A Lifted l_1/l_2 Constraint for Sparse Blind Deconvolution

Ernie Esser, Tim Lin, Rongrong Wang and Felix Herrmann

December 9, 2014

Sparse Blind Deconvolution

Problem: Given n traces f_j, estimate source wavelet w and sparse reflectivities x_j

Models:

$$f_j = x_j * w \qquad \text{(BD)}$$

$$f_j = x_j * w - x_j * f_j \qquad \text{(EPSI)}$$

Common Assumptions:

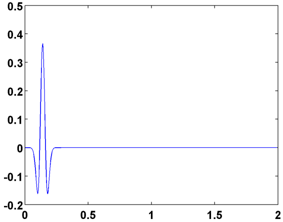
- w is short in time
- w is approximately band limited
- w is minimum phase or impulsive
- good initial guess for w
- x_j is statistically white
- x_j is sparse

Goal: Solve while only assuming sparsity of x_j

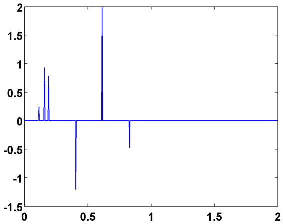
Strategy: New lifted implementation of sparsity promoting constraint $\frac{\|x_j\|_1}{\|x_j\|_2} \leq \sqrt{k}$

Synthetic Linear Convolution Data

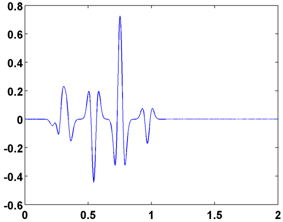
w (peak frequency 10Hz)



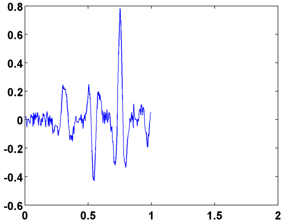
x_1



$x_1 * w$



$f_j = x_1 * w + \eta \ , \ j = 1, ..., n$



Problem Setup and Motivation

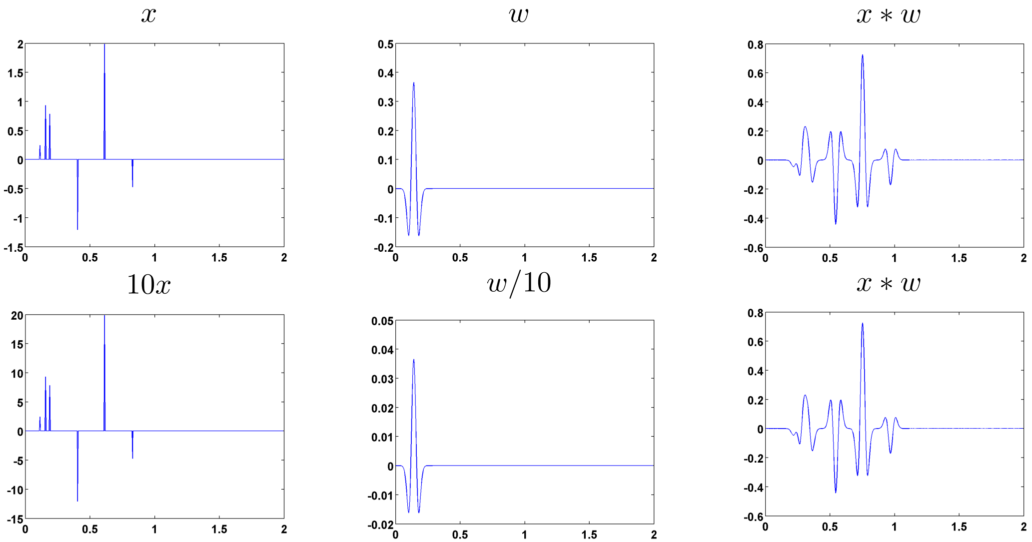
$$f_j = \mathcal{X}(x_j * w) + \eta_j \qquad \mathcal{X} \in \mathbb{R}^{N \times L} \qquad \mathcal{X} = \begin{bmatrix} \mathrm{I}_N & 0 \end{bmatrix}$$

$$w = Bh \qquad B \in \mathbb{R}^{L \times K} \qquad B = \begin{bmatrix} \mathrm{I}_K \\ 0 \end{bmatrix}$$

