Resolving Scaling Ambiguities with the ℓ_1/ℓ_2 Norm in a Blind Deconvolution Problem with Feedback

Ernie Esser[#†], Rongrong Wang[*], Tim T.Y. Lin[†], and Felix J. Herrmann[†]

[†]Department of Earth, Ocean and Atmospheric Sciences
University of British Columbia
2020-2207 Main Mall
Vancouver, BC V6T 1Z4 , Canada

[*]Department of Mathematics
University of British Columbia
1984 Mathematics Rd
Vancouver, BC V6T 1Z2, Canada

*

November 26, 2015

Abstract

Compared to more mundane blind deconvolution problems, blind deconvolution in seismic applications involves a feedback mechanism related to the free surface. The presence of this feedback mechanism gives us an unique opportunity to remove ambiguities that have plagued blind deconvolution for a long time. While beneficial, this feedback by itself is insufficient to remove the ambiguities even with ℓ_1 constraints. However, when paired with an ℓ_1/ℓ_2 constraint the feedback allows us to resolve the scaling ambiguity under relatively mild assumptions. Inspired by lifting approaches, we propose to split the sparse signal into positive and negative components and apply an ℓ_1/ℓ_2 constraint to the difference, thereby obtaining a constraint that is easy to implement. Numerical experiments demonstrate robustness to the initialization as well as to noise in the data.

1 Introduction

Deconvolution problems often arise in physical measurements, where the experimental apparatus introduces a linear time-invariant transfer function from the true underlying signal to the measured data. This effect often degrades the fidelity of measurements, and counteracting it will necessitate deconvolving from the data a blurring kernel that describes the measurement system. Also, in many realistic scenarios the blurring kernel is imprecisely or not at all known a-priori, in which case the deconvolution is *blind*. Obtaining a useful solution in blind deconvolution often requires imposing tight bounds on the subspaces in which the kernel and the underlying signal lie, making the process susceptible to heavy influences from any prior assumptions on the signal model.

In addition to a linear transfer function, some measurement processes also incorporates a feedback mechanism, where the reflections of the signal off the recording apparatus causes further excitation the measured system. One example of this type of system is in active seismological surveys where both the artificial sources and the detectors are located at the Earth's surface, from which recorded seismic waves can be assumed to scatter perfectly back into the Earth as secondary sources. As a consequence, the observed data contains both primary reflections and its echoes (called *multiples* in the geophysical context). These echoes are essentially high-order moments (under convolution) of the underlying signal, which is a non-linear relation that is decoupled from the transfer function of the measurement system. Therefore we might consider whether the feedback effect imposes enough additional structure to the blind deconvolution problem in a way that can lessen the set of prior assumptions needed to constrain the signal model.

[*#]Ernie Esser (1980-2015), who passed away under tragic circumstances, was an extremely enthusiastic and talented researcher who was, above all, immensely generous with his time and ideas. This work is an attempt to share his work and is a reflection of what a true privilege it has been to have had the opportunity to work with Ernie.

This work will mainly consider how the feedback convolution model interacts with a common regularizer in deconvolution, in which a sparsity assumption is imposed on the underlying signal. Specifically, an ℓ_1-norm constraint or minimization formulation is popular due to a favourable mix of sparsifying efficacy and the wide availability of solution algorithms. We show that the amplitude sensitivity of ℓ_1 (or even $\ell_{p<1}$) norms do not lead to the resolution of the basic scaling ambiguity in blind deconvolution even with the feedback system, but more importantly that the unique combination of a feedback system and a scaling-insensitive ℓ_1/ℓ_2 norm ratio can actually have a local minimum at the correct scaling. We also propose an algorithm to solve this problem in a feasible way using lifting and the method of multipliers.

2 Theoretical result

2.1 Blind deconvolution

In many applications, we record data that are the convolution (represented by operator $\otimes$) of the signal of interest $x \in \mathbb{R}^N$ with some (unknown) blurring kernel $w \in \mathbb{R}^N$ plus additive noise $n \in \mathbb{R}^N$—i.e., we have

$$y(t) = w(t) \otimes x(t) + n(t). \tag{1}$$

Given full knowledge of this kernel, we can under certain conditions deconvolve y and retrieve x despite the fact that this inverse problem is ill-posed because w is band-limited. In this case, there will be frequencies θ such that $\hat{w}(\theta) = 0$. Hence, for noise-free data ($n = 0$) we have $\hat{y}(\theta) = \hat{w}(\theta)\hat{x}(\theta) = 0$. As a consequence, the missing bands of w turn (1) into an underdetermined system of equations.

