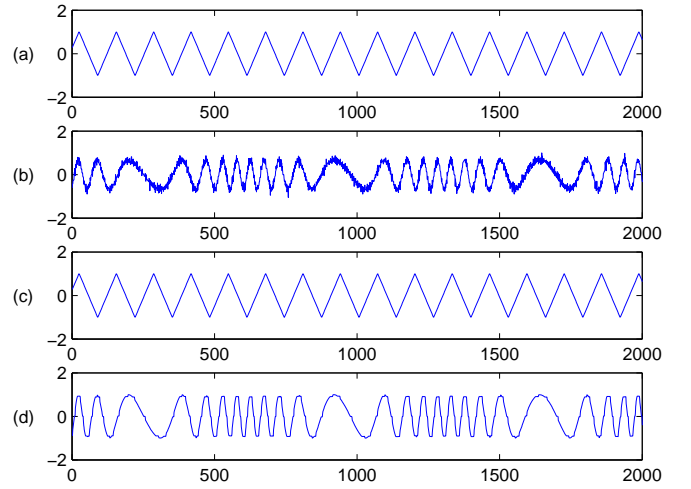


Fig. 7: Reconstruction of the sources using conventional and proposed $\Sigma\Delta$ with $OSR=1024$

III. NUMERICAL EVALUATION OF $\Sigma\Delta$ LEARNING

We first verify the improvement achievable using $\Sigma\Delta$ learning as predicted by the equations (8). For this controlled experiment we have chosen two synthetic signals which have been also used in previous studies [24].

$$\begin{aligned} s_1(t) &= 480t - \lfloor 480t \rfloor \\ s_2(t) &= \sin(800t + 6 \cos(90t)) \end{aligned} \quad (33)$$

Each of these source signals were mixed using a random ill-conditioned matrix \mathbf{A} to obtain the two-dimensional signals which were then processed by the $\Sigma\Delta$ learner. The outputs of the $\Sigma\Delta$ learner were then used to reconstruct the source signals according to

$$\tilde{s}_d = \mathbf{A}^{-1} \mathbf{x} \quad (34)$$

$$\tilde{s}_m = \mathbf{A}^{-1} \mathbf{B}^{-1} \mathbf{x} \quad (35)$$

$$\tilde{s}_{m^+} = \mathbf{A}^{-1} \mathbf{B}^{-1} \Lambda^{-1} \mathbf{x} \quad (36)$$

assuming that the un-mixing matrix \mathbf{A}^{-1} can be perfectly determined.

The equations (34)- (36) represent the following three cases: (a) \tilde{s}_d which is the signal reconstructed using a $\Sigma\Delta$ modulator without any learning (denoted by *without*); (b) \tilde{s}_m which is the signal reconstructed using $\Sigma\Delta$ learning without resolution enhancement (denoted by *with*); and (c) \tilde{s}_{m^+} which is the signal reconstructed using $\Sigma\Delta$ learning with resolution enhancement (denoted by *with+*). For this experiment, the condition number of the mixing matrix was chosen to be 1000 and the oversampling ratio (OSR), which is defined as the sampling frequency/Nyquist frequency, was chosen to be 1024. For the signal in (33) the Nyquist frequency was chosen to be 10 KHz. Figure 7 shows the reconstructed signals obtained with and without the application of $\Sigma\Delta$ learning. The quantization artifacts can be clearly seen Fig. 7(b) which is the signal recovered using $\Sigma\Delta$ modulator without learning. However, the signals obtained when $\Sigma\Delta$ learning is applied does not show any such artifacts indicating improvement in resolution. To quantify this improvement, we compared the signal-to-error

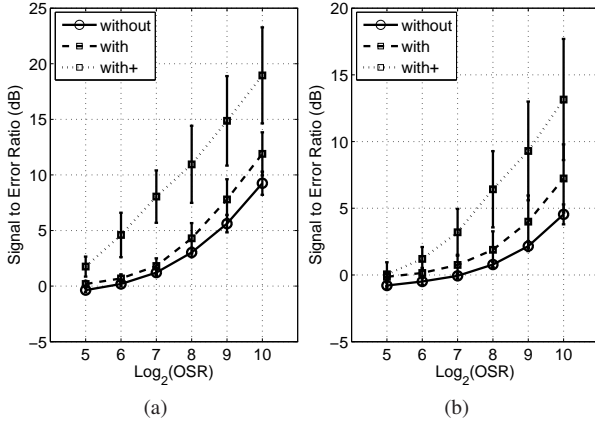


Fig. 8: Evaluating the reconstruction of the sources for classical (without), learning (with), and learning with resolution enhancement (with+) $\Sigma\Delta$ at different OSR for $\log_2(\text{condition number})$ of (a) 10 and (b) 12

