Resolving Scaling Ambiguities with the L_1/L_2 Norm in a Blind Deconvolution Problem with Feedback

Ernie Esser, Rongrong Wang, Tim Lin, and Felix Herrmann

Oct 27, 2015

SLIM

UBC

In memory of our dear friend:

John "Ernie" Esser (May 19, 1980 - March 8, 2015)

Sparse Blind Deconvolution

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

Basic Model:

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

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 and short duration in w

Surface Related Multiples

With a free surface boundary condition, density $\rho = 0$ at the surface.

Reflection coefficient is $R = \frac{v_1\rho_1 - v_0\rho_0}{v_1\rho_1 + v_0\rho_0} = -1$.

Primary reflections $x_j * w$ act like a new source with the opposite sign when reflecting off the surface:

$$f_j = x_j * w - x_j * x_j * w + x_j * x_j * x_j * w - + ...$$

which sums to

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

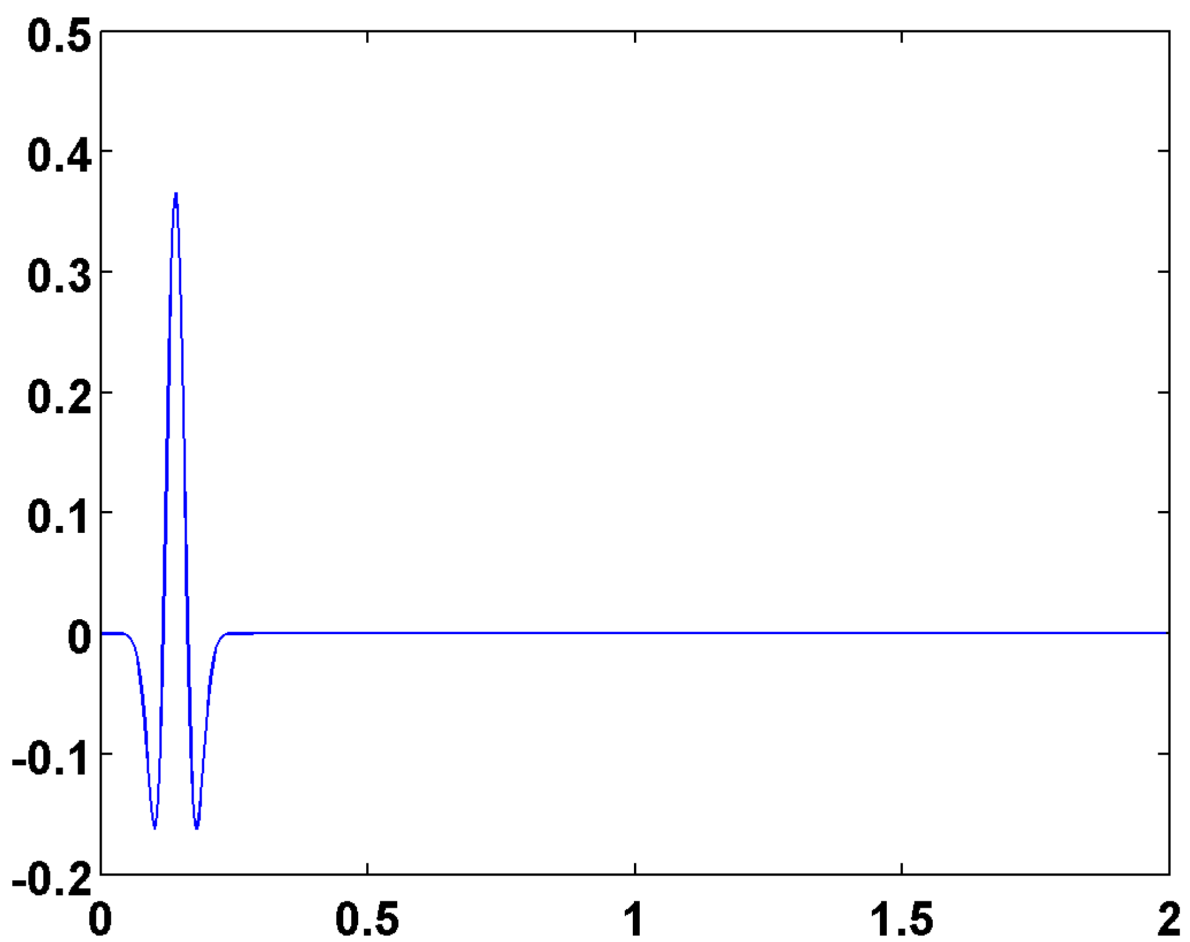
Main points of this talk

We address that

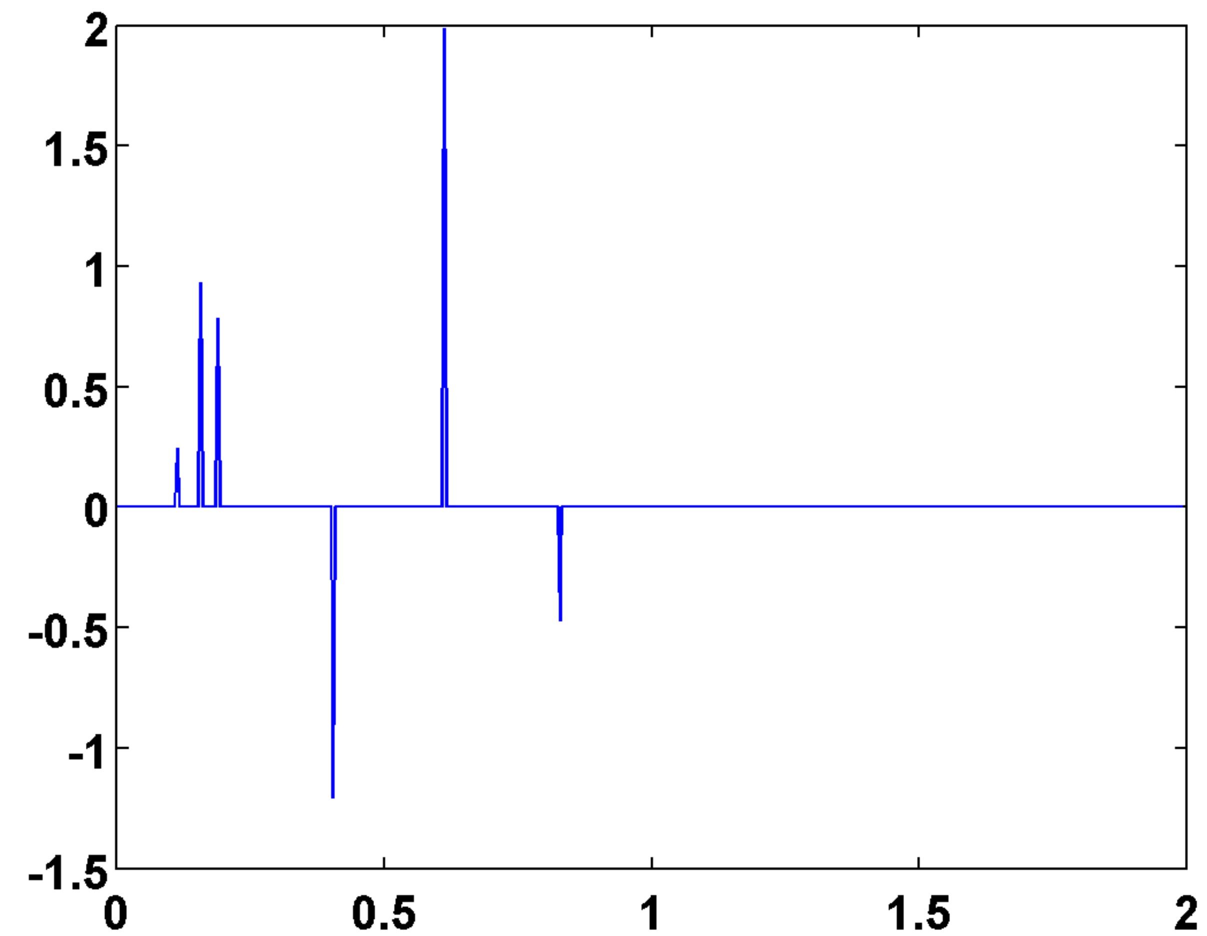
- the multiples can help resolving the scaling and shift ambiguities which are intrinsic in the non-feedback system.
- the l_1/l_2 penalty is more effective than the l_1 norm.
- the proposed rank lifting techniques can mitigate the local minimum potentially existing in any bilinear optimization problems.