When x is sparse, or permits some compressible representation (with fast decay for the magnitude-sorted coefficients), we can overcome this ill-posedness by imposing structure-promoting regularizations on x such as the ℓ_1-norm penalty—i.e., we solve the following program

$$\min_{x=Cz} \|y - w \otimes x\| + \lambda \|z\|_1,$$

where C is some sparsifying basis and λ some control parameter balancing the data validity and the ℓ_1-norm penalty.

While sparsity-promoting deconvolutions have been carried out with great success, these problems become drastically more challenging when the convolution kernel is unknown. In this case, the problem has two unknowns and is known as blind deconvolution, which suffers from scaling and shift ambiguities—i.e. if the pair $(\hat{w}, \hat{x})$ is a solution of (1) so is $\left(\alpha\hat{w}(t-\beta), \frac{1}{\alpha}\hat{x}(t+\beta)\right)$ for any $0 \neq \alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$. Moreover, it has been shown by [1] that minimizing the data misfit with an additional ℓ_1-norm penalty on x and a shift-invariant norm penalty on w never returns the true solution unless x is a delta function. Possible remedies include imposing support constraints on w and x, or using other regularizers such as the ratio of ℓ_1 to ℓ_2 norm.

2.2 The ratio of ℓ_1 to ℓ_2 norm

Intuitively, the ℓ_1 norm is not minimized at the true solution because it is biased towards small scalings. As a remedy, we may put a similarly biased penalty on w to balance the scaling between the unknowns w and x, but this approach raises the question how to choose these relative weights. As recognized by [1, 7, 6, 3], the ratio between the ℓ_1 and ℓ_2-norms is scale invariant, thus is arguably more stable for the blind deconvolution problem.

2.3 Convolution systems with feedback

Mathematically, the expression of this feedback system (introduced in [9] in the geophysical context) in 1D reads

$$\begin{aligned} y &= w \otimes x - w \otimes x \otimes x + w \otimes x^{\otimes 3} - w \otimes x^{\otimes 4} + ... \\ &= w \otimes x - x \otimes y, \end{aligned} \tag{2}$$

where the alternating signs are related to the minus sign of the free-surface reflection coefficient, and $^{\otimes n}$ stands for order n auto-convolutions of x with itself.

We recognize the first term on the RHS as the convolution of the primary reflections with the wavelet while all other terms are echoes—i.e., multiples. Because these multiples can be wrongly identified as layers, one of the primary tasks in seismic imaging is to remove these multiples, which may overlay late primaries.

In the context of geophysics, many attempts have been made to solve the typical blind deconvolution problem with approaches that assume the underlying signal to be white and the kernel to be minimum phase, or by maximizing the kurtosis of the underlying signal. A more recent approach known as Estimation of Primaries by Sparse Inversion (EPSI) model used in ([8, 5]) aims to invert the feedback system (2) using a block coordinate-descent algorithm that imposes ℓ_1 regularization on the underlying signal. While this latter approach has been demonstrated to work successfully, it requires relatively ad-hoc, carefully-performed initialization and normalization steps to recover the correct relative scaling between x and w.

2.4 Resolving the scaling ambiguity of feedback system with l1/l2 norm

The physical feedback mechanism that leads to the system described in (2) removes the trivial scale ambiguity, which plagues the traditional blind deconvolution problem. However, as we state in the following theorem this scaling issue can never be truly resolved when imposing the ℓ_1-norm on the primary reflections. For simplicity of presentation, we work in the continuous domain and use $\hat{x}$, $\hat{w}$ to denote the continuous Fourier transform of x and w and $L^1(\mathbb{R}) := \{f : \int_t |f|dt < \infty\}$ represents the set of integrable functions of real variables.

Theorem 2.1. *Suppose $x, w, \hat{x}, \hat{w} \in L^1(\mathbb{R})$, and $\|x\|_1 < 1$ then the right-hand side of the following quantity exists and is integrable*

$$y := \sum_{i=1}^{\infty} (-1)^{i-1} w \otimes x^{\otimes i}. \tag{3}$$

Moreover, there exist a sequence $\alpha_k \in (1, \infty)$ and a sequence of functions $x_k \in L^1(\mathbb{R})$ such that

- $\alpha_k \to 1^+$,
- *$(\alpha_k w, x_k)$ is consistent with y (in the sense that $(\alpha_k w, x_k, y)$ satisfies (3)),*

and $\|x_k\|_1 < \|x\|_1$ for each k. In other words, the true solution (w, x) is not a local minimum to the optimization problem

$$\min_{\widetilde{w}, \widetilde{x}} \|\widetilde{x}\|_1 \text{ subject to } y = \sum_{i=1}^{\infty} (-1)^{i-1} \widetilde{w} \otimes \widetilde{x}^{\otimes i} \tag{4}$$

Remark 1. The constraint $\|x\|_1 < 1$ prevents the energy from diverging when performing higher and higher order convolutions on the RHS of (3).