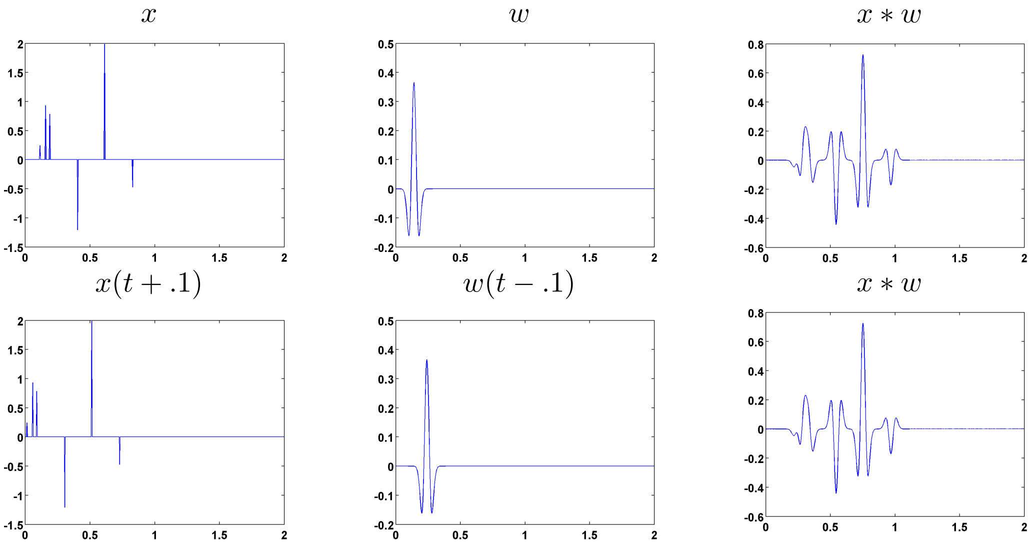
$$x_j = Cm_j \qquad C \in \mathbb{R}^{L \times N} \qquad C = \begin{bmatrix} \mathrm{I}_N \\ 0 \end{bmatrix}$$

- Bilinear convolution measurements $f = Cm * Bh$ are linear in the matrix hm^T
- For good B, C matrices, a lifted convex problem can be solved for a rank one matrix corresponding to hm^T [Ahmed, Recht and Romberg 2012]
- For our B and C matrices, more assumptions are needed such as sparsity of m
- $\frac{\|m\|_1}{\|m\|_2} \leq \sqrt{k}$ can be lifted to a linear constraint
- We will combine lifted blind deconvolution ideas with an l_1/l_2 sparsity constraint

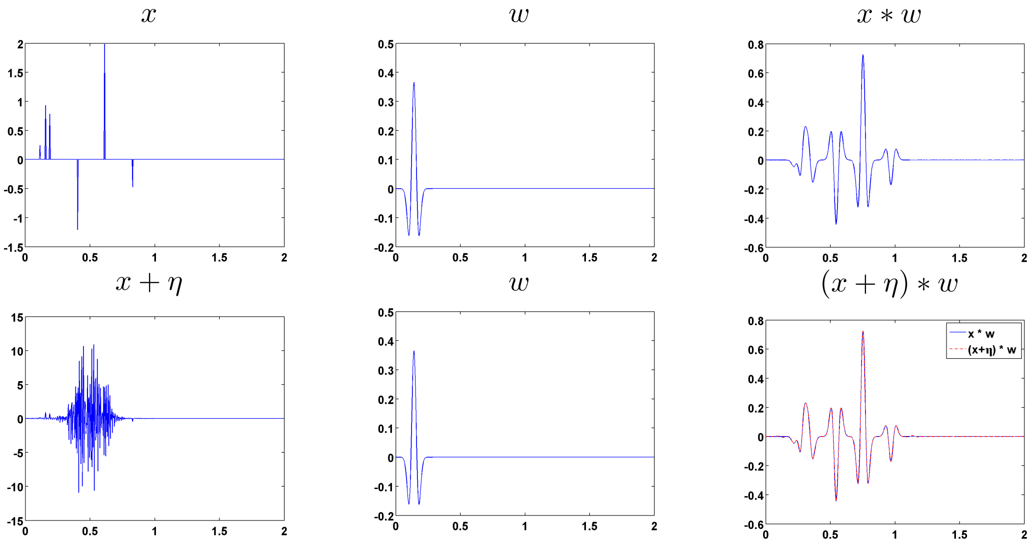
Fundamental Ill-Posedness – Scaling Ambiguity



Fundamental Ill-Posedness – Shift Ambiguity



Fundamental Ill-Posedness – Other Ambiguity



l_1 Regularization

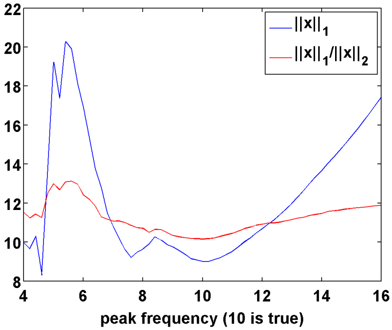
$$\min_{x,w} \frac{\lambda}{2}\|f - x * w\|^2 + \|x\|_1 + \beta\|w\|$$

- Global minimum is trivial: $x \sim \delta$ [Benichoux, Vincent and Gribonval 2013]
- Local minima may or may not be good