Synthetic Linear Convolution Data

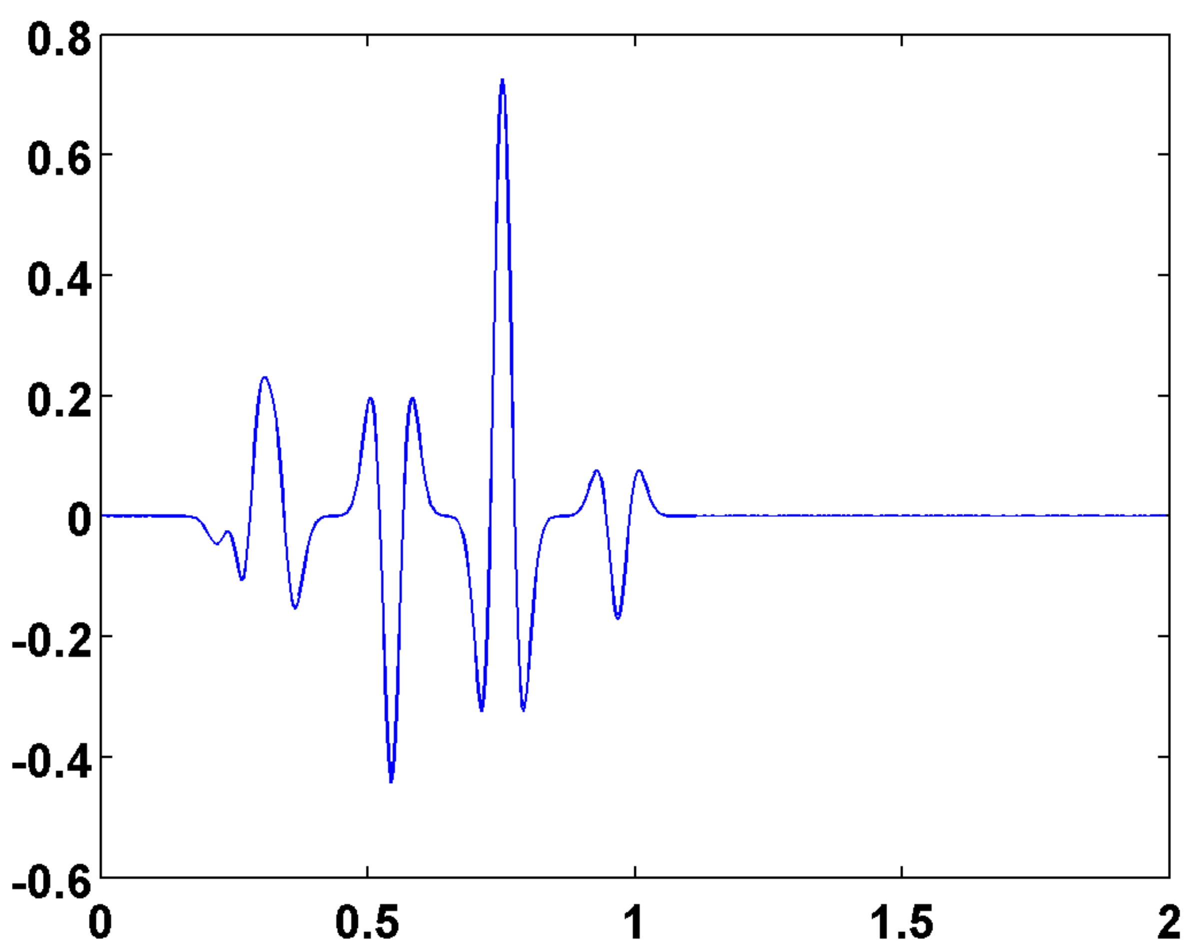
w (peak frequency 10Hz)