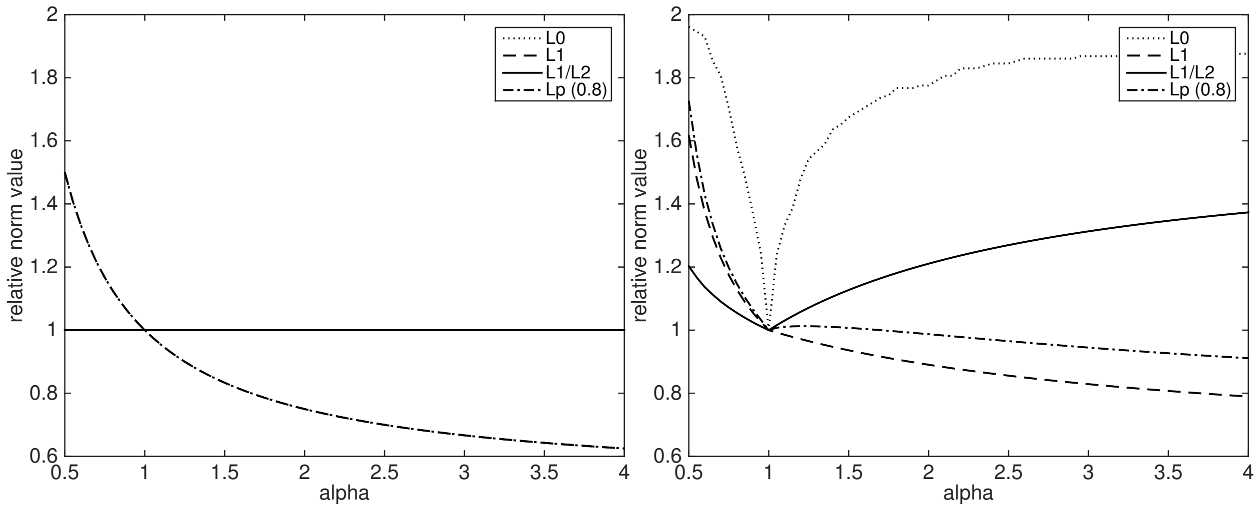
Remark 2. For any signal x, we can find a consistent $\widetilde{x}$ with a smaller ℓ_1 norm according to this theorem. Therefore the ℓ_1 penalty still exhibits bias towards small scaling as in the non-feedback case. Moreover, since αw has the same support as w, then we can not expect any support assumptions on w to prevent α from going into the wrong direction.

Now let us look at the non-convex ℓ_1/ℓ_2-norm and whether its scale invariance properties can provide a better measure to overcome the scaling ambiguity to a certain degree. (In the expressions below, supp(x) stands for the support of x.)

Theorem 2.2. *Assume $supp(x) \cap supp(x \otimes x) = \emptyset$. Then there is* no *sequence $\alpha_k \to 1$ and x_k, such that*

- *$(\alpha_k w, x_k)$ is consistent with y, and*

•



(a) typical convolution (b) feedback convolution

Figure 1: Different norm measures (normalized and shifted) of a signal x_α that results in the same observation when convolved with a scaled kernel αw. For the typical convolution (1) shown in panel (a), all norms here are ineffective at indicating the unscaled kernel (where α=1). They are either completely insensitive to scaling (ℓ_0, dotted line; ℓ_1/ℓ_2, solid line; overlapping in panel a), or do not reach a minimum at α=1 (ℓ_1, dash line; $\ell_{p=0.8}$, dash-dot line; overlapping in panel a). For the feedback-type convolution (2) shown in panel (b), the scale sensitive ℓ_1 and ℓ_p norms remain ineffective, but the scale invariant ℓ_0 and ℓ_1/ℓ_2 norms identify the unscaled kernel with clear minima.

- $\|x_k\|_1/\|x_k\|_2 < \|x\|_1/\|x\|_2$.

Consequently, the scaling is locally correct when minimizing the ℓ_1/ℓ_2 penalty.

Figure 1b empirically confirms the above two theorems for the discrete case. It contains plots of the ℓ_1 and ℓ_1/ℓ_2 norms of solutions x_α for various scalings of w with α. The original signal has a sparsity of 5, and we observe with the ℓ_0 measure that any perturbation of α around 1 tends to make x_α less sparse. For this type of scaling, (2) still permits an explicit Fourier-domain expression for any given pair $(\alpha w, x_\alpha)$ fitting the data with the scaled model $\hat{x}_\alpha = \hat{y}/(\alpha\hat{w} - \hat{y})$. The true solution for $\alpha = 1$ corresponds to a local minimum for the ℓ_1/ℓ_2 norm of x_α, but it is not a local minimum for both the ℓ_1 norm and the ℓ_p norm with $p < 1$. While the ℓ_p norm is even more biased towards sparse signals than the ℓ_1 norm, it does not seem to achieve the same level of amplitude invariance as the ℓ_1/ℓ_2 norm.

