Multichannel Blind Deconvolution using Low Rank Recovery

Justin Romberg[a], Ning Tian[a], and Karim Sabra[b]

[a] School of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta, GA, 30332

[b] School of Mechanical Engineering
Georgia Institute of Technology, Atlanta, GA, 30332

ABSTRACT

We introduce a new algorithm for multichannel blind deconvolution. Given the outputs of K linear time-invariant channels driven by a common source, we wish to recover their impulse responses without knowledge of the source signal. Abstractly, this problem amounts to finding a solution to an overdetermined system of quadratic equations. We show how we can recast the problem as solving a system of underdetermined linear equations with a rank constraint. Recent results in the area of low rank recovery have shown that there are effective convex relaxations to problems of this type that are also scalable computationally, allowing us to recover 100s of channel responses after a moderate observation time. We illustrate the effectiveness of our methodology with a numerical simulation of a passive "noise imaging" experiment.

Keywords: channel estimation, blind deconvolution, passive imaging, low rank recovery

1. INTRODUCTION

This paper introduces a new framework for multichannel blind deconvolution. The problem is illustrated in Figure 1. A common *source signal* $s(t)$ drives K different channels with impulse responses $h_1(t), h_2(t), \ldots, h_K(t)$. We observe samples of the outputs $y_1(t), \ldots, y_K(t)$ of these channels over a period of time, and from these observations we wish to estimate the $h_k(t)$. Without knowing anything about the source signal $s(t)$ or the channel responses $h_k(t)$, this problem does not have a unique solution. We will show (using numerical experiments) that if we make some structural assumptions about the channels, namely that they live in known subspaces, then this problem can become well-posed. Moreover, there is a fast, scalable algorithm for solving it.

One of the challenging aspects of the multichannel blind deconvolution problem is that it essentially amounts to solving a system of *quadratic equations* — we are treating both the source $s(t)$ and the channels $h_k(t)$ as unknowns, and our observations consist of linear combinations of entires of s and the h_k multiplied by one another. To make these statements precise, we will write down the inverse problem using the language of linear algebra. We will assume that either the $s(t)$ or the $h_k(t)$ are bandlimited, and so recovering the samples $h_k[n]$ of the $h_k(t)$, spaced at the corresponding Nyquist rate or closer, is the same as recovering the responses themselves (limited to this frequency band). We have the general observational model

$$y_k[\ell] = \sum_{n=-\infty}^{\infty} h_k[n]s[\ell - n], \qquad \text{for } \ell = 0, 1, \ldots, L-1. \tag{1}$$

It should be clear (and it is indeed well known) that without any structural assumptions on s or h_k, there can be many sources and channel responses which explain the observations $y_k[\ell]$. Even if we were to observe an infinite number of samples $\{y_k[\ell],\ \ell \in \mathbb{Z}\}$ over an infinite amount of time, the h_k and s are still not identifiable. In this case, (1) can be written in the frequency domain as $\hat{y}(\omega) = \hat{s}(\omega) \cdot \hat{h}_k(\omega)$, where $\hat{s}(\omega)$ and $\hat{h}_k(\omega)$ are the discrete-time Fourier transforms of $s[n]$ and $h_k[n]$. We are left with the problem of identifying two periodic functions $\hat{s}(\omega)$ and $\hat{h}_k(\omega)$ on the interval $-\pi \leq \omega \leq \pi$ from their multiplication. Knowing nothing else, this is hopeless, and the situation does not improve if we treat the channels jointly.

E-mail: jrom@ece.gatech.edu, ningtian@gatech.edu, karim.sabra@me.gatech.edu. This work was partially supported by ONR Grant N00014-11-1-0459 and a Packard Fellowship.

