Using Signal Invariants to Perform Nonlinear Blind Source Separation

David N. Levin

University of Chicago
d-levin@uchicago.edu

Abstract. Given a time series of multicomponent measurements $x(t)$, the usual objective of nonlinear blind source separation (BSS) is to find a "source" time series $s(t)$, comprised of statistically independent combinations of the measured components. In this paper, the source time series is required to have a density function in $(s, \dot{s})$-space that is equal to the product of density functions of individual components. This formulation of the BSS problem has a solution that is unique, up to permutations and component-wise transformations. Separability is shown to impose constraints on certain locally invariant (scalar) functions of x, which are derived from local higher-order correlations of the data's velocity $\dot{x}$. The data are separable if and only if they satisfy these constraints, and, if the constraints are satisfied, the sources can be explicitly constructed from the data. The method is illustrated by using it to separate two speech-like sounds recorded with a single microphone.

1 Introduction

Consider a time series of data $x(t)$, where x is a multiplet of N measurements (x_k for $k = 1, 2, \ldots, N$). The usual objectives of nonlinear BSS are: 1) determine if these data are instantaneous mixtures of N statistically independent source components $s(t)$

$$x(t) = f[s(t)], \tag{1}$$

where f is a possibly nonlinear, invertible N-component mixing function, and, if this is the case, 2) compute the mixing function. In other words, the problem is to find a coordinate transformation f^{-1} that transforms the observed data $x(t)$ from the measurement-defined coordinate system (x) on state space to a source coordinate system (s) in which the components of the transformed data are statistically independent. In the source coordinate system, let the state space probability density function (PDF) $\rho_S(s)$ be defined so that $\rho_S(s)ds$ is the fraction of total time that the source trajectory $s(t)$ is located within the volume element ds at location s. In the usual formulation of the BSS problem [1], the source components are required to be statistically independent in the sense that their state space PDF is the product of the density functions of the individual components

$$\rho_S(s) = \prod_{k=1}^{N} \rho_k(s_k), \tag{2}$$

However, it is well known that this criterion is so weak that this form of the BSS problem always has many solutions (see [2] and references therein).

The issue of non-uniqueness can be circumvented by considering the data's trajectory in $(s, \dot{s})$-space instead of state space (i.e., s-space). First, define the PDF $\rho_S(s, \dot{s})$ in this space so that $\rho_S(s, \dot{s})dsd\dot{s}$ is the fraction of total time that the location and velocity of the source trajectory are within the volume element $dsd\dot{s}$ at location $(s, \dot{s})$. Earlier papers [3, 4] described a formulation of the BSS problem in which this PDF was required to be the product of the density functions of the individual components

$$\rho_S(s, \dot{s}) = \prod_{k=1}^{N} \rho_k(s_k, \dot{s}_k). \quad (3)$$

This type of statistical independence is satisfied by almost all classical non-interacting physical systems. Furthermore, because separability in $(s, \dot{s})$-space is a stronger requirement than separability in state space, the corresponding BSS problem can be shown to have a unique solution [3, 4].

It was previously demonstrated [3, 4] that the $(s, \dot{s})$-space PDF of a time series induces a Riemannian geometry on the data's state space, with the metric equal to the local second-order correlation matrix of the data's velocity. Nonlinear BSS can be performed by computing this metric in the x coordinate system (i.e., by computing the second-order correlation of $\dot{x}(t)$ at point x), as well as its first and second derivatives with respect to x. However, if the dimensionality of state space is high, it must be covered by a great deal of data in order to calculate these derivatives accurately. The current paper shows how to perform nonlinear BSS by computing higher-order local correlations of the data's velocity, instead of computing derivatives of its second-order correlation. This approach is advantageous because it requires less data for an accurate computation.

In addition to using a stronger criterion for statistical independence, there are technical differences between the proposed method and conventional ones. First of all, the technique in this paper exploits statistical constraints on the data that are *locally* defined in state space, in contrast to the usual criteria for statistical independence that are *global* conditions on the data time series or its time derivatives [5]. Furthermore, unlike many other methods [6], the mixing function is constructed in a "deterministic", non-parametric manner, without using probabilistic learning methods and without parameterizing it with a neural network architecture or other means.

The next section describes the theoretical framework of the method. Section 3 illustrates the method by using it to separate two simultaneous speech-like sounds that are recorded with a single microphone. The implications of this work are discussed in the last section.