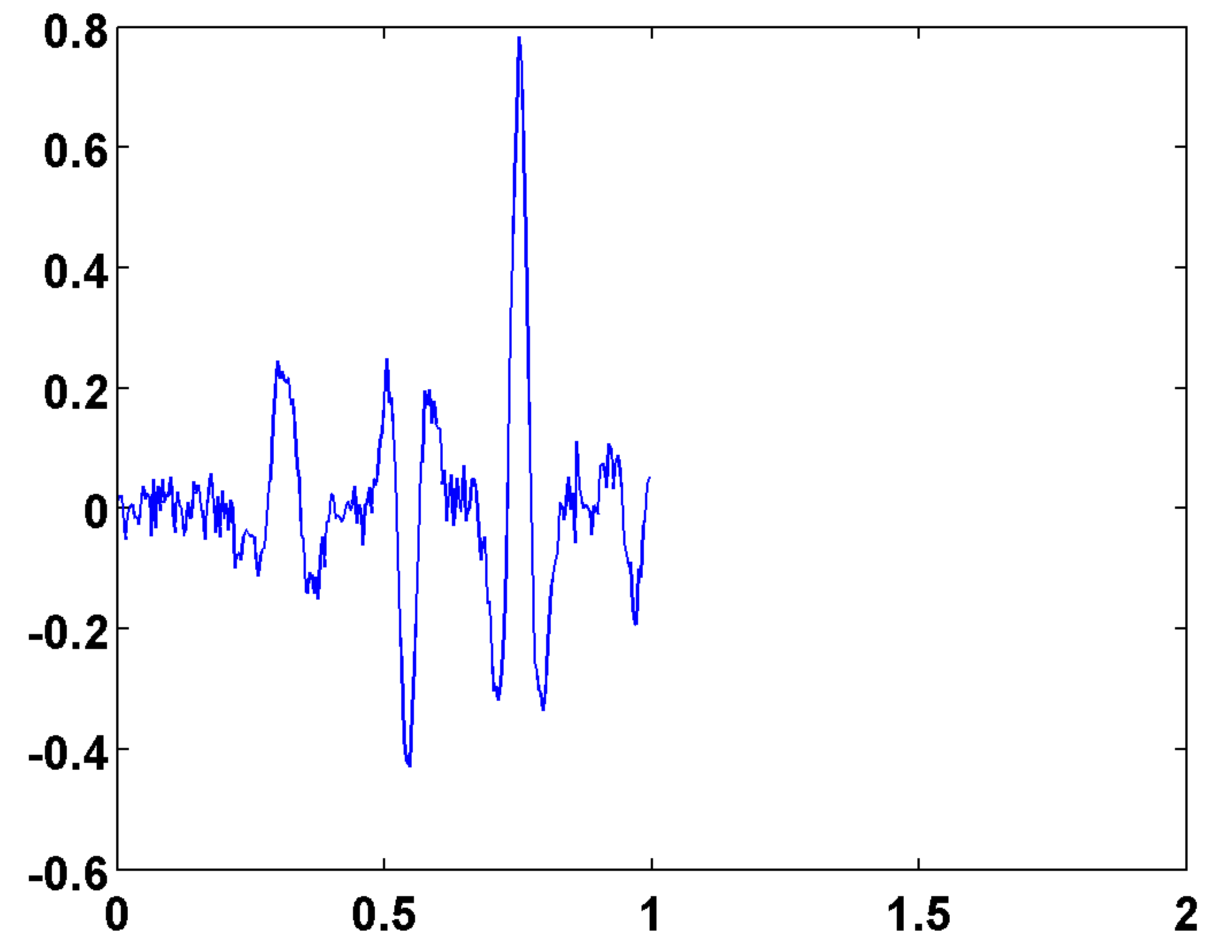
x_1



$x_1 * w$



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