Independent Component Analyses, Compressive Sampling, Wavelets, Neural Net, Biosystems, and Nanoengineering XI,
edited by Harold H. Szu, Proc. of SPIE Vol. 8750, 87500E · © 2013 SPIE
CCC code: 0277-786X/13/$18 · doi: 10.1117/12.2018550

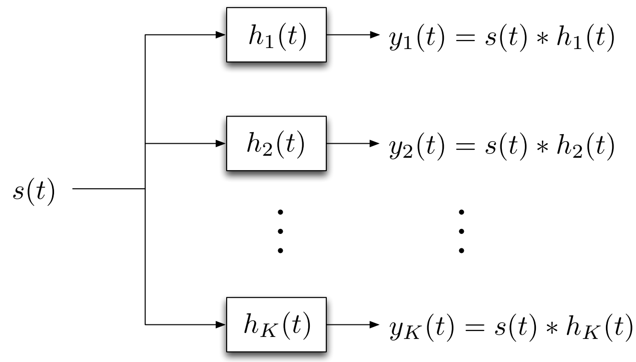


Figure 1: *Problem illustration. An unknown common source drives K linear time-invariant channels with different impulse responses $h_1, \ldots, h_K$. From the outputs $y_1, \ldots, y_K$ of these channels we wish to determine both $s(t)$ and the $h_k(t)$.*

The story does improve, however, if we make even mild assumptions about the channel responses $h_k[n]$. To begin, let us assume that the $h_k[n]$ are time-limited in that $h[n]$ can only be nonzero for $0 \leq n \leq N-1$. We can now limit the sum in the observational model (1):

$$y_k[\ell] = \sum_{n=0}^{N-1} h_k[n]s[\ell - n], \qquad \text{for } \ell = 0, 1, \ldots, L-1. \tag{2}$$

There are $N + (L + N - 1) = L + 2N - 1$ variables involved in (2), and L quadratic equations which are combining these variables in different ways: each sample of y_k involves a sum over products of different combinations of $\{h_k[0], \ldots, h_k[N-1]\}$ and $\{s[-N+1], \ldots, s[0], \ldots, s[L-1]\}$. Treating the channels jointly adds more equations and more unknowns. But since the channels are all being driven by a common source, the number of unknowns increases more slowly than the number of equations as more channels are considered. The system of equations corresponding to the observations $\{y_k[\ell],\ \ell = 0, \ldots, L-1;\ k = 1, \ldots, K\}$ has KL equations in $KN + (L+N-1) = (K+1)N + L - 1$ unknowns. Thus the number of equations we have can in fact be much larger than the number of unknowns for $L > \text{Const} \cdot N$ when $K \geq 2$.

Simply comparing the number of equations to the number of unknowns does not tell the whole story, especially since the equations are quadratic. The observations above are simply meant to show that in some loose sense as the number of observations grows, so does the ratio of the information we have versus the information we do not. In the next section, we take another step towards formalizing this trade off by recasting the quadratic equations in (2) as a system of linear equations with a rank constraint.

2. SOLVING LARGE-SCALE SYSTEMS OF QUADRATIC EQUATIONS

Let $\mathbf{s} \in \mathbb{R}^{L+N-1}$ contain the samples of the common source that are used in (2), and let $\mathbf{h}_1, \ldots, \mathbf{h}_K \in \mathbb{R}^N$ be vectors containing the nonzero elements of the channel responses:

$$\mathbf{s} = \begin{bmatrix} s[-N+1] \\ \vdots \\ s[-1] \\ s[0] \\ s[1] \\ \vdots \\ s[L-1] \end{bmatrix}, \qquad \mathbf{h}_k = \begin{bmatrix} h_k[0] \\ h_k[1] \\ \vdots \\ h_k[N-1] \end{bmatrix}.$$