2 Method

The local correlation of the data's velocity is

$$C_{kl\ldots}(x) = < (\dot{x}_k - \bar{\dot{x}}_k)(\dot{x}_l - \bar{\dot{x}}_l) \ldots >_x, \quad (4)$$

where $\bar{\dot{x}} = < \dot{x} >_x$ is the local time average of $\dot{x}$, where the bracket denotes the time average over the trajectory's segments in a small neighborhood of x, where $1 \leq k, l \leq N$, and where "..." denotes possible additional indices on the left side and corresponding factors of $\dot{x} - \bar{\dot{x}}$ on the right side. The definition of the PDF implies that this velocity correlation is one of its moments

$$C_{kl\ldots}(x) = \frac{\int \rho(x, \dot{x})(\dot{x}_k - \bar{\dot{x}}_k)(\dot{x}_l - \bar{\dot{x}}_l)\ldots d\dot{x}}{\int \rho(x, \dot{x}) d\dot{x}}, \tag{5}$$

where $\rho(x, \dot{x})$ is the PDF in the x coordinate system. Incidentally, although Eq. (5) is useful in a formal sense, in practical applications, all required correlation functions can be computed directly from local time averages of the data (Eq. (4)), without explicitly computing the data's PDF. Also, note that the analogous velocity "correlation" with a single subscript vanishes identically.

Next, let $M(x)$ be a local $N\,x\,N$ matrix, and use it to define transformed velocity correlations

$$I_{kl\ldots}(x) = \sum_{1 \leq k', l', \ldots \leq N} M_{kk'}(x) M_{ll'}(x) \ldots C_{k'l'\ldots}(x), \tag{6}$$

where "..." denotes possible additional indices of I and C, as well as corresponding factors of $M(x)$. Because $C_{kl}(x)$ is positive definite at any point x, it is always possible to find an $M(x)$ such that

$$I_{kl}(x) = \delta_{kl} \tag{7}$$

$$\sum_{1 \leq m \leq N} I_{klmm}(x) = D_{kl}(x), \tag{8}$$

where $D(x)$ is a diagonal $N\,x\,N$ matrix. As long as D is not degenerate, $M(x)$ is unique, up to arbitrary *local* permutations and reflections. In almost all realistic applications, the velocity correlations will be continuous functions of the state space coordinate x. Therefore, in any neighborhood of state space, there will always be a continuous solution for $M(x)$, and this solution is unique, up to arbitrary *global* reflections and permutations.

Now, imagine doing the same computation in some other coordinate system x'. An M-matrix that satisfies Eqs. (7, 8) in the x' coordinate system is given by

$$M'_{kl}(x') = \sum_{1 \leq m \leq N} M_{km}(x) \frac{\partial x_m}{\partial x'_l}(x'). \tag{9}$$

This can be understood in the following manner: because velocity correlations transform as contravariant tensors, the partial derivative factor transforms correlations from the x' coordinate system to the x coordinate system, and the factor $M(x)$ then transforms these correlations into the functions on the left sides of Eqs. (7, 8). All other solutions for $M'(x')$ differ from this one by global reflections and permutations. Similar reasoning shows that the functions $I'_{kl\ldots}(x')$, derived by using Eq. (6) in the x' coordinate system, equal the functions $I_{kl\ldots}(x)$, up to possible global permutations and reflections. In other words,

$$I'_{kl\ldots}(x') = \sum_{1 \leq k', l', \ldots \leq N} P_{kk'} P_{ll'} \ldots I_{k'l'\ldots}(x), \tag{10}$$

where $P_{kk'}$ denotes an element of a product of permutation, reflection, and identity matrices. Essentially, the functions $I_{kl\ldots}(x)$ transform as scalar functions on the state space, except for possible reflections and index permutations.

We now assume that the system is separable and derive some necessary conditions in the source coordinate system (s). Then, the above-described scalar functions are used to transfer these separability conditions to the measurement-defined coordinate system (x), where they can be tested with the data. In order to make the notation simple, it is assumed that $N = 2$. However, as described below, the methodology can be generalized in order to separate higher-dimensional data into possibly multidimensional source variables.