However, if w is known, then l_1 regularization can be used to resolve sparse well separated spikes [Claerbout and Muir 1973], [Santosa and Symes 1986], [Donoho 1992], [Dossal and Mallat 2005]

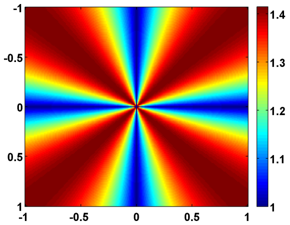
l_1/l_2 Can Evaluate Partially Blind Weiner Deconvolution Results

- Parameterize Ricker wavelet $w(v)$ by peak frequency v
- Use Weiner deconvolution to estimate $x(v)$ such that $f \approx x(v) * w(v)$
- Use l_1 and l_1/l_2 to evaluate the quality of $x(v)$

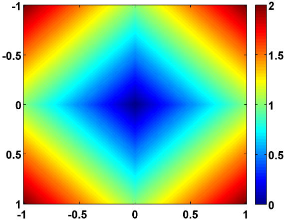


Applications where l_1/l_2 Can Outperform l_1

$\frac{\|x\|_1}{\|x\|_2}$ in two dimensions



$\|x\|_1$ in two dimensions



- Blind image deconvolution [Krishnan, Tay and Fergus 2011], [Ji, Li, Shen and Wang 2012]
- Sparse nonnegative least squares [Esser, Lou and Xin 2013]
- Compressed sensing [Yin, Lou, He and Xin 2014]
- Blind seismic deconvolution [Repetti, Pham, Duval, Chouzenoux and Pesquet 2014] (They smooth an l_1/l_2 penalty and use alternating forward backward iterations)

Connections to Classical Methods

- Minimum Entropy Deconvolution [Wiggins 1978]
  - Maximizes kurtosis $\frac{\|x^2\|_2^2}{\|x^2\|_1^2}$
  - Like minimizing l_1/l_2 applied to x^2 instead of to $|x|$
- Variable Norm Deconvolution [Gray 1979]
  - Maximizes $\frac{\sum_j |x_j|^\alpha}{(\sum_j x_j^2)^{\frac{\alpha}{2}}}$
  - Kurtosis if $\alpha = 4$
  - $\frac{\|x\|_1}{\|x\|_2}$ if $\alpha = 1$, but we would want to minimize to promote sparsity for $\alpha < 2$

Lifted l_1/l_2 Constraint

$$\frac{\|x\|_1}{\|x\|_2} \leq \sqrt{k} \qquad \Leftrightarrow \qquad \|x\|_1^2 - k\|x\|_2^2 \leq 0$$

which can be lifted to

$$1^T|xx^T|1 - k\,\mathrm{tr}(xx^T) \leq 0$$

[D'Aspremont 2011], [Long, Solna and Xin 2014]

Split x into positive and negative parts: $x = x_p - x_m$, $x_p \geq 0$, $x_m \geq 0$ so that $|x| = x_p + x_m$

Obtain a linear lifted l_1/l_2 constraint

$$1^T(x_p + x_m)(x_p + x_m)^T 1 - k\,\mathrm{tr}((x_p + x_m)(x_p + x_m)^T) \leq 0$$

Lifted Blind Deconvolution

$w = Bh$, $x = Cm$, $f = x * w \to f = \mathcal{A}_f(hm^T)$ for linear $\mathcal{A}_f$.

For good B, C matrices, $\min_X \|X\|_*$ s.t. $\mathcal{A}_f(X) = f$ has a unique rank one solution at $X = hm^T$ [Ahmed, Recht and Romberg 2012]

Lift $\begin{bmatrix} h \\ m_p \\ m_m \end{bmatrix}$ to $Z = \begin{bmatrix} h \\ m_p \\ m_m \end{bmatrix} \begin{bmatrix} h^T & m_p^T & m_m^T \end{bmatrix}$ and combine:

- Lifted data constraint
- Lifted sparsity constraint
- Constraint to ensure m_p and m_m have nonoverlapping support and consistent signs
- Normalization constraint on h
- A low rank penalty such as $\mathrm{tr}(Z) - \|Z\|_F$

Rank r Approximation

Solving for Z is too expensive. Compromise by solving for rank r H, M_p and M_m in the factorization

$$Z = \begin{bmatrix} H \\ M_p \\ M_m \end{bmatrix} \begin{bmatrix} H^T & M_p^T & M_M^T \end{bmatrix}$$

Constraints for each measurement:

- Data: $\|f - \mathcal{A}_f(HM_p^T - HM_m^T)\| \leq \epsilon$
- Sparsity: $1^T(M_p + M_m)(M_p + M_m)^T 1 - k\,\mathrm{tr}((M_p + M_m)(M_p + M_m)^T) \leq 0$
- Support and signs: $\mathrm{tr}(M_p M_m^T) = 0$, $M_p \geq 0$, $M_m \geq 0$

Additional constraints and penalties:

- Wavelet normalization $\|h\| = 1$ via $\mathrm{tr}(HH^T) = 1$
- Low rank penalty $\|H\|_F^2 + \|M_p\|_F^2 + \|M_m\|_F^2 - \|H^T H + M_p^T M_p + M_m^T M_m\|_F$
- Optional regularization penalties $\|\Gamma H\|_F^2 + \|M_p + M_m\|_F^2$

Method of Multipliers

$$\min_x F(x) \qquad \text{s.t.} \qquad h_i(x) \in C_i$$

Assume C_i is convex and F, h_i are differentiable with Lipschitz continuous gradient.