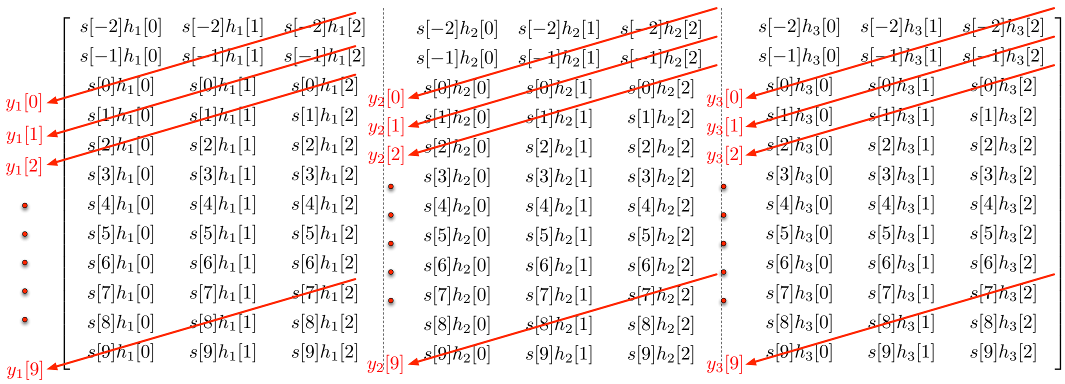


Figure 2: *The matrix above is* $\mathbf{X}_0$ *in* (3) *with a number of measurements per channel of* $L = 9$*, channel length* $N = 3$*, and number of channels* $K = 3$*. Each observation* $y_k[\ell]$ *is a sum along one of the skew diagonals of the a submatrix of* $\mathbf{X}_0$*, as illustrated by the red lines above. In general, it is impossible to recover a* $L + N - 1 \times NK$ *matrix from measurements of this type, but explicitly incorporating the fact that* $\mathbf{X}_0$ *has rank 1 into the recovery makes this possible for appropriate values of* L, N, K*.*

We can arrange all pairs of variables appearing in the sum in (2) in a large $(L + N - 1) \times NK$ matrix $\mathbf{X}_0$:

$$\mathbf{X}_0 = \begin{bmatrix} \mathbf{s} \end{bmatrix} \begin{bmatrix} \mathbf{h}_1^{\mathrm{T}} & \mathbf{h}_2^{\mathrm{T}} & \cdots & \mathbf{h}_K^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} \mathbf{s}\mathbf{h}_1^{\mathrm{T}} & \mathbf{s}\mathbf{h}_2^{\mathrm{T}} & \cdots & \mathbf{s}\mathbf{h}_K^{\mathrm{T}} \end{bmatrix}. \tag{3}$$

Each observation $y_k[\ell]$ is now a linear combination of the entries in $\mathbf{X}_0$ — this is shown more precisely in Figure 2. Concatenating the observations from each channel into a single vector $\mathbf{y} \in \mathbb{R}^{KL}$, we have a linear system of equations

$$\mathbf{y} = \mathcal{A}(\mathbf{X}_0), \tag{4}$$

where in this case, $\mathcal{A}$ takes sums over skew diagonals of the submatrices $\mathbf{s}\mathbf{h}_k^{\mathrm{T}}$.

We can now treat the recovery of the channel responses $\mathbf{h}_k$ and the source $\mathbf{s}$ as a *matrix recovery* problem. The vector $\mathbf{y}$ contains KL different linear combinations of entries of a $(L+N-1) \times NK$ matrix. No matter what L, N, and K are, there will always be fewer equations than entries in the unknown matrix, meaning that the system of linear equations in (4) is underdetermined. But we know that the underlying matrix $\mathbf{X}_0$ has special structure — since it can be written as the outer product of two vectors (as in (3)), its rank is one. To recover $\mathbf{X}_0$ from $\mathbf{y}$, then, we might search for a $L + N - 1 \times NK$ matrix that satisfies

$$\mathcal{A}(\mathbf{X}) = \mathbf{y}, \qquad \operatorname{rank}(\mathbf{X}) = 1.$$

For an arbitrary $\mathbf{y}$, there may or may not be a point that satisfies the constraints above. Given $\mathbf{y}$, we might search for such a feasible point by solving

$$\min_{\mathbf{X}} \ \operatorname{rank}(\mathbf{X}) \quad \text{subject to} \quad \mathcal{A}(\mathbf{X}) = \mathbf{y}. \tag{5}$$

While the rank minimization program above is difficult to solve directly, it has a natural and very effective convex relaxation.[1] We simply replace $\operatorname{rank}(\mathbf{X})$ with the *nuclear norm* $\|\mathbf{X}\|_*$, which is the sum of the singular values of $\mathbf{X}$. We can then solve

$$\min_{\mathbf{X}} \ \|\mathbf{X}\|_* \quad \text{subject to} \quad \mathcal{A}(\mathbf{X}) = \mathbf{y}. \tag{6}$$