Separability implies that there is a transformation f^{-1} from the x coordinate system to a source coordinate system (s) in which Eq. (3) is true. Because of Eq. (5), the velocity correlation functions in the s coordinate system are products of correlations of independent sources

$$C_{S1\ldots2\ldots}(s_1, s_2) = C_{S1\ldots}(s_1)\, C_{S2\ldots}(s_2), \tag{11}$$

where $1\ldots$ and $2\ldots$ denote arbitrary numbers of indices equal to 1 and 2, respectively. It follows that, in the s coordinate system, Eqs. (7, 8) are satisfied by a block-diagonal matrix M, in which each "block" is the $1\,x\,1$ M "matrix" satisfying Eqs. (7, 8) for one of the subsystems. Therefore, in the s coordinate system, the functions $I_{Skl\ldots}(s)$, which are defined by Eq. (6) with all subscripts $kl\ldots$ equal to 1 (2), must equal the corresponding functions derived for subsystem 1 (2), and these latter functions depend on s_1 (s_2) alone. Although these constraints were derived in the s coordinate system, Eq. (10) implies that they are true in all coordinate systems, except for possible permutations and reflections. Therefore, in the measurement-defined coordinate system (x), the functions $I_{kl\ldots}(x)$, which are defined by Eq. (6) with all subscripts equal to 1, must be functions of either $s_1(x)$ *or* $s_2(x)$. Likewise, the functions $I_{kl\ldots}(x)$, which are defined by Eq. (6) with all subscripts equal to 2, must be functions of the other source variable ($s_2(x)$ or $s_1(x)$, respectively).

This coordinate-system-independent consequence of separability can be used to perform nonlinear BSS in the following manner:

1. Use Eq. (4) to compute velocity correlations, $C_{kl\ldots}(x)$, from the data $x(t)$.
2. Use linear algebra to find a continuous matrix $M(x)$ that satisfies Eqs (7, 8) (or that satisfies similar algebraic constraints that determine $M(x)$ uniquely, except for permutations and reflections).
3. Use Eq. (6) to compute the functions $I_{kl\ldots}(x)$.
4. Plot the values of the triplets

$$I_A(x) = \{I_{111}(x),\, I_{1111}(x),\, I_{11111}(x)\} \tag{12}$$

$$I_B(x) = \{I_{222}(x),\ I_{2222}(x),\ I_{22222}(x)\}, \tag{13}$$

as x varies over the measurement-defined coordinate system.

5. If the plotted values of I_A and/or I_B *do not* lie in one-dimensional subspaces within the three-dimensional space of the plots, $I_A(x)$ and/or $I_B(x)$ cannot be functions of single source components ($s_1(x)$ or $s_2(x)$) as required by separability, and the data are not separable.
6. If the plotted values of both I_A and I_B *do* lie on one-dimensional manifolds, define one-dimensional coordinates (σ_A and σ_B, respectively) on those subspaces [7]. Then, compute the function $\sigma(x) = (\sigma_A(x), \sigma_B(x))$ that maps each coordinate x onto the value of σ, which parameterizes the point $(I_A(x), I_B(x))$. Notice that, because of the Takens' embedding theorem [8], x is invertibly related to the six components of $I_A(x)$ and $I_B(x)$, and, therefore, it is invertibly related to σ.
7. Transform the PDF (or correlations) of the measurements from the x coordinate system to the σ coordinate system. The data are separable if and only if the PDF factorizes (the correlations factorize) in the σ coordinate system.

The last statement can be understood in the following manner. As shown above, separability implies that $I_A(x)$ must be a function of a single source variable ($s_1(x)$ or $s_2(x)$), and the Takens theorem implies that this function is invertible. Because I_A is also an invertible function of σ_A, it follows that σ_A must be invertibly related to one of the source variables, and, in a similar manner, σ_B must be invertibly related to the other source variable. Thus, separability implies that σ_A and σ_B are themselves source variables. It follows that the data are separable if and only if the PDF factorizes in the σ coordinate system.

This procedure can be generalized to determine whether data with $N > 2$ are a mixture of two, possibly multidimensional source variables. Consider any partition of the x indices ($k = 1, 2, \ldots, N$) into two groups: d_A "A" indices and $d_B = N - d_A$ "B" indices. Let $I_A(x)$ ($I_B(x)$) be any set of more than $2d_A$ ($2d_B$) of the functions $I_{kl\ldots}(x)$ for which all subscripts belong to the A (B) group. Now, suppose that the data are separable. Then, there must be such a partition for which the values of $I_A(x)$ ($I_B(x)$) lie in a d_A-dimensional (d_B-dimensional) subspace, as x varies over the N-dimensional data space. Furthermore, if σ_A and σ_B are some coordinates on those subspaces, a set of source variables is given by $\sigma(x) = (\sigma_A(x), \sigma_B(x))$, the values of σ_A and σ_B that parameterize the point $(I_A(x),\ \ I_B(x))$. Therefore, to perform BSS: 1) systematically examine all possible index partitions and determine if the data-derived functions, $I_A(x)$ and $I_B(x)$, map x onto subspaces with the required dimensions; 2) if they do comprise such maps, construct the function $\sigma(x)$ and determine if the data's PDF factorizes in the σ coordinate system. The data are separable if and only if at least one such index partition leads to a $\sigma(x)$ that factorizes the PDF. If the data are separable, the same procedure can then be used to determine if each multicomponent independent variable (σ_A or σ_B) can be further separated into lower-dimensional independent variables.