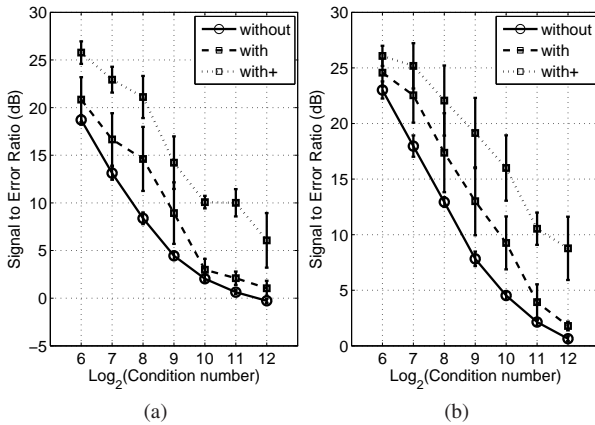


Fig. 9: Evaluating the reconstruction of the sources for classical (without), learning (with), and learning with resolution enhancement (with+) $\Sigma\Delta$ at different condition number for OSR of (a) 256 and (b) 512

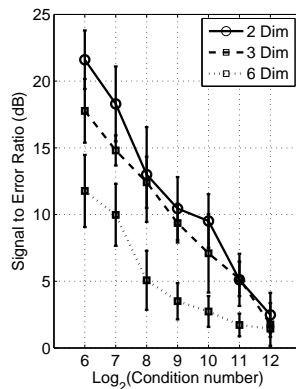


Fig. 10: Evaluating the reconstruction of sources at different dimension for the learning $\Sigma\Delta$ at different condition number for OSR of 128

ratio (SER) for the separated signals. SER is defined as

$$SER = \log_2 \left\{ \frac{\|s\|_2}{\|s - \tilde{s}\|_2} \right\} \quad (37)$$

where s and \tilde{s} is based on the definition in (34)- (36). To compute the mean SER and its variance 10 different mixing matrices with a fixed condition number were chosen and the mean/variances were calculated across different experimental runs. Fig. 8(a) compares the SER obtained when the mixing matrix with condition number 2^{10} was chosen for different values of OSR. Figure 8(b) compares the SER obtained when the mixing matrix with condition number 2^{12} was chosen. It can be seen in Fig. 8(a) and (b) that as the OSR of the $\Sigma\Delta$ modulator increases, the SER increases. This is consistent with results reported for $\Sigma\Delta$ modulators where OSR is directly related to the resolution of the ‘‘analog-to-digital’’ conversion. However, it can be seen that for all conditions of OSR, $\Sigma\Delta$ learning with resolution enhancement outperforms the other two approaches.

Figure 9(a) and (b) compares the performance of $\Sigma\Delta$ learner when the condition number of the mixing matrix is varied for fixed over-sampling ratios of 256 and 512. The results again show that $\Sigma\Delta$ learner (with and without resolution enhancement) demonstrates consistent performance improvement over the traditional $\Sigma\Delta$ modulator. Also, as expected the SER performance for all the three cases deteriorates with the increasing condition number, which indicates that the mixing becomes more singular. Figure 10 evaluates the SER achieved by $\Sigma\Delta$ learning (with resolution enhancement) when the dimensionality of the mixing matrix is increased. For this experiment, the number of source signals are increased by randomly selecting signals which were mutually independent with respect to each other. It can be seen from Fig. 10 that the response of the $\Sigma\Delta$ learning is consistent across different signal dimensions with larger SER when the dimension is lower.

In the next set of experiments the performance of $\Sigma\Delta$ learning is evaluated for the task of source separation when the un-mixing matrix is estimated using an ICA algorithm. Speech samples were obtained from TIMIT database and were synthetically mixed using an ill-conditioned matrix with different condition number. The instantaneous mixing parameters are shown in Fig. 11 which simulates the ‘‘near-far’’ scenario where one of the speech sources is assumed to be much closer to the microphone array than the other. This scenario was emulated by scaling one of the signals by $-50dB$ with respect to another, as shown in Fig. 11. The speech mixture is then presented to the $\Sigma\Delta$ learner and its output is then processed by a second-order blind inference (SOBI) [25] and by an efficient FastICA (EFICA) [26] algorithms. The performance metrics chosen for this experiment is based on source-to-distortion ratio (SDR) [27] where the estimated signal $\hat{s}_j(n)$ is decomposed into

$$\hat{s}_j(n) = s_{target}(n) + e_{interf}(n) + e_{artif}(n) \quad (38)$$

with $s_{target}(n)$ being the original signal, and $e_{interf}(n)$ and $e_{artif}(n)$ denote the interference and artifacts errors, respectively. The SDR metric is a global performance metric which