In the past five years, there has been significant progress in our understanding of when the solutions to (6) and (5) agree. We mention here one seminal result. If $\mathbf{X}_0$ is a $N_1 \times N_2$ matrix with rank R, and $\mathcal{A}$ is a linear operator that returns M *random* linear combinations of all the entries in $\mathbf{X}_0$ with weights chosen independently from a normal distribution*, then (6) will (with appropriately high probability) recover $\mathbf{X}_0$ exactly when[2,3]

$$M \gtrsim R(N_1 + N_2). \tag{7}$$

That is, from a number of random measurements within a constant of the number of degrees of freedom of a rank R matrix, we can recover the matrix using convex programming.

This low rank recovery result can be interpreted in the context of solving systems of quadratic equations as follows. In a manner similar to that described earlier in this section, we can write a system of quadratic equations in variables $v_1, \ldots, v_{N_1}$ as a linear operator acting on $\mathbf{v}\mathbf{v}^{\mathrm{T}}$. Then (7) tells us that "most" slightly overdetermined systems of quadratic equations (a number of equations that is a constant times the number of variables) specified by $(\mathcal{A}, \mathbf{y})$ which have a feasible solution can be solved using (6).

Of course, in the multichannel blind deconvolution problem, the measurements are completely deterministic (as illustrated in Figure 2), meaning that the analytic techniques used for the majority of research in low-rank recovery cannot be applied directly to derive results for this problem. There is, however, a natural way to inject randomness into this problem. In so-called *noise imaging* applications,[4,5] $s(t)$ is a random process. This means that while the system $\mathcal{A}$ is deterministic, $\mathbf{X}_0$ is random. A first set of numerical simulations using this model are presented in Section 4. Deriving analytic bounds on L versus K and N that guarantee that we can untangle the s and the h_i is a topic of current research.

3. RELATED QUADRATIC SYSTEMS

The general method of taking an overdetermined system of quadratic equations and recasting it as a system of linear equations with a rank constraint has been studied in other contexts in the recent literature. In particular, the problem of phase retrieval[6] can be approached in precisely this manner. The general problem considered in that work is to reconstruct a signal $x \in \mathbb{R}^L$ from the *magnitudes* of a series of linear measurements $|\langle a_\ell, x\rangle|$. It was shown that an arbitrary vector in $\mathbb{R}^L$ can be recovered (with high probability) from $\sim L \log L$ such measurements if the a_ℓ are chosen randomly.

Another recent work[7] looks at solving quadratic equations using low-rank recovery in the context of single channel blind deconvolution. In that work, the general problem considered is to recover two unknown signals $x, w \in \mathbb{R}^L$ from their convolution $y = x * w$. The structural assumption made on the signals is that they live in known *subspaces* of $\mathbb{R}^L$ of dimension N and K, respectively. This linear constraints can easily be incorporated into the low-rank recovery problem, and it is shown that if one of the signals lives in a subspace which is *incoherent* with the Fourier domain, and the other signal lives in a *generic* (i.e. randomly chosen) subspace, then they can be untangled when $(N + K) \sim L/\log^3 L$. That is, the sum of the dimensions of the subspaces is within a logarithmic factor of the dimension of the ambient space.

In the present work, we are not making any structural assumptions on the source signal $s(t)$ other than that it is bandlimited so we can discretize it by sampling; in the numerical experiments in the next section, we takes these samples to be a sequence of independent Gaussian random variables. Roughly speaking, we can do without structure on the source since it is involved in every measurement we make.

We do not present any theoretical results here for multichannel blind deconvolution using low rank recovery. Rather we will demonstrate its effectiveness in the next section through two illustrative numerical examples.