3 Numerical Example: Separating Two Speech-Like Sounds Recorded with a Single Microphone

This section describes a numerical experiment in which two speech-like sounds were synthesized and then summed, as if they were simultaneously recorded with a single microphone. Each sound simulated an "utterance" of a vocal tract resembling a human vocal tract, except that it had fewer degrees of freedom (one degree of freedom instead of the 3-5 degrees of freedom of the human vocal tract). The methodology of Section 2 was blindly applied to a time series of two features extracted from the synthetic recording, in order to recover the time dependence of the state variable of each vocal tract (up to an unknown transformation on each voice's state space). BSS was performed with only 16 minutes of data, instead of the hours of data required to separate similar sounds using a differential geometric method [3, 4].

The glottal waveforms of the two "voices" had different pitches (100 Hz and 160 Hz), and the "vocal tract" response of each voice was characterized by a damped sinusoid, whose amplitude, frequency, and damping were linear functions of that voice's state variable. For each voice, a 16 minute utterance was produced by using glottal impulses to drive the vocal tract's response, which was determined by the time-dependent state variable of that vocal tract. The state variable time series of each voice was synthesized by smoothly interpolating among successive states, which were randomly chosen at 100-120 msec intervals. The resulting utterances had energies differing by 2.4 dB, and they were summed and sampled at 16 kHz with 16-bit depth. Then, this "recorded" waveform was pre-emphasized and subjected to a short-term Fourier transform (using frames with 25 msec length and 5 msec spacing). The log energies of a bank of 12 mel-frequency filters between 0-8000 Hz were computed for each frame, and these were then averaged over pairs of consecutive frames. These log filterbank outputs were nonlinear functions of the two vocal tract state variables, which were chosen to be statistically independent of each other.

In order to blindly analyze these data, we first determined if any data components were redundant in the sense that they were simply functions of other components. Figure 1a shows the first three principal components of the log filterbank outputs during a typical short recording of the simultaneous utterances. Inspection showed that these data lay on an approximately two-dimensional surface within the ambient 12-D space, making it apparent that they were produced by a "hidden" dynamical system with two degrees of freedom. The redundant components were eliminated by using dimensional reduction (principal components analysis in small overlapping neighborhoods of the data) to establish a coordinate system x on this surface and to find $x(t)$, the trajectory of the recorded sound in that coordinate system. The next step was to determine if $x(t)$ was a nonlinear mixture of two source variables that were statistically independent of one another. Following steps 1-4 of the BSS procedure in Section 2, $x(t)$ of the entire recording was used to compute "invariants" $I_{kl\ldots}(x)$ with up to five indices, and the related functions $I_A(x)$ and $I_B(x)$ were plotted, as illustrated in Fig. 1b. It was evident that the plotted values of both I_A and I_B did lie in

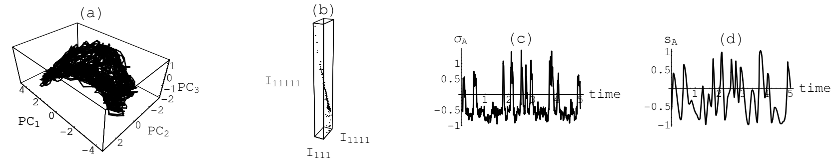

Fig. 1. (a) The first three principal components of log filterbank outputs of a typical short recording of two simultaneous speech-like sounds. (b) The distribution of the values of $I_A(x)$ (Eq. (12)), as x varied over the approximately two-dimensional manifold in (a). (c) The time dependence of one of the source variables, blindly computed from a typical five-second segment of the data's trajectory $x(t)$, by finding the coordinates of $I_A[x(t)]$ on the approximately one-dimensional manifold in (b). (d) The state variable time series originally used to generate one of the speech-like sounds during the five-second recording analyzed in (c).

approximately one-dimensional subspaces. Following step 6 of the BSS procedure, a dimensional reduction procedure [7] was used to define coordinates (σ_A and σ_B) on these one-dimensional manifolds, and a numerical representation of $\sigma(x) = (\sigma_A(x), \sigma_B(x))$ was derived. If the data were separable, σ must be a set of source variables, and $\sigma[x(t)]$ must describe the evolution of the underlying vocal tract states (up to invertible component-wise transformations). As illustrated in Figs. 1c-d, the time courses of the putative source variables $(\sigma_A[x(t)], \sigma_B[x(t)])$ did resemble the time courses of the state variables that were originally used to generate the voices' utterances (up to an invertible transformation on each state variable space). Thus, it was apparent that the information encoded in the time series of each vocal tract's state variable was blindly extracted from the simulated recording of the superposed utterances.