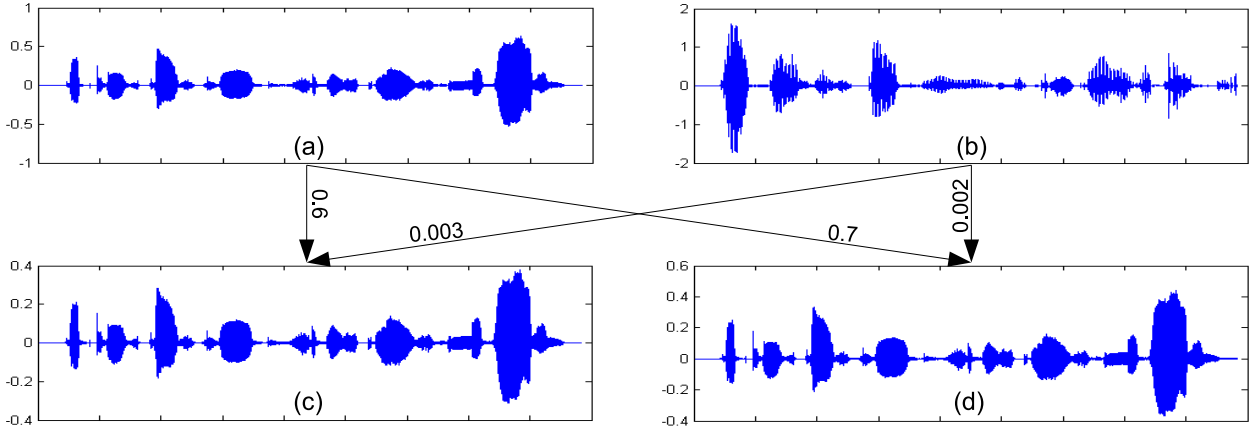


Fig. 11: Source signals s_1 and s_2 (subplots a and b) and mixed signals x_1 and x_2 (subplots c and d).

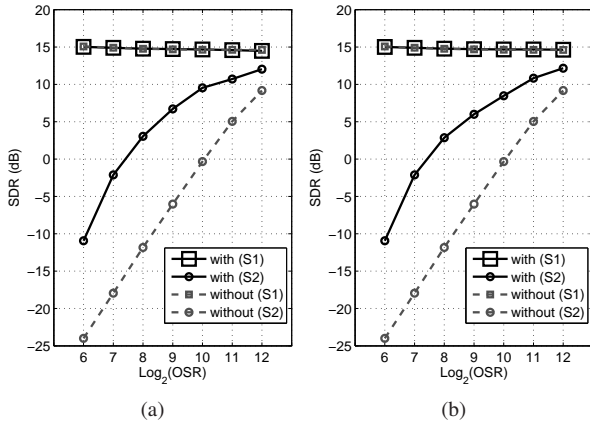


Fig. 12: SDR corresponding to with/without $\Sigma\Delta$ learning for the near-far recording conditions illustrated in Fig 11 using (a) SOBI and (b) EFICA algorithms.

then measures both source-to-interference ratio (SIR) - the amount of interferences from non wanted sources and also other artifacts like quantization and musical noise. The SDR is defined as:

$$SDR = 10 \log \frac{\|s_{target}\|^2}{\|e_{interf} + e_{artif}\|^2} \quad (39)$$

The speech sources s_1 and s_2 in Fig. 11 consists of 44200 samples with a sampling rate of 16KHz which is also the Nyquist rate. Figure 11 also shows that, after mixing, one of the sources is being completely masked by the other which is consistent with the “near-far” effect. Figure 12(a) and (b) shows the SDR obtained using the $\Sigma\Delta$ learning when the OSR is varied from 128 to 4096. Also shown in Fig. 12 (a) and (b) are the SDR metrics obtained when a conventional $\Sigma\Delta$ algorithm is used. It can be seen from the Fig. 12 that the SDR corresponding to the stronger source is similar for both cases (with and without $\Sigma\Delta$ learning), where as for the masked source the SDR obtained using $\Sigma\Delta$ learning is superior. This is consistent with the results published in [28]. However, the approach in [28] is applied after quantization

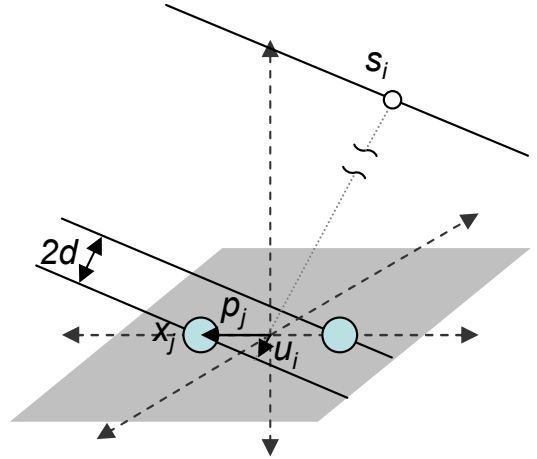


Fig. 13: Far-field recording using a two-element sensor array

and hence according to formulation in section I, is limited by the condition number of the mixing matrix. It should also be noted that the $\Sigma\Delta$ learning only enhances the resolution of the measured signals. The ability to successfully recover the weak source under “near-far” conditions, however, is mainly determined by the choice of the ICA algorithm.