3 A new algorithm for blind deconvolution

Despite the fact that the mathematical properties of the systems with or without feedback have little in common, algorithms initially designed for blind deconvolution ([7, 6, 3, 4]) can be applied to systems with feedback. We discuss some related work with the one over two norm penalty. The optimization problem in these works has the general form

$$\min_{w,x} \|y - w \otimes x\|_2 + \lambda\phi\left(\frac{\|x\|_1}{\|x\|_2}\right) + \rho(x, w), \tag{5}$$

where ϕ is a non-decreasing function, $\rho(x, w) = \rho_1(x) + \rho_2(w)$ with convex ρ_1, ρ_2 representing additional constraints on x and w. Efforts made to tackle the non-smoothness and non-convexity of (5) include [4, 6, 3, 7].

[4] considered the image deblurring problem with an ℓ_1/ℓ_2 penalty on the objective. Here, the authors described the ℓ_1/ℓ_2 penalty as a normalized ℓ_1 norm and propose a reweighed algorithm that alternates between updates on x, w, and $z := \|x\|_2$. [6] formulated the ℓ_1/ℓ_2 in the constraint and lifted the problem to a semidefinite programming. They also employed a special treatment to deal with exponentially growing number of linear constraints. Another approach by [3] and [7] focuses on smoothing the ℓ_1/ℓ_2 norm at the origin by replacing it with a majorant function with parameterizable smoothness. However, in the case with multiple observation traces, this method is susceptible to over-regularizing particular model traces x_i that have a relative small ℓ_2 norm.

In this paper, we introduce the first denoising version of the blind deconvolution problem where the ℓ_1 over ℓ_2 norm appear in conjunction with a data constraint. Moreover, our formulation is flexible enough to include other constraints on the model (such as bounds on ℓ_2 norms of w) or to solve penalty-based formulations such as (5).

Our idea is to use a splitting technique separating x into positive and negative parts, which enables the use of the method of multipliers. To describe our method, We start with the following feasibility formulation:

$$\begin{aligned} &\|\hat{y} - \hat{w} \odot \hat{x} + \hat{y} \odot \hat{x}\|_2 \leq \epsilon \\ &\|x\|_1/\|x\|_2 \leq k \\ &w = Bh, \end{aligned} \tag{6}$$

where $\hat{w}$ stands for the Fourier coefficients, $\odot$ for the pointwise product of two vectors. The parameter k represents prior knowledge on the effective sparsity and $B = [I, 0]^T$ restricts the wavelet to a predefined time interval. Using this formulation, one can easily check whether the algorithm is stuck at a local minimum by examining if all its constraint are satisfied. The obvious drawback of this formulation is its reliance on the estimate of k.

For a given k and (6), we rearrange the constraint $\|x\|_1/\|x\|_2 \leq k$ to $\|x\|_1^2 \leq k^2\|x\|_2^2$, and split x into positive and negative parts $x = x_+ - x_-$, with an additional constraint $\langle x_+, x_- \rangle = 0$ to ensure that the two parts do not overlap. With these manipulations, we arrive at

$$\begin{aligned} &\|\hat{y} - \hat{w} \odot (\hat{x}_+ - \hat{x}_-) + \hat{y} \odot (\hat{x}_+ - \hat{x}_-)\|_2 \leq \epsilon \\ &\langle x_+, x_- \rangle = 0 \\ &(\sum x_+ + \sum x_-)^2 - k^2\|x_+ - x_-\|_2^2 \leq 0 \\ &x_-, x_+ \geq 0 \\ &w = Bh. \end{aligned} \tag{7}$$

Notice that (7) can be represented linearly by the entries of the following rank 1 matrix

$$\begin{bmatrix} w \\ x_+ \\ x_- \\ 1 \end{bmatrix} \begin{bmatrix} w^T & x_+^T & x_-^T & 1 \end{bmatrix} = \begin{bmatrix} ww^T & wx_+^T & wx_-^T & w \\ x_+w^T & x_+x_+^T & x_+x_-^T & x_+ \\ x_-w^T & x_-x_+^T & x_-x_-^T & x_- \\ w^T & x_+^T & x_-^T & 1 \end{bmatrix}.$$