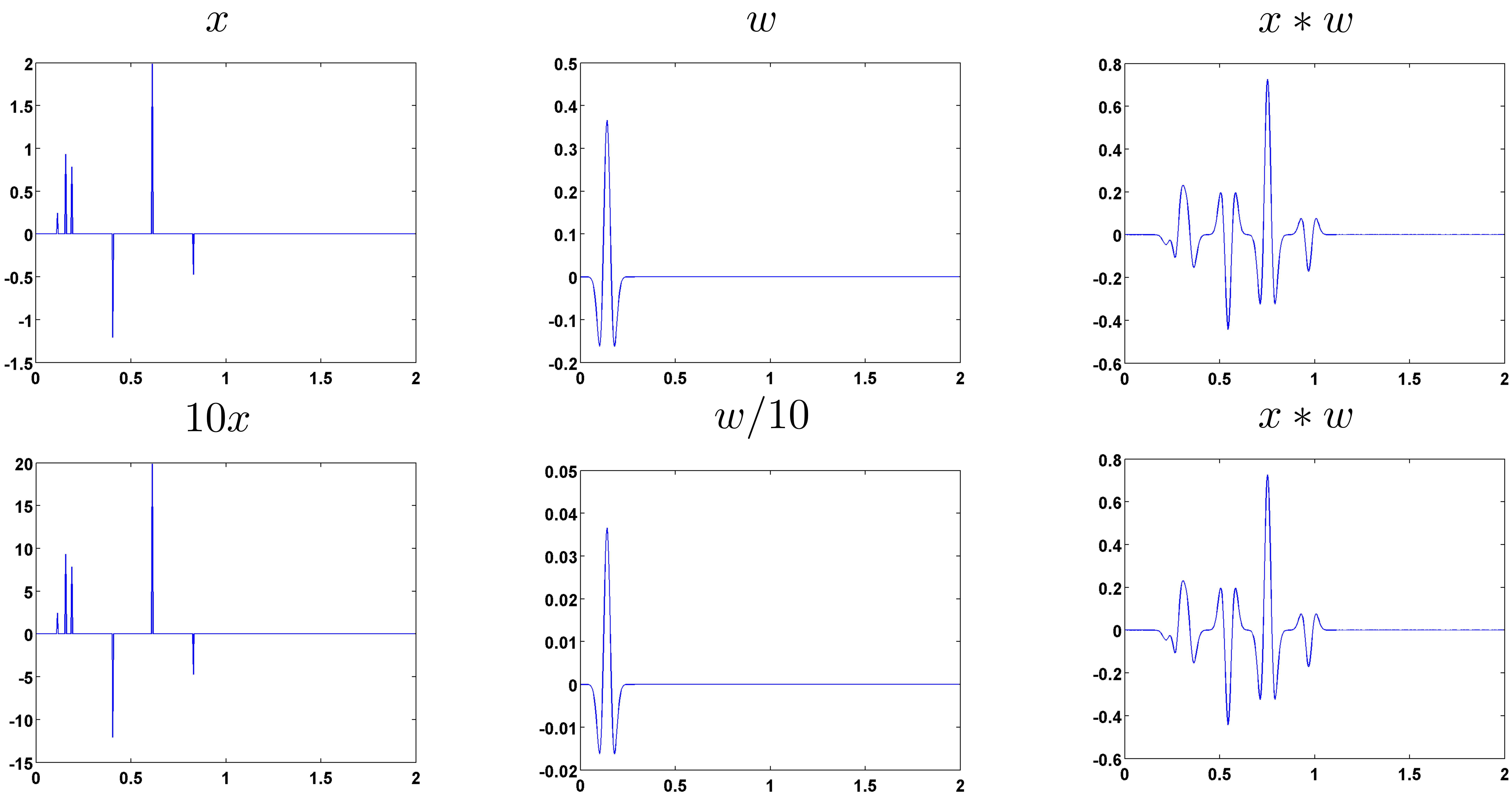
Problem Setup

Without the sparsity assumption on x_j, we are solving a feasibility problem for x_j and w,