Find a saddle point of the augmented Lagrangian

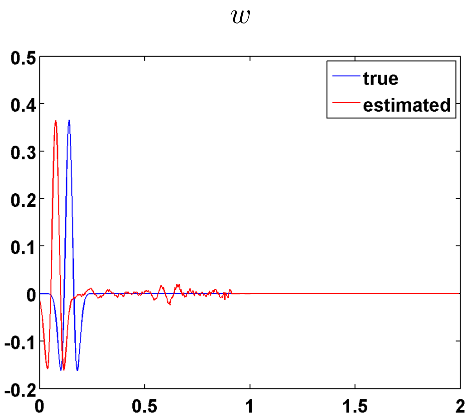
$$L(x,p) = F(x) + \sum_i \frac{1}{2\delta_i} \|D_{C_i}(p_i + \delta_i h_i(x))\|^2 - \frac{1}{2\delta_i}\|p_i\|^2$$

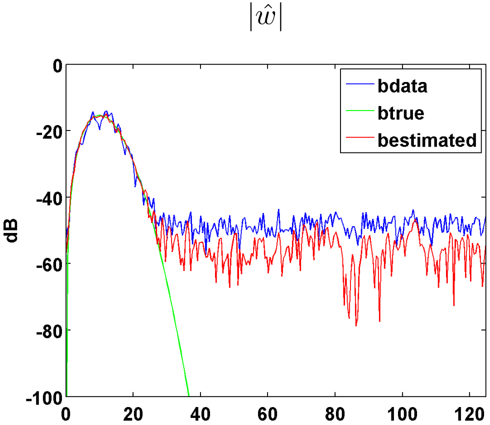
where $D_{C_i}(p) = p - \Pi_C(p)$ (distance from p to C_i)

by iterating

$$x^{k+1} = \arg\min_x L(x, p^k)$$
$$p_i^{k+1} = D_{C_i}(p_i^k + \delta_i h_i(x^{k+1}))$$

Recovered Wavelet for $n = 5$, $r = 1$, SNR $= 23.6$

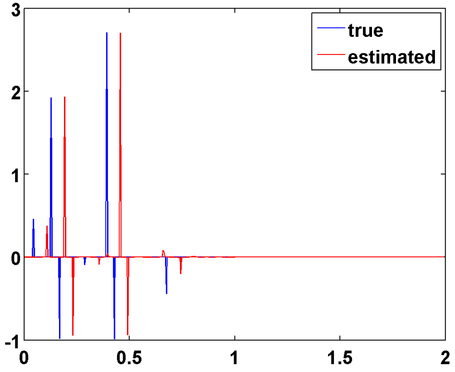




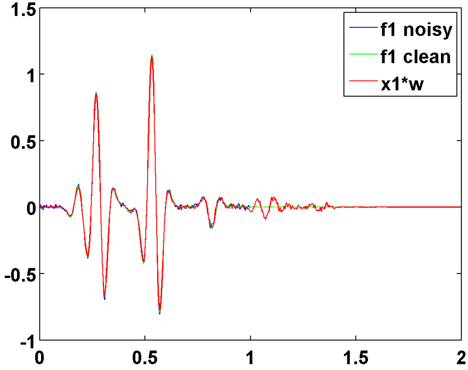
(included $\|\Gamma H\|_F^2$ to promote impulsive wavelet)