Although (3) can be solved directly by the method of multipliers, this approach is ineffective at resolving issues with problematic local minima that arises from its non-convexity. However, a lifting approach similar to that used in [2] to expand the search space of the model could be applied. We replace the outer product of the two identical vectors with two identical low-rank matrices to expand the space of search directions (note that we cannot get a rank one solution unless an extra low-rank penalty is added.) Concretely, suppose $w, x, y \in \mathbb{R}^N$, we lift these vectors to rank r matrices and solve the following optimization problem:

$$
\begin{aligned}
\min_{R_1,\cdots,R_4} \quad & \text{Trace}(X) - \|X\|_F \ \text{ subject to} \qquad (8) \\
& \|\hat{y} - \text{diag}(\hat{R}_1\hat{R}_2^{\ T} - \hat{R}_1\hat{R}_3^{\ T}) + \hat{y} \odot (\hat{R}_2\hat{R}_4^{\ T} - \hat{R}_3\hat{R}_4^T))\|_2 \le \epsilon \\
& \text{Trace}(R_2R_3^T) = 0 \\
& \mathbf{1}^T(R_2R_2^T + R_3R_3^T + 2R_3R_2^T)\mathbf{1} - k^2\text{Trace}(R_2R_2^T + R_3R_3^T) \le 0 \\
& R_4R_1^T(BB^T - I) = 0 \\
& R_4R_2^T, R_4R_3^T \ge 0 \\
& R_4R_4^T = 1,
\end{aligned}
$$

where $R_1, R_2, R_3 \in \mathbb{R}^{N,r}$ and $R_4 \in \mathbb{R}^{1,r}$. The matrix $\hat{R}_1$ contains the Fourier transform for each column of R_1 and the vector $\mathbf{1}$ contains all ones, I is the identity matrix, and

$$
X = \begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ R_4 \end{bmatrix} \begin{bmatrix} R_1^T & R_2^T & R_3^T & R_4^T \end{bmatrix}.
$$

Since k is not always known, we can also move the sparse penalty to the objective and solve

$$
\begin{aligned}
\min_{R_1,\cdots,R_4} \lambda(\text{Trace}(X) - \|X\|_F) + \log[\mathbf{1}^T(R_2R_2^T + R_3R_3^T \\
+ 2R_3R_2^T)\mathbf{1}] - \log[\text{Trace}(R_2R_2^T + R_3R_3^T))]. \qquad (9)
\end{aligned}
$$

Equations (8) and (9) are special cases of

$$
\min_x F(x) \text{ subject to } \ h_i(x) \in C_i
$$

with C_i convex and F and h_i differentiable with Lipschitz continuous gradients, so the method of multipliers applies.

4 Numerical Experiments

Systems without feedback are less sensitive to initialization because their ambiguity forms a plane consisting of shift and scalings. Systems with feedback, on the other hand, are more challenging to initialize because the existence of multiples effectively shrinks the degrees of freedom of the minimizers. Consequently, we need to be more careful and refrain from biasing the system by initializing the kernel with zeros. This is to be preferred over initializing with a kernel of the wrong scaling and/or location and fits our goal of lessening reliance on priors.

In our experiment, we discretized the real axis with a grid separated by $\Delta t = 0.004s$ and we subsequently generate a sparse signal of 1.2 seconds by setting each of the entries to 0 with probability 0.97. The amplitude of each nonzero element is i.i.d. uniform on [-0.2,0.2]. This box constraint ensures convergence of the feedback system but is not requirement for the algorithm itself. The kernel function is assumed to be the Ricker wavelet with peak frequency at 20Hz. We restrict the wavelet support to 0-0.2s during the reconstruction, and use 0's for its initial guess. We initialize x also with zeros, except that we made x_+, x_- nonzero to make sure that the initial gradient is not zero. For the lifted problem, we initialize R_1 to R_4 by replicating the aforementioned vectors over r columns.