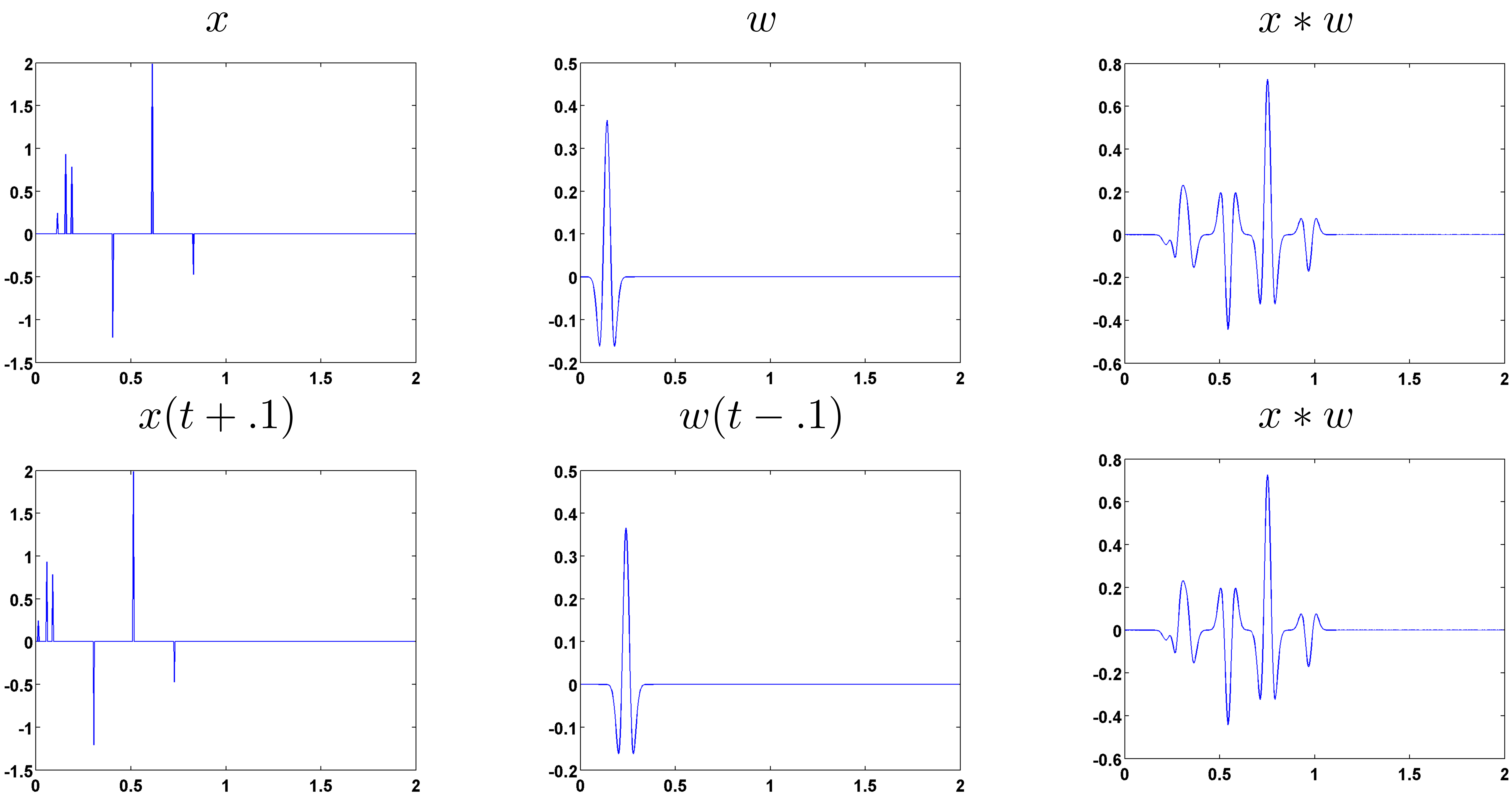
$$\begin{cases} f_j = x_j * w + \eta_j & \mathcal{X} \in \mathbb{R}^{N \times L} \\ w = Bh & B \in \mathbb{R}^{L \times K}, \quad B = \begin{bmatrix} \mathrm{I}_K \\ 0 \end{bmatrix} \end{cases}$$

The problem suffers from scaling, shift and other ambiguities.

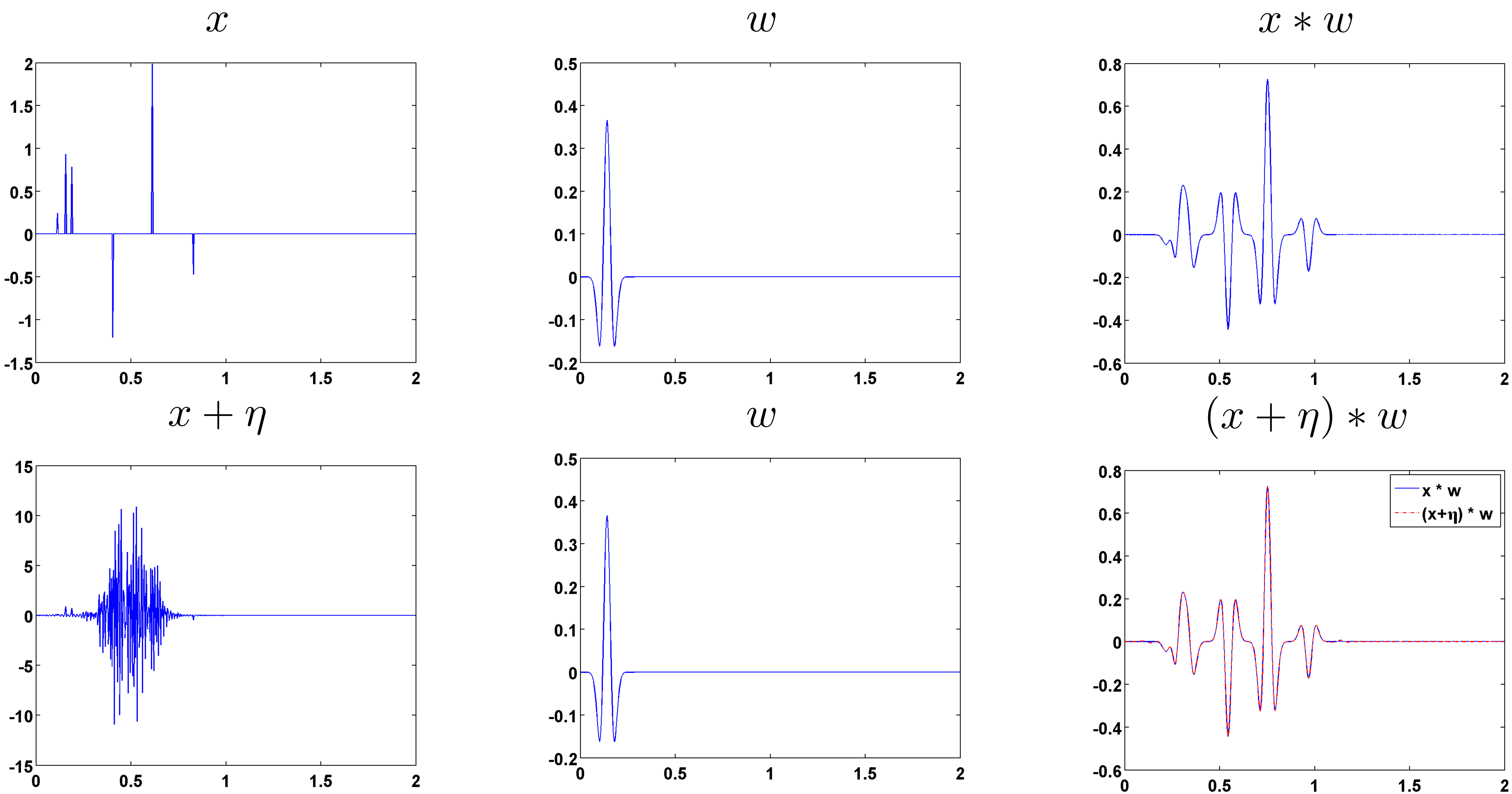
Fundamental Ill-Posedness – Scaling Ambiguity



Fundamental Ill-Posedness – Shift Ambiguity



Fundamental Ill-Posedness – Other Ambiguity



l_1 Regularization

With the l_1 regularization to promote sparsity

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

- 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 Regularization

Replace the l_1 norm with a l_1/l_2 penalty

$$\min_{x,w\in \mathrm{span}(B)} \frac{\lambda}{2}\|f - x * w\|^2 + \frac{\|x\|_1}{\|x\|_2} + \beta\|w\|$$

Or the denoising version

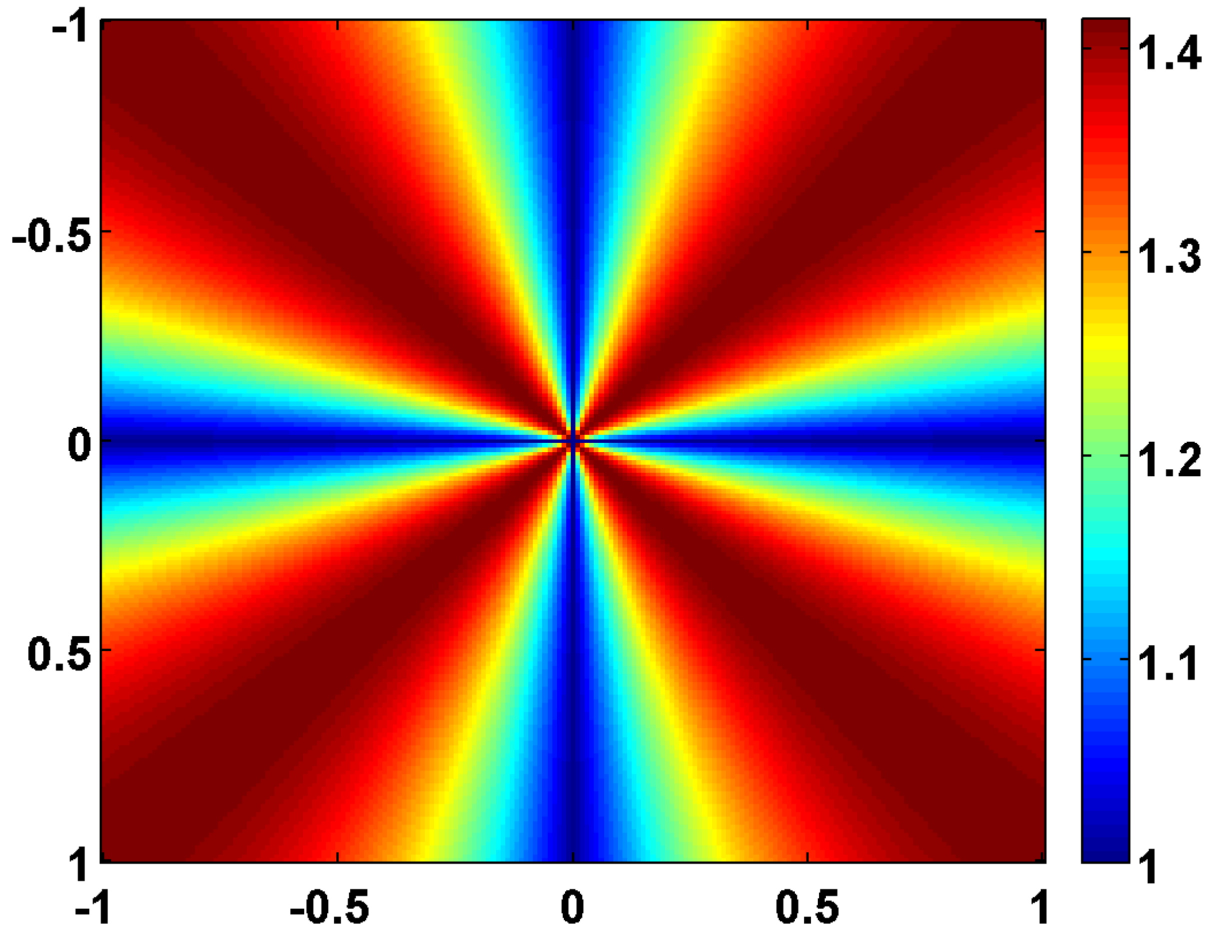
$$\min_{x,w} \log \frac{\|x\|_1}{\|x\|_2}$$

$$\text{subject to } \|f - x * w\|_2 \leq \epsilon$$