Recovered Sparse Signal for $n = 5$, $r = 1$, SNR $= 23.6$

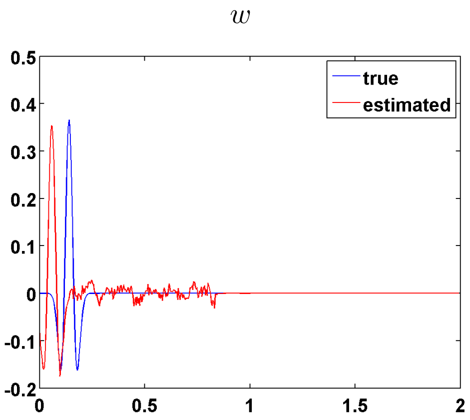
x_1

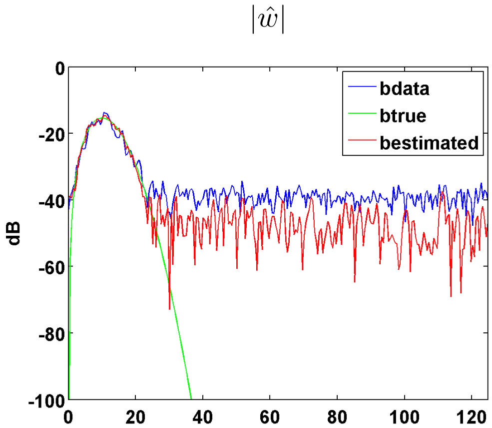


f_1



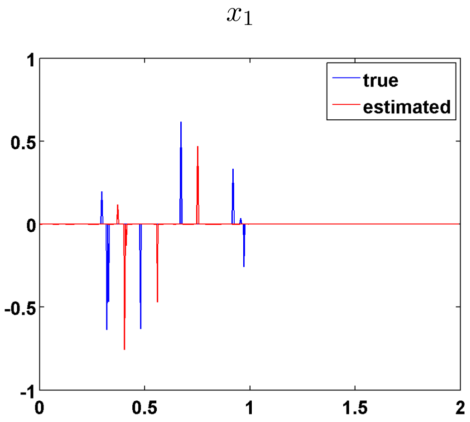
Recovered Wavelet for $n = 5$, $r = 1$, SNR $= 13.5$

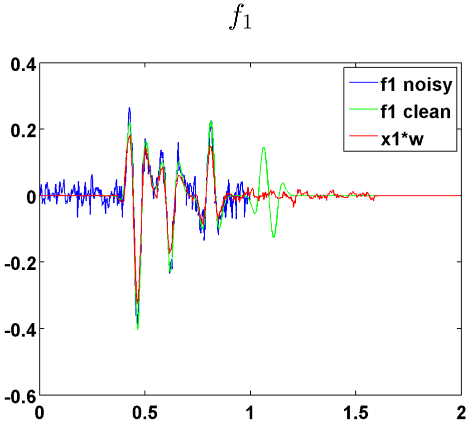




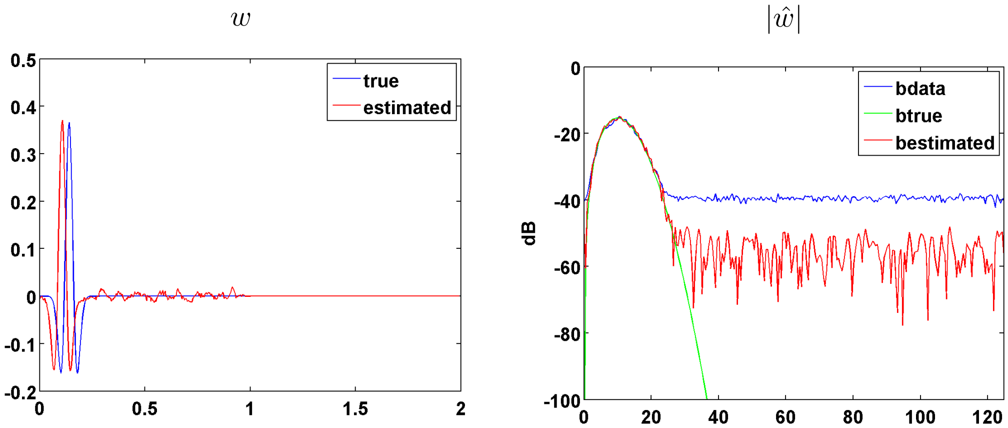
(included $\|\Gamma H\|_F^2$ to promote impulsive wavelet)