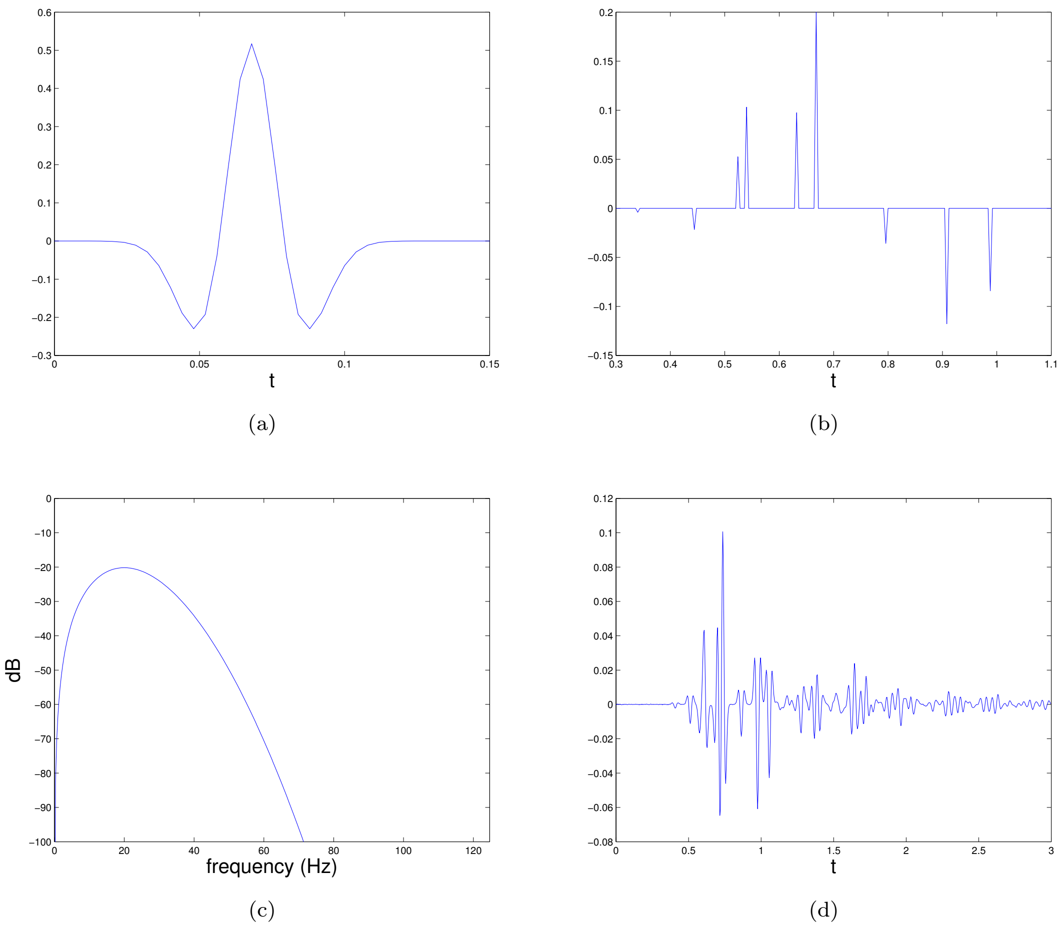
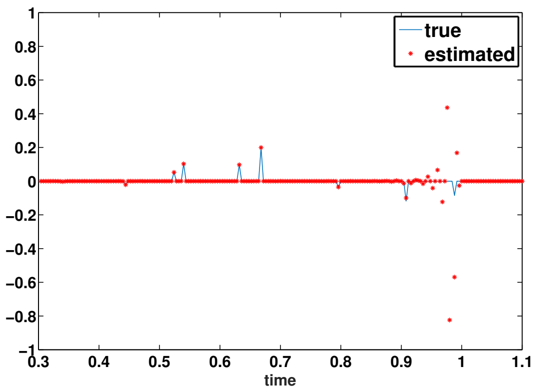
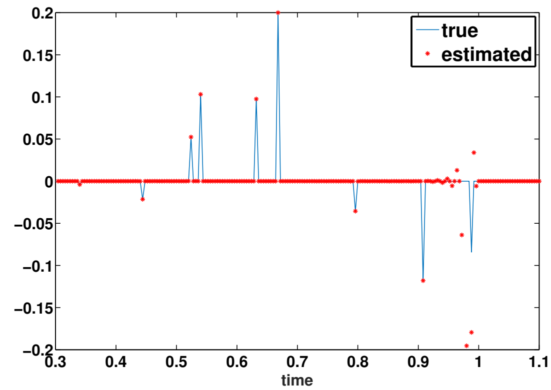


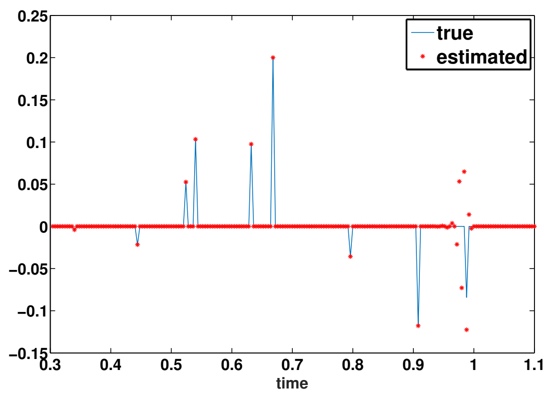
Figure 2: (a) True wavelet, (b) True signal, (c) Spectrum of the wavelet measured in dB, (d) Data record with feedback.