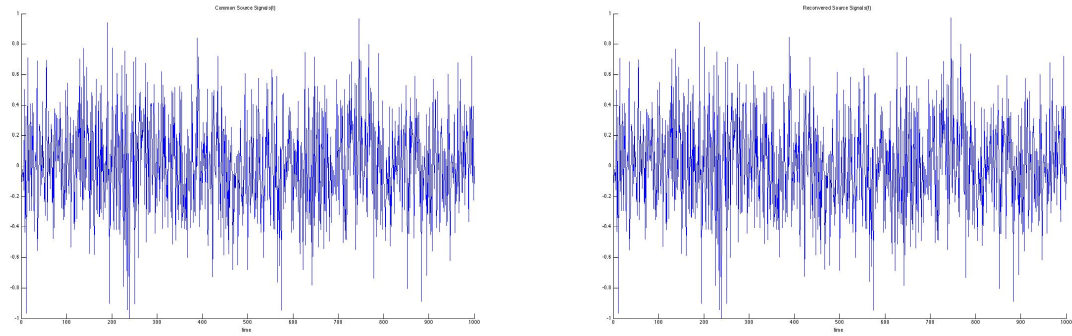
4. NUMERICAL RESULTS

We implemented this multichannel blind deconvolution framework on two different types of channel structure to show preliminary numerical results. In our experiments, the common driving source signal $\mathbf{s} \in \mathbb{R}^{L+N-1}, (L + N - 1 = 1000)$ is completely unstructured (Figure 3a); it is simply a sequence of independent and identically distributed Gaussian random variables. This type of model for the source signal is not uncommon in passive imaging.[4] We assume two different types of channel structure, which we will describe below. In both cases, the

*There are many types of distributions from which the weights might be chosen;[2] the result generalizes in a straightforward manner as long as they are independent.

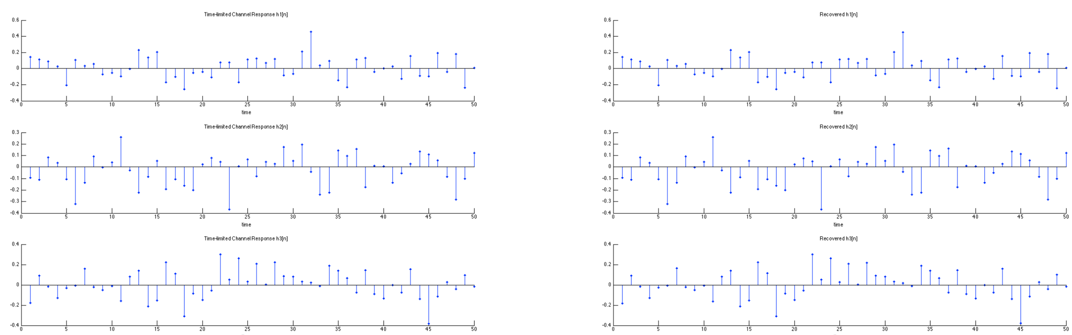
recovery algorithm solves a variant[8] of (6) that explicitly takes into account the fact that we expect the solution to have very small rank, allowing the algorithm to scale to moderate vales of N, L, and K.

In our first experiment, we assume that the channel impulse responses $\mathbf{h}_1, \ldots, \mathbf{h}_K \in \mathbb{R}^N$ are "short" (time-limited) in that $\mathbf{h}_k[n]$ is only nonzero for $0 \leq n \leq N-1$. This leads to the measurement model illustrated in Figure 2. Figure 3 shows the a typical recovery result for a channel length of $N = 50$ and $K = 100$ channels with a source length of 1000. Even though the channels are being driven with unstructured noise, both the source signals and all of the channel responses (6000 variables total) are recovered with relative error within two digits from the 100 observed outputs ($\approx 100,000$ total samples).