Recovered Sparse Signal for $n = 5$, $r = 1$, SNR $= 13.5$

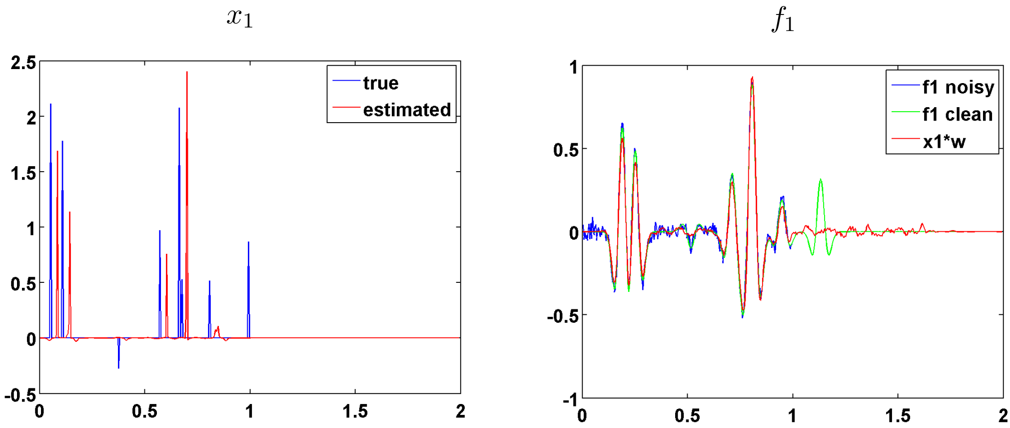




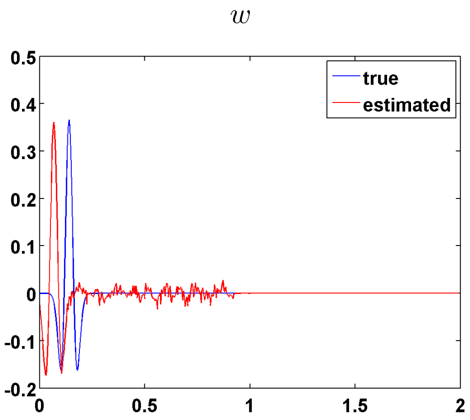
Recovered Wavelet for $n = 50$, $r = 1$, SNR $= 13.5$

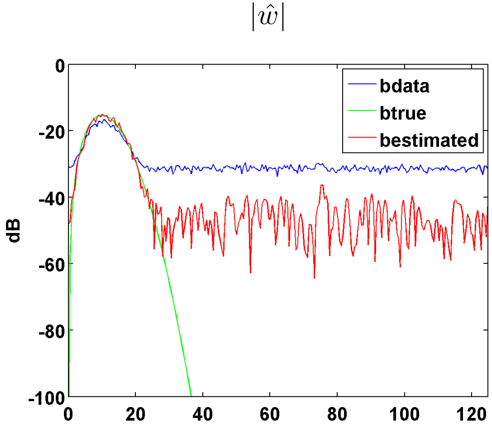


Recovered Sparse Signal for $n = 50$, $r = 1$, SNR $= 13.5$



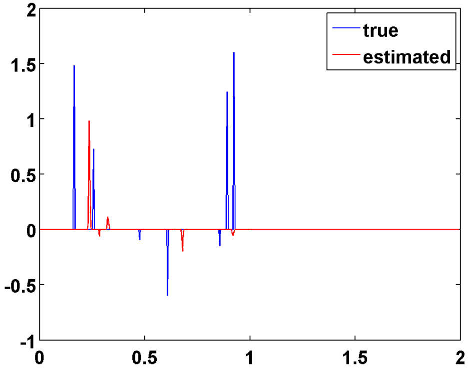
Recovered Wavelet for $n = 50$, $r = 1$, SNR $= 5.25$