IV. FAR-FIELD EXPERIMENTS USING MINIATURE MICROPHONE ARRAYS

In this section we first present a mathematical model for a miniature microphone array and show that the recorded signals from the array can be near singular. Then, we use this model with the $\Sigma\Delta$ learner to compare the performance of the algorithm with a traditional source separation technique. We then verify the proposed approach using speech data collected using a prototype miniature microphone array under real recording conditions.

A. Far-field model for a miniature microphone array

We illustrate a specific instance of singular mixing that occurs in the case of miniature microphone arrays. For this

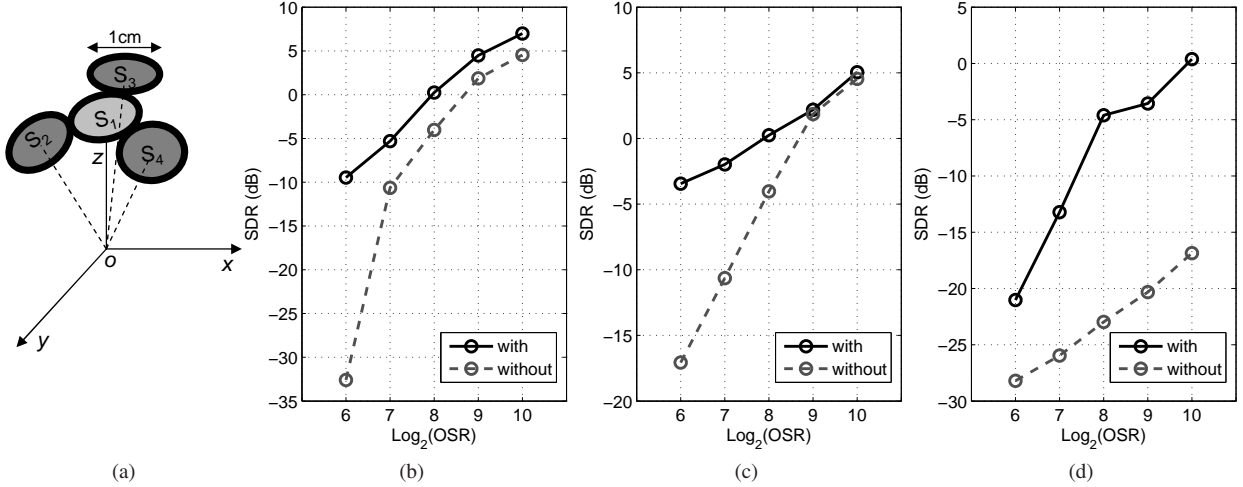


Fig. 14: $\Sigma\Delta$ performance with and without learning for three speech signals corresponding to the far-field model

modeling we resort to the far-field wave propagation models that have been extensively studied within the context of plenacoustic signal processing [17]. For modeling purposes, consider a sensor array shown in Fig. 13 that consists of two recording elements. If the inter-element distance is much less than the wavelength of the sensor signal of interest, the signals recorded at each of the sensor elements can be approximated using far-field models. Also, for miniature sensor array the distance to the sources from the center of the array can be assumed to be larger than the inter-element distance. We express the signal $x_j(\mathbf{p}_j, t)$ recorded at j^{th} sensor as a superposition of i independent sources $s_i(t)$ ($i \in 1, \dots, D$), each of which are referenced with respect to the center of the array [17]. This can be written as

$$x_j(\mathbf{p}_j, t) = \sum_i c_i(\mathbf{p}_j) s_i(t - \tau_i(\mathbf{p}_j)) \quad (40)$$

where $c_i(\mathbf{p}_j)$ and $\tau_i(\mathbf{p}_j)$ denotes the attenuation and delay, for the source $s_i(t)$ at the position \mathbf{p}_j , measured relative to the center of the sensor array. \mathbf{p}_j in equation (40) denotes the position vector of the j^{th} microphone. Equation (40) can be approximated using Taylor's series expansion as

$$x_j(\mathbf{p}_j, t) = \sum_i c_i(\mathbf{p}_j) \sum_{k=0}^{\infty} \frac{(-\tau_i(\mathbf{p}_j))^k}{k!} s_i^{(k)}(t) \quad (41)$$

Under far-field conditions it can be assumed that $c_i(\mathbf{p}_j) \approx c_i$ is constant across all the sensor elements. Also, the higher-order terms in the series expansion (41) can be ignored and can be expressed as

$$x_j(\mathbf{p}_j, t) \approx \sum_i c_i s_i(t) - \sum_i c_i \tau_i(\mathbf{p}_j) \dot{s}_i(t). \quad (42)$$

The component $x_c(t) = \sum_i c_i s_i(t)$ signifies the common-mode signal common to all the recording elements and the second part of the RHS signifies an instantaneous mixture of the derivative of the source signals. The common-mode component can be canceled using a differential measurement [18]

under which equation (42) becomes

$$\Delta x_j(\mathbf{p}_j, t) = x_j(\mathbf{p}_j, t) - x_c(t) \approx - \sum_i c_i \tau_i(\mathbf{p}_j) \dot{s}_i(t) \quad (43)$$

and can be expressed in a matrix-vector form as

$$\Delta \mathbf{x}(t) \approx -\mathbf{A} \dot{\mathbf{s}}(t) \quad (44)$$

where $\mathbf{A} = \{c_i \tau_i(\mathbf{p}_j)\}$ denotes the instantaneous mixing matrix. Under far-field approximation, the time delays can be expressed as

$$\tau_i(\mathbf{p}_j) = \mathbf{u}_i^T \mathbf{p}_j / v \quad (45)$$

where \mathbf{u}_i is the unit normal vector of the wavefront of source i and v is the speed of sound ($v = 340\text{m/s}$ in air). Thus equations (44) and (45) show that for miniature recording array, recovery of the desired sources \mathbf{s} or $\dot{\mathbf{s}}$ entails solving a linear source separation problem [19]. However, equation (45) reveals that sources that are located closer to the sensor array can completely mask the sources located away from the sensor array, resulting in a near-singular mixing \mathbf{A} . As shown in the following description that under near-singular mixing conditions, conventional methods of signal acquisition and source separation fail to deliver robust performance.

B. Experiments with Far-field Model

In this setup, we simulated recording conditions that consisted of four closely spaced microphones. The arrangement is shown in Fig. 14(a) where each circle represents a microphone of dimension 1cm. In this arrangement, three of the microphones were placed along a triangle, whereas the fourth microphone was placed at the centroid and act as the reference sensor which records the common signal. The set up is similar to the conditions that have been reported in [19] where the simulation have been shown to be consistent with recording in the real-life conditions. The outputs of each microphone along the triangle were subtracted from the reference microphone to produce three differential outputs. In our experiments we used three independent speech signals as

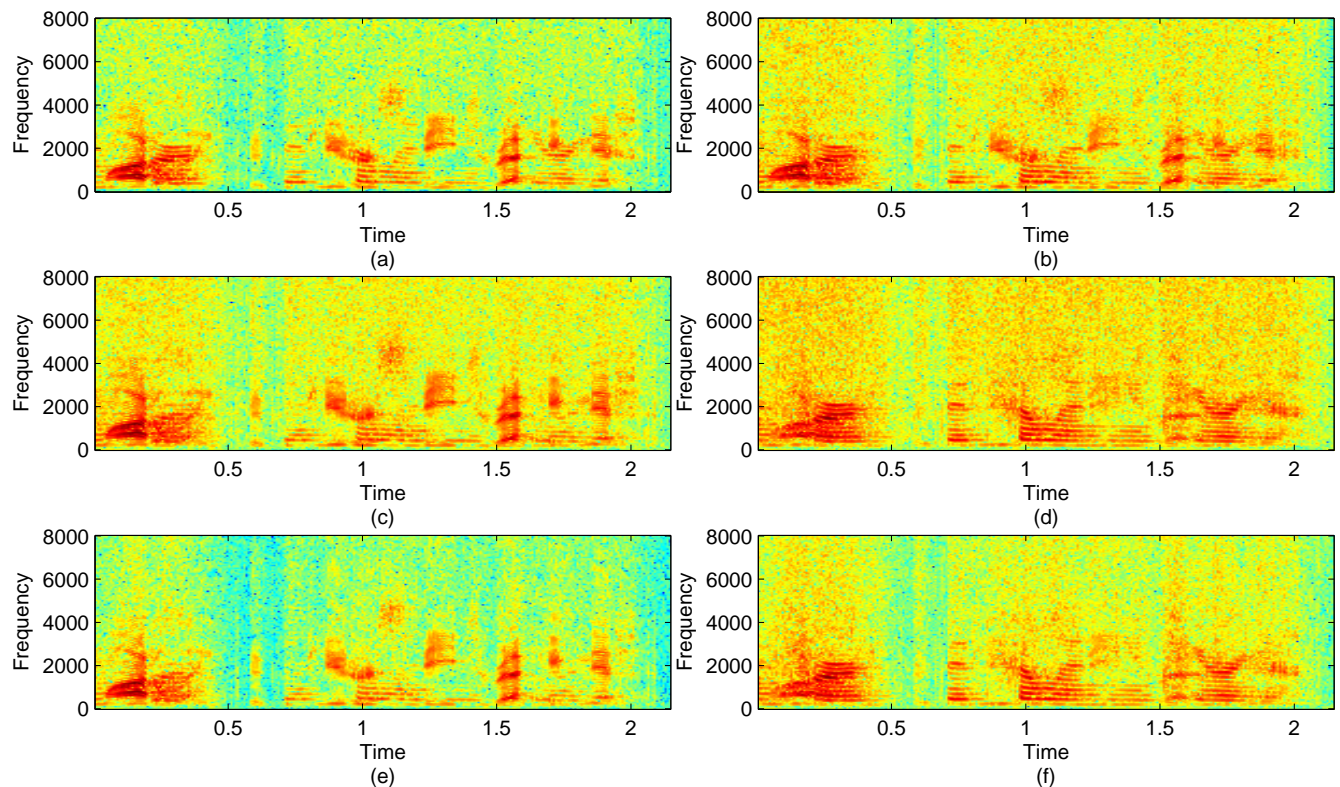


Fig. 15: Spectrogram of the recorded signals (subplots (a) and (b)) and recovered signals using $\Sigma\Delta$ without learning (subplots (c) and (d)) and with learning (subplots (e) and (f))

far-field sources. The differential outputs of the microphone array were first presented to the proposed $\Sigma\Delta$ learner, and the outputs of $\Sigma\Delta$ learner array were then used as inputs to the SOBI algorithm. A benchmark used for comparative study consisted of $\Sigma\Delta$ converters that directly quantized the differential outputs of the microphones. Figure 14(b)-(d) summarizes the performance of source separation (with and without $\Sigma\Delta$ learning) for different orientation of the acoustic sources. For the three experiments, only the bearing of the sources were varied but their respective distances to the center of the microphone was kept constant. It can be seen that for each of these orientations, source separation algorithm that uses $\Sigma\Delta$ learning as a front-end “analog-to-digital” converter delivers superior performance compared to the algorithm that does not use $\Sigma\Delta$ learning. Also, from Fig. 14(b)-(d) it can be seen that the improvement in SDR performance significantly increases when the sampling frequency (resolution) decreases showing that $\Sigma\Delta$ learning efficiently utilizes the available resolution (due to coarse quantization) to capture information necessary for source separation.

C. Experiments with real microphone recordings

In this set of experiments, we have applied $\Sigma\Delta$ learning to speech data that was recorded using a prototype miniature microphone array, similar to the set up described in [18]. The four omni-directional microphones *Knowles FG3629* were mounted on a circular array with radius 0.5 cm. The differential microphone signals were pre-processed using second-order bandpass filters with low-frequency cutoff at 130 Hz

and high-frequency cutoff at 4.3 kHz. The signals were also amplified by 26dB. The speech signals were presented through loudspeakers positioned at 1.5 m distance from the array and the sampling frequency of the National Instruments data acquisition card was set to 32 KHz. Male and female speakers from TIMIT database were chosen as sound sources and was replayed through standard computer speakers. The data was recorded from each of the microphones, archived, and then presented as inputs to the $\Sigma\Delta$ learning and the SOBI algorithm. Figure 15(a) and (b) shows the spectrogram of the speech signals recorded from the microphone array. The two spectrograms look similar, thus emulating a “near-far” recording scenario where a dominant source masks the background weak source. Also it can be seen from the spectrogram in Fig. 15(b) that one of the recordings is more noisy than the other (due to microphone mismatch). Figure 15(c) and (d) show the spectrogram of the separated speech signals obtained without $\Sigma\Delta$ learning. Figure 15(e) and (f) show the spectrogram of the separated speech signals obtained with $\Sigma\Delta$ learning. A visual comparison of the spectrograms show that separated speech signal (without $\Sigma\Delta$ learning) contain more quantization artifacts which can be seen as the broadband noise in Fig. 15(c) and (d). The table I summarizes the SDR performance (for different OSR) for each of the sources in these two cases (with and without $\Sigma\Delta$ learning), showing that $\Sigma\Delta$ learning indeed improves the performance of the source separation algorithm. Also, from table I, it can be seen that when the OSR increases, the performance differences between

TABLE I: Performance (SDR (dB)) of the proposed $\Sigma\Delta$ for the real data for different over-sampling ratio.

	OSR=4		OSR=8		OSR=16		OSR=32		OSR=64	
	with	without	with	without	with	without	with	without	with	without
S_1	1.03	-0.72	1.33	0.41	1.34	0.94	1.3	1.15	1.28	1.21
S_2	-12.52	-13.15	-9.77	-10.17	-8.69	-8.88	-8.29	-8.32	-8.12	-8.1

the two cases becomes insignificant. This artifact is due to the limitations in the SOBI algorithm for separating sources with high fidelity, noise in the microphones and ambient recording conditions.

V. CONCLUSIONS AND DISCUSSIONS

In this paper, we have argued that the classical approach of signal quantization followed by DSP based source separation fails to deliver robust performance when processing signals recorded using miniature sensor arrays. We proposed a framework that combined statistical learning with $\Sigma\Delta$ modulation and can be used for designing “smart” high-dimensional analog-to-digital converters that can exploit spatial correlations to resolve acute differences between signals recorded by miniature sensor array. Even though in this paper $\Sigma\Delta$ learning has been demonstrated to improve the performance of speech based source separation algorithms, the proposed technique is general and can be applied to any sensor array. The potential applications include microphone array hearing aids, microelectrode array in neuroprosthetic devices, miniature radio-frequency antenna arrays and for radar applications. The experimental results presented in this paper demonstrated that the $\Sigma\Delta$ learning algorithm consistently led to a superior performance over the classical approach and hence is attractive for implementing high-dimensional analog-to-digital converters. The future work in this area will include: (a) exploring higher-order noise-shaping $\Sigma\Delta$ modulators for improving the performance of resolution enhancement; (b) extending $\Sigma\Delta$ learning to non-linear signal transforms by embedding kernels into the optimization framework; (c) extending $\Sigma\Delta$ learning to integrate ICA with signal quantization.

ACKNOWLEDGMENT

The authors would like to thank the Prof. Milutin Stancevic from Stony Brook University for providing us with the real microphone data to which we have applied the $\Sigma\Delta$ learning algorithm. This work was supported in part by a grant from the National Science Foundation (IIS:0836278).

REFERENCES

- [1] M.J. Madou, M.J., *Fundamentals of Microfabrication: The Science of Miniaturization*, Boca Raton, FL: CRC Press, 2002.
- [2] K.D. Wise, D.J. Anderson, J.F. Hetke, D.R. Kipke, and K. Najafi, “Wireless implantable microsystems: High-density electronic interfaces to the nervous system,” *Proceedings of the IEEE*, vol. 92, no. 1, pp. 138-145, Jan. 2004.
- [3] C.T. Nordhausen, E.M. Maynard, and R.A. Normann, “Single unit recording capabilities of a 100-microelectrode array,” *Brain Research*, vol. 726, pp. 129-140, 1996.
- [4] K.D. Wise and K. Najafi, “Microfabrication techniques for integrated sensors and microsystems,” *Science*, vol. 254, pp. 1335-1342, Nov. 1991.
- [5] R.N. Miles and R.R. Hoy, “The development of a biologically-inspired directional microphone for hearing aids,” *Audiology and Neuro-Otology*, vol. 11, no. 2, pp. 86-94, 2006.
- [6] J.E. Greenberg and P.M. Zurek, “Microphone array hearing aids in microphone arrays: Signal Processing Techniques and Application,” *Springer*, Berlin, pp. 229-253, 2001.
- [7] B. Edwards, “The future of hearing aid technology,” *Trends in Amplification*, vol. 11, no. 1, pp. 31-45, Mar. 2007.
- [8] C. Jutten and J. Héroult, “Blind separation of sources, Part 1: an adaptive algorithm based on neuromimetic architecture,” *Signal Processing*, vol. 24, no. 1, pp. 1-10, July 1991.
- [9] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, New York: Wiley, 2001.
- [10] A. Hyvärinen and E. Oja, “Independent component analysis: Algorithms and applications,” *Neural Networks*, vol. 13, pp. 411-430, 2000.
- [11] P. Comon, “Independent component analysis - a new concept?,” *Signal Processing*, vol. 36, pp. 287-314, 1994.
- [12] A.J. Bell and T.J. Sejnowski, “An Information-Maximization Approach to Blind Separation and Blind Deconvolution,” *Neural Computation*, vol. 7, pp. 1129-1159, 1995.
- [13] S. Amari, A. Cichocki, and H.H. Yang, “A new learning algorithm for blind signal separation,” in *Adv. Neural Information Processing Systems (NIPS)*, vol. 8, pp. 757-763, Cambridge: MIT Press, 1996.
- [14] J.-F. Cardoso and B.H. Laheld, “Equivariant adaptive source separation,” *IEEE Trans. Signal Process.*, vol. 44, no. 12, pp. 3017-3030, 1996.
- [15] E. Oja, “Principal components, minor components, and linear neural networks,” *Neural Networks*, vol. 5, no. 6, pp.927-935, 1992.
- [16] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, New York, NY: Wiley, 2002.
- [17] M.N. Do, “Toward sound-based synthesis: the far-field case,” in *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Canada, 2004.
- [18] A. Celik, M. Stanacevic, and G. Cauwenberghs, “Gradient flow independent component analysis in micropower VLSI,” *Adv. Neural Information Processing Systems (NIPS)*, vol. 8, pp. 187-194, Cambridge: MIT Press, 2006.
- [19] J. Barrère, and G. Chabriel, “A Compact Sensor Array for Blind Separation of Sources,” *IEEE Trans. Circuits and Systems: Part I*, vol. 49, no. 5, pp. 565-574, 2002.
- [20] J.C. Candy and G.C. Temes, “Oversampled methods for A/D and D/A conversion,” in *Oversampled Delta-Sigma Data Converters*, Piscataway, NJ: IEEE Press, pp. 1-29, 1992.
- [21] L. Bottou, “Stochastic learning,” in *Advanced Lectures on Machine Learning, Lecture Notes in Artificial Intelligence*, vol. 3176, O. Bousquet and U. von Luxburg, Ed. Berlin: Springer Verlag, 2004, pp. 146-168.
- [22] A. Fazel, S. Chakrabarty, “Sigma-Delta resolution enhancement for far-field acoustic source separation,” in *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, NV, 2008.
- [23] W. Rudin, *Functional Analysis*, New York: McGraw-Hill, 1991.
- [24] A. Cichocki, R.E. Bogner, L. Moczynski, and K. Pope, “Modified Hrault-Jutten algorithms for blind separation of sources,” *Digital Signal Processing*, vol. 7, no. 2, pp. 80-93, 1997.
- [25] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, E. Moulines, “A blind source separation technique using second-order statistics,” *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 4344-444, 1997.
- [26] Z. Koldovský, P. Tichavský and E. Oja, “Efficient Variant of Algorithm FastICA for Independent Component Analysis Attaining the Cramér-Rao Lower Bound,” *IEEE Trans. on Neural Networks*, Vol. 17, No. 5, Sept 2006.
- [27] E. Vincent, R. Gribonval, and C. Fvotte, “Performance measurement in Blind Audio Source Separation,” *IEEE Trans. Audio, Speech, and Lang. Processing*, vol. 14, no. 4, pp. 1462-1469, July 2006.
- [28] M. Gupta and S. C. Douglas, “Performance Evaluation of Convolutional Blind Source Separation of Mixtures of Unequal-Level Speech Signals,” in *Proc. Int. Symposium on Circuits and Systems (ISCAS)*, New Orleans, Louisiana, 2007.



Amin Fazel (S'07) received the B.Sc. degree in computer science and engineering from Shiraz University, Shiraz, Iran, in 2002 and the M.Sc. degree in computer engineering from Sharif University of Technology, Tehran, Iran, in 2005. Currently, he is pursuing the Ph.D. degree in electrical and computer engineering at Michigan State University. His research interests include speech processing, robust speech/speaker recognition, acoustic source separation, and analog-to-information converters.



Amit Gore received the B.E. degree in instrumentation engineering from the University of Pune, Pune, India, in 1998, and the M.S. and Ph.D. degrees in electrical and computer engineering from Michigan State University, East Lansing, in 2002 and 2008, respectively. Currently, he is with General Electric Global Research, Niskayuna, New York. His research interests are low-power sigma-delta converters, analog signal processing and low-power mixed-signal VLSI design.



Shantanu Chakrabarty (M'96-SM'09) received his B.Tech degree from Indian Institute of Technology, Delhi in 1996, M.S and Ph.D in Electrical Engineering from Johns Hopkins University, Baltimore, MD in 2001 and 2004 respectively. He is currently an assistant professor in the department of electrical and computer engineering at Michigan State University. From 1996-1999 he was with Qualcomm Incorporated, San Diego and during 2002 he was a visiting researcher at University of Tokyo. His current research interests include low-power analog and digital VLSI systems, hardware implementation of machine learning algorithms with application to biosensors and biomedical instrumentation. Dr. Chakrabarty was a recipient of The Catalyst foundation fellowship from 1999-2004 and received the best undergraduate thesis award in 1996. He is currently a member for IEEE BioCAS technical committee, IEEE Circuits and Systems Sensors technical committee and serves as an associate editor for Advances in Artificial Neural Systems from Hindawi publications.