(a) Common Source Signal

(b) Recovered Source Signal

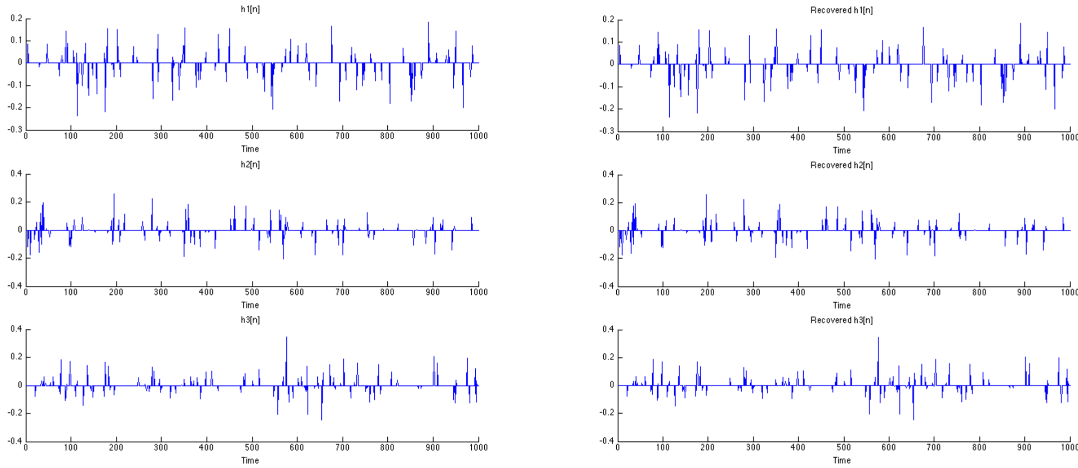


(c) Time-limited Channel Impulse Response

(d) Recovered Channel Response

Figure 3: *Time-limited channel response recovery. (a) The common source signal* $s[n]$ *is white Gaussian noise. We observe the convolution of this source signal with* $K = 100$ *different channels that have* $N = 50$ *taps. (b) Recovered source signal, the relative error is* 1.128^{-2}. *(c) The true underlying impulse responses of three of the channels. (d) Recovered response of the same three channels. The total relative error across all 100 channels in this case is* 1.046×10^{-2}.

In our second experiment, we choose the channels to be *sparse* instead of short. That is, instead of having N components all concentrated in locations in $[0, N-1]$, they are spread out at (known) locations $n_{k,1}, n_{k,2}, \ldots, n_{k,N}$ — in this case, we will let there be at most $N = 100$ different delays, and these delays can be different from channel to channel. (We pause to stress that in this particular experiment, we are assuming we know the delays, but not the responses at the delays. Jointly solving for the delays and the response coefficients is a current topic of research.) Figure 4 shows the (partial) results of at $K = 20$ channel system being recovered along with a source input of length 1000. (In this case, we recover 2000 total variables, 1000 for the source and 50 for each of the 20 channels, from $20,000$ total samples.) Again, we are able to perform the deconvolution very accurately, achieving a relative accuracy of almost 4 digits.



(a) (b)

Figure 4: *Sparse (with known support) channel recovery. The simulated system has 20 channels, each channel has* $N = 100$ *non-zeros. (a) Impulse response of three of the 20 channels used in the experiment. (b) From the outputs of the 20 channels, we are able to recover the channel responses and the source signal exactly. The total relative error over the 20 recovered channels is* 1.261×10^{-4}.

REFERENCES

[1] M. Fazel, *Matrix rank minimization with applications.* PhD thesis, Stanford University, March 2002.

[2] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review* **52**(3), pp. 471–501, 2010.

[3] E. Candès and Y. Plan, "Tight oracle bounds for low-rank matrix recovery from a minimal number of random measurements," *IEEE Trans. Inform. Theory* **57**(4), pp. 2342—2359, 2011.

[4] K. Sabra and D. Dowling, "Blind deconvolution in oceanic waveguides using artificial time reversal," *J. Acoust. Soc. Am.* **116**, pp. 262–271, 2004.

[5] R. Snieder and K. Wapenaar, "Imaging with ambient noise," *Physics Today* , pp. 44–49, September 2010.

[6] E. Candès, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *To appear in Comm. Pure Appl. Math.* , 2012.

[7] A. Ahmed, B. Recht, and J. Romberg, "Blind deconvolution using convex programming," *Submited to IEEE Trans. Inform. Theory* , November 2012.

[8] S. Burer and R. Monteiro, "A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization," *Math. Program., Ser. B* **95**, pp. 329–357, 2003.