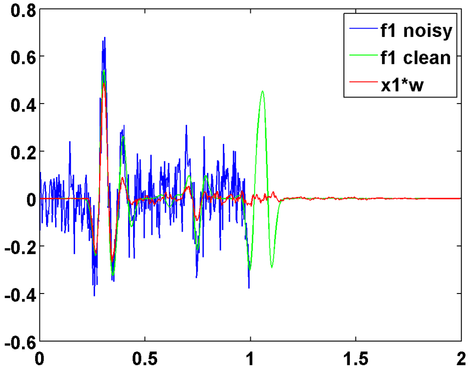


Recovered Sparse Signal for $n = 50$, $r = 1$, SNR $= 5.25$

x_1

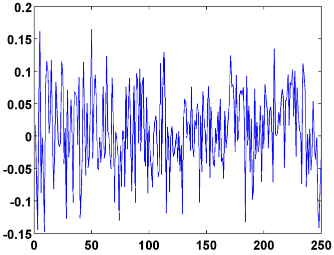


f_1

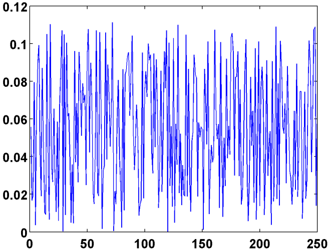


Random Initial Guess

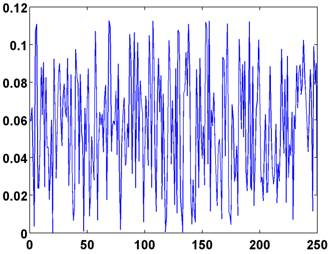
Initial H



Initial M_p



Initial M_m

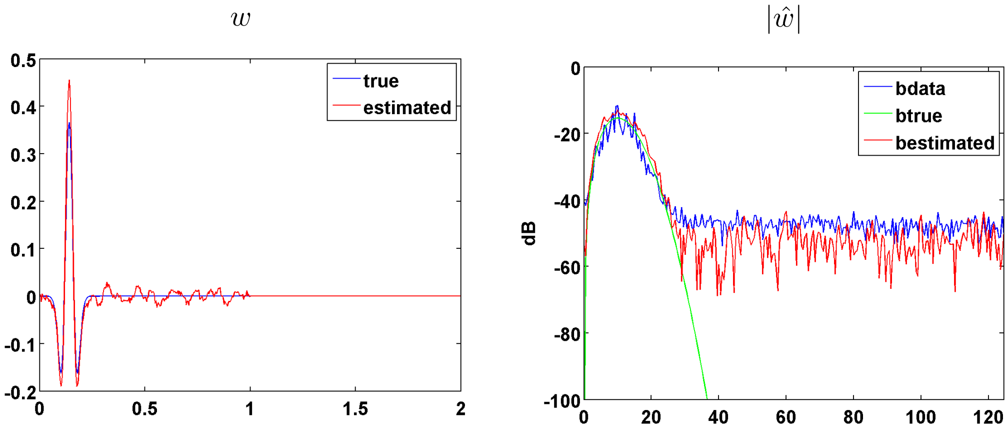


Modifications for EPSI Model

$$g_j = x_j * w - x_j * g_j$$

- Change data constraint to $\|g - \mathcal{A}_f(HM_p^T - HM_m^T) + \mathcal{A}_g(GM_p^T - GM_m^T)\| \le \tilde{\epsilon}$
- Replace $\mathrm{tr}(HH^T) = 1$ with $\mathrm{tr}(HH^T) \ge c$
- Must prevent spikes at early times. Since $x = Cm$, let $C = \begin{bmatrix} 0 \\ \mathrm{I} \\ 0 \end{bmatrix}$

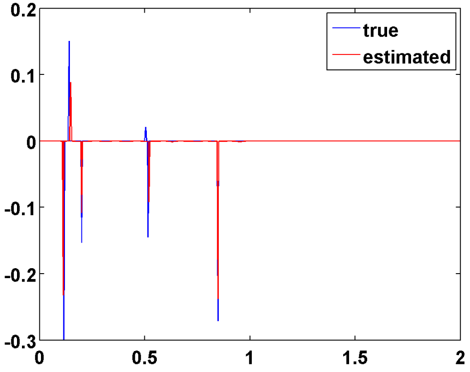
EPSI Recovered Wavelet for $n = 5$, $r = 1$, SNR $= 20.8$



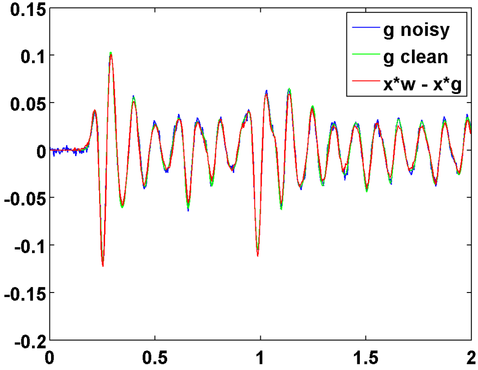
(included $\|\Gamma H\|_F^2$ to promote impulsive wavelet)

EPSI Recovered Sparse Signal for $n = 5$, $r = 1$, SNR $= 20.8$

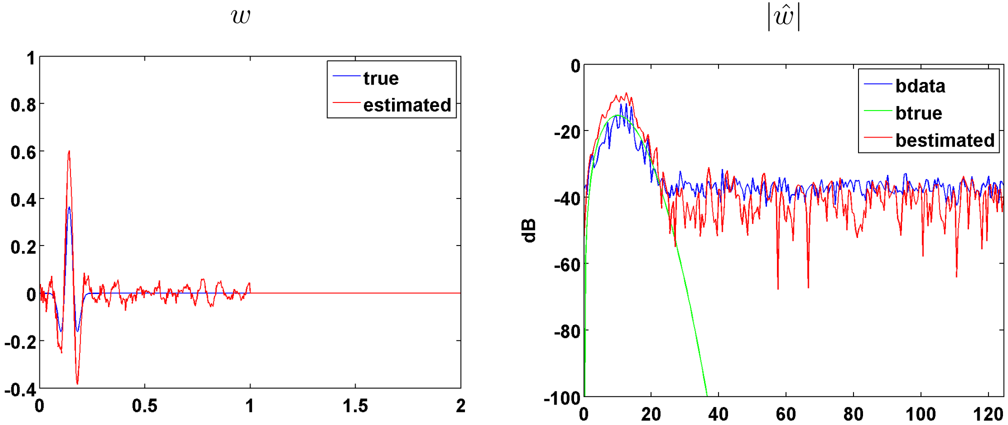
x_1



f_1

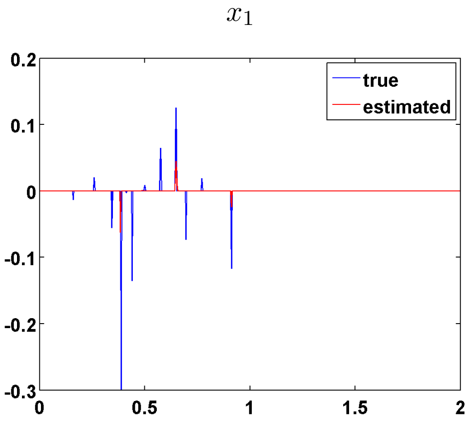


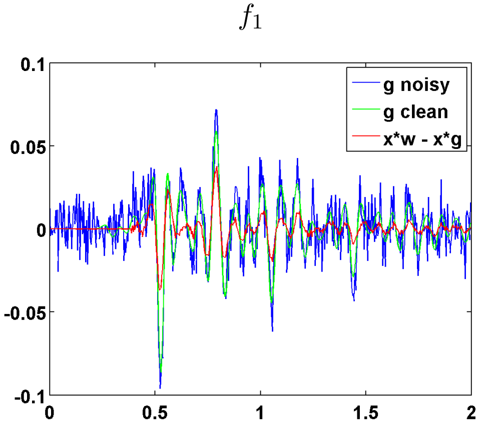
EPSI Recovered Wavelet for $n = 5$, $r = 1$, SNR $= 9.82$



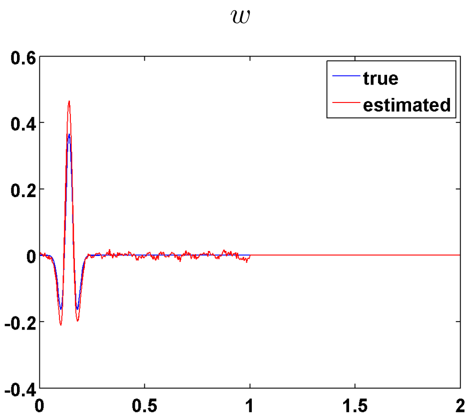
(included $\|\Gamma H\|_F^2$ to promote impulsive wavelet)

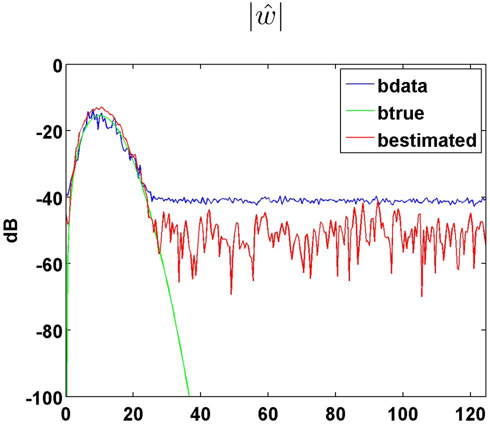
EPSI Recovered Sparse Signal for $n = 5$, $r = 1$, SNR $= 9.82$



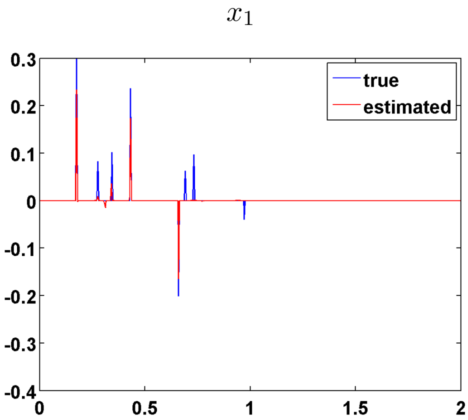


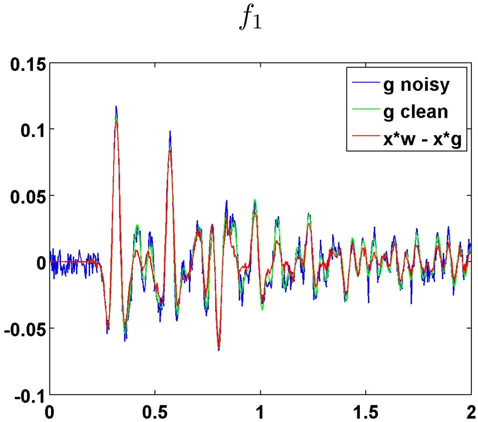
EPSI Recovered Wavelet for $n = 50$, $r = 1$, SNR $= 14.2$





EPSI Recovered Sparse Signal for $n = 50$, $r = 1$, SNR $= 14.2$





Conclusions and Future Work

- Method of Multipliers implementation of a lifted l_1/l_2 sparsity constraint can solve EPSI and standard 1D blind deconvolution problems
- Works with a random initial guess
- With more measurements, results improve and data can be noisier
- Higher rank $r > 1$ (not shown) works, but results so far are slightly worse than $r = 1$
- Compare to alternative approaches and other l_1/l_2 implementations
- Incorporate into multilevel EPSI algorithm at the coarsest level, where the EPSI deconvolution problems are smaller but more difficult
- Prepare future software release