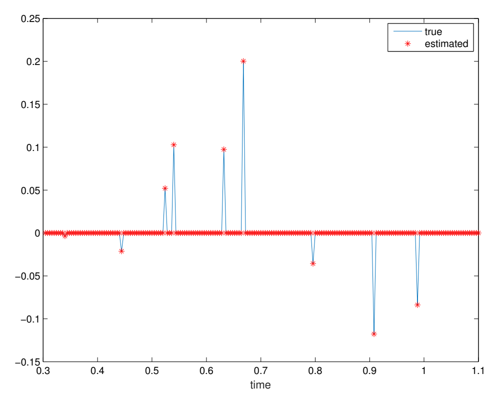
(a) rank 1



(b) rank 2



(c) rank 3



(d) rank 4

Figure 3: Reconstruction of x using (9) with $\lambda = 5 \times 10^4$. One rank is added after every 20 iterations of the outer loop of the method of multipliers.

References

[1] A. Benichoux, E. Vincent, and R. Gribonval. A fundamental pitfall in blind deconvolution with sparse and shift-invariant priors. In *ICASSP-38th International Conf. on Acoustics, Speech, and Signal Processing*, 2013.

[2] A. Cosse, S. D. Shank, and L. Demanet. A short note on rank-2 relaxation for waveform inversion. 2015.

[3] E. Esser, Y. Lou, and J. Xin. A method for finding structured sparse solutions to nonnegative least squares problems with applications. *SIAM Journal on Imaging Sciences*, 6(4):2010–2046, 2013.

[4] D. Krishnan, T. Tay, and R. Fergus. Blind deconvolution using a normalized sparsity measure. *in Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, pages 233–240, 2011.

[5] T. T. Y. Lin and F. J. Herrmann. Robust estimation of primaries by sparse inversion via one-norm minimization. *Geophysics*, 78(3):R133–R150, May 2013.

[6] X. Long, K. Solna, and J. Xin. Two l1 based nonconvex methods for constructing sparse mean reverting portfolios. *UCLA CAM Reports*, pages 1–29, 2014.

[7] A. Repetti, M. Pham, and L. Duval. Euclid in a taxicab: Sparse blind deconvolution with smoothed l1/l2 regularization. *Signal Processing Letters, IEEE*, 22(5):539–543, 2015.

[8] G. J. van Groenestijn and D. J. Verschuur. Estimating primaries by sparse inversion and application to near-offset data reconstruction. *Geophysics*, 74(3):A23–A28, 2009.

[9] D. J. Verschuur, A. J. Berkhout, and C. P. A. Wapenaar. Adaptive surface related multiple elimination. *Geophysics,*, 57(9):1166–1177., 1992.

5 Appendix

Proof of Theorem 2.1. Let $\odot$ and $\oslash$ denote the pointwise multiplication and division, respectively. The assumption $x, w \in L^1$ implies the following facts

1. $w \otimes x^{\otimes k} \in L^1(R)$ for any integer k;

2. $\hat{w} \odot \hat{x}^{\odot k}$ exists everywhere;

3. $$\|y\|_1 \leq \sum_{k=1}^{\infty} \|w * x^{\otimes k}\|_1 \leq \|w\|_1 \sum_{k=1}^{\infty} \|x\|_1^k \leq \frac{\|w\|_1 \|x\|_1}{1 - \|x\|_1};$$

4. $\|\hat{x}\|_\infty \leq \|x\|_1 < 1$;

5. the infinity series converges
$$\sum_{k=1}^{\infty} (-1)^{k-1} \hat{w} \odot \hat{x}^{\odot k} = \hat{w} \odot \hat{x} \oslash (1 - \hat{x}); \tag{10}$$

6. $$\|\hat{w} \odot \hat{x} \oslash (1 - \hat{x})\|_1 \leq \frac{\|\hat{x}\|_\infty}{1 - \|\hat{x}\|_\infty} \|\hat{w}\|_1 \leq \frac{\|x\|_1}{1 - \|x\|_1} \|\hat{w}\|_1;$$

7. $$y = \mathcal{F}^{-1}(\hat{w} \odot \hat{x} \oslash (1 - \hat{x})).$$
   where $\mathcal{F}^{-1}$ is the inverse Fourier transform operator.

These results ensure that the Fourier domain representation of the feedback system (3) obeys

$$\hat{y} = \hat{w} \odot \hat{x} - \hat{x} \odot \hat{y}. \tag{11}$$

We examine whether a simple scaling of w will lower the ℓ_1 norm of the corresponding x. For this purpose, let $v(\alpha) := \alpha w$ be a scaling of w, and defined $z(\alpha)$ to be the reflectivity whose convolution with $v(\alpha)$ is consistent with the data, i.e., $(\widehat{v(\alpha)}, \widehat{z(\alpha)})$ satisfies (11). Clearly, given $v(\alpha) = \alpha w$, the expression of $z(\alpha)$ can be solved as

$$\widehat{z(\alpha)} = \hat{y} \oslash (\widehat{v(\alpha)} - \hat{y}) = \hat{x} \oslash (\alpha + \hat{x}(\alpha - 1)). \tag{12}$$

The true solution corresponds to $\alpha = 1$, i.e., $z(1) = x$.