4 Discussion

In previous papers [3, 4], the nonlinear BSS problem was formulated in (state, state velocity)-space, instead of state space as in conventional formulations. This approach is attractive because: 1) this type of statistical independence is satisfied by almost all classical non-interacting physical systems; 2) this form of the BSS problem has a unique solution in the following sense: either the data are inseparable, or they can be separated by a mixing function that is unique, up to permutations and transformations of independent source variables. This paper shows how to perform this type of nonlinear BSS by computing local higher-order correlations of the data's velocity $\dot{x}(t)$, instead of computing derivatives of its local second-order correlation as was previously proposed [3, 4]. This is advantageous because it requires less data for an accurate computation, as demonstrated in a numerical example in which BSS was performed with minutes (instead of hours) of data.

The BSS procedure in Section 2 shows how to compute $(\sigma_A[x(t)], \sigma_B[x(t)])$, the trajectory of each independent subsystem in a specific coordinate system on that subsystem's state space. In many practical applications, a pattern recognition "engine" has been trained to recognize the meaning of trajectories of one subsystem (e.g., "A") in another coordinate system (e.g., s_A) on that subsystem's state space. In order to use this information, it is necessary to know the transformation to this coordinate system ($\sigma_A \rightarrow s_A$). For example, subsystem A may be the vocal tract of speaker A, and subsystem B may be a noise generator of some sort. In this example, we may have trained an automatic speech recognition (ASR) engine on the quiet speech of speaker A (or, equivalently, on the quiet speech of another speaker who mimics A in the sense that their state space trajectories are related by an invertible transformation when they speak the same utterances). In order to recognize the speaker's utterances in the presence of B, we must know the transformation from the vocal tract coordinates provided by BSS (σ_A) to the coordinates used to train the ASR engine (s_A). This mapping can be determined by using the training data to compute more than $2d_A$ invariants (like those in Eq. (6)) as functions of s_A. These must equal the invariants of one of the subsystems identified by the BSS procedure, up to a global permutation and/or reflection (Eq. (10)). This global transformation can be determined by permuting and reflecting the distribution of invariants produced by the training data, until it matches the distribution of invariants of one of the subsystems produced by the BSS procedure. Then, the mapping $\sigma_A \rightarrow s_A$ can be determined by finding paired values of σ_A and s_A that correspond to the same invariant values within these matching distributions. This type of analysis of human speech data is currently underway.

References

1. Jutten, C., Karhunen, J.: Advances in Nonlinear Blind Source Separation. In: Proceedings of the 4^{th} International Symposium on Independent Component Analysis and Blind Signal Separation, Nara, Japan, April 2003
2. Hyvärinen, A., Pajunen, P.: Nonlinear Independent Component Analysis: Existence and Uniqueness Results. Neural Networks 12, 429-439 (1999)
3. Levin, D. N.: Using State Space Differential Geometry for Nonlinear Blind Source Separation. In: Davies, M. E., James, C. J., Abdallah, S. A., Plumbley, M. D. (eds) ICA 2007. LNCS, vol. 4666, pp. 65-72. Springer, Heidelberg (2007)
4. Levin, D. N.: Using State Space Differential Geometry for Nonlinear Blind Source Separation. J. Applied Physics 103, art. no. 044906 (2008)
5. Lagrange, S., Jaulin, L., Vigneron, V., Jutten, C.: Analytic Solution of the Blind Source Separation Problem Using Derivatives. In: Puntonet, C. G., Prieto, A. G. (eds) ICA 2004. LNCS, vol. 3195, pp. 81-88. Springer, Heidelberg (2004)
6. Yang, H. H., Amari, S-I., Cichocki, A.: Information-Theoretic Approach to Blind Separation of Sources in Non-linear Mixture. Signal Processing 64, 291-300 (1998)
7. Roweis, S. T., Saul, L. K.: Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science 290, 2323-2326 (2000)
8. Sauer, T., Yorke, J. A., Casdagli, M.: Embedology. J. Statistical Physics 65, 579-616 (1991)