If the statement of theorem is not true, then $\|z_\alpha\|_1$ has a local minimum w.r.t. α at $\alpha = 1$. Hence to prove the theorem, we only need to show that there is no local minimum at this point.

Since the scaling operator is continuous, there is a neighborhood U of 1 such that $z(\alpha)$ is smooth in U. One can also verify that the second order derivative $|z''(\alpha)| \le M$ is bounded in U and the Taylor expansion of z at $\alpha = 1$ has the form

$$z(\alpha) = \sum_{k=0}^{\infty} (-1)^k x \otimes (\delta + x)^{\otimes k} \alpha^k.$$

where δ is the delta distribution at 0.

Now assume on the contrary that $\|z_\alpha\|_1 \ge \|x\|_1$ for any $\alpha \in U \cap [1, \infty)$. It means that we have

$$\begin{aligned} &\|x\|_1 \le \|z_\alpha\|_1 = \|x + (z_\alpha - z_1)\|_1 = \int sign(z_\alpha)(x + (z_\alpha - z_1))dt \\ &\le \|x\|_1 + \int sign(z_\alpha)(z_\alpha - z_1)dt. \end{aligned}$$

Here the *sign* function is defined to be the sign of the input if it is nonzero and to be 0 otherwise. Canceling the $\|x\|_1$ on both sides to obtain

$$\int sign(z_\alpha)(z_\alpha - z_1)dt \ge 0, \quad \forall \alpha.$$

Inserting (12) to the above and divide both sides by $1 - \alpha$, it becomes

$$\int sign(z_\alpha)\mathcal{F}^{-1}((\hat{x} + \hat{x} \odot \hat{x}) \oslash (\alpha + \hat{x}(\alpha - 1)))dt \le 0.$$

Let α approaches 1 and bring the limit inside (the interchangeability of the limit and integral sign is due to the dominated convergence theorem), we get

$$\int sign(x)(x + x * x) = \int sign(z_1) z'(1)dt \le 0$$

This implies

$$\|x\|_1 - \|x\|_1^2 \le 0,$$

which contradicts to the assumption that $\|x\|_1 < 1$. □

Proof of Theorem 2.2. WLOG, assume $\alpha \in [1, \infty)$ (the proof for $\alpha \le 1$ follows the same line of argument). We prove the result by contradiction. Suppose $\alpha = 1$ is not a local minimizer, then for any $\delta > 1$, there exists $\alpha < \delta$, such that

$$\frac{\|z(\alpha)\|_1}{\|z(\alpha)\|_2} \le \frac{\|x\|_1}{\|x\|_2}.$$

This further means that there exists a sequence $\alpha_n \to 1$ such that

$$\lim_{n\to\infty} \frac{\|z(\alpha_n)\|_1^2 \|x\|_2^2 - \|z(\alpha_n)\|_2^2 \|x\|_1^2}{\alpha_n - 1} \le 0. \tag{13}$$

As above, writing $z(\alpha) = x + z(\alpha) - z(1)$, it can be verified that the (13) is equivalent to

$$\lim_{n\to\infty} \frac{1}{\alpha_n - 1} (\|x\|_1 \|x\|_2^2 \langle sign(z(\alpha)), z(\alpha) - z(1)\rangle + |\langle z(\alpha) - z(1), sign(z(\alpha))\rangle|^2 \|x\|_2^2 \\ - \|z(\alpha) - z(1)\|_2^2 \|x\|_1^2 - 2\langle z_1, z(\alpha) - z_1\rangle \|x\|_1^2) \leq 0. \tag{14}$$

Using the facts that

$$\lim_{\alpha\to 1^+} \frac{z(\alpha) - z(1)}{\alpha - 1} = x + x \otimes x,$$

and

$$\lim_{\alpha\to 1} sign(z(\alpha)_T) = sign(x + x \otimes x) \tag{15}$$

where $T = \text{supp}(x + x \otimes x)$ and $z(\alpha)_T$ denotes the vector $z(\alpha)$ restricted to the set T, (14) becomes

$$2\|x\|_1 \|x\|_2^2 (\|x\|_1 + \|x\|_1^2) - 2\|x\|_1^2 \|x\|_2^2 \leq 0,$$

which cannot be true, hence we arrive at a contradition. Note that in (15) we have used the assumption that $\text{supp}(x) \cap \text{supp}(x \otimes x) = \emptyset$.

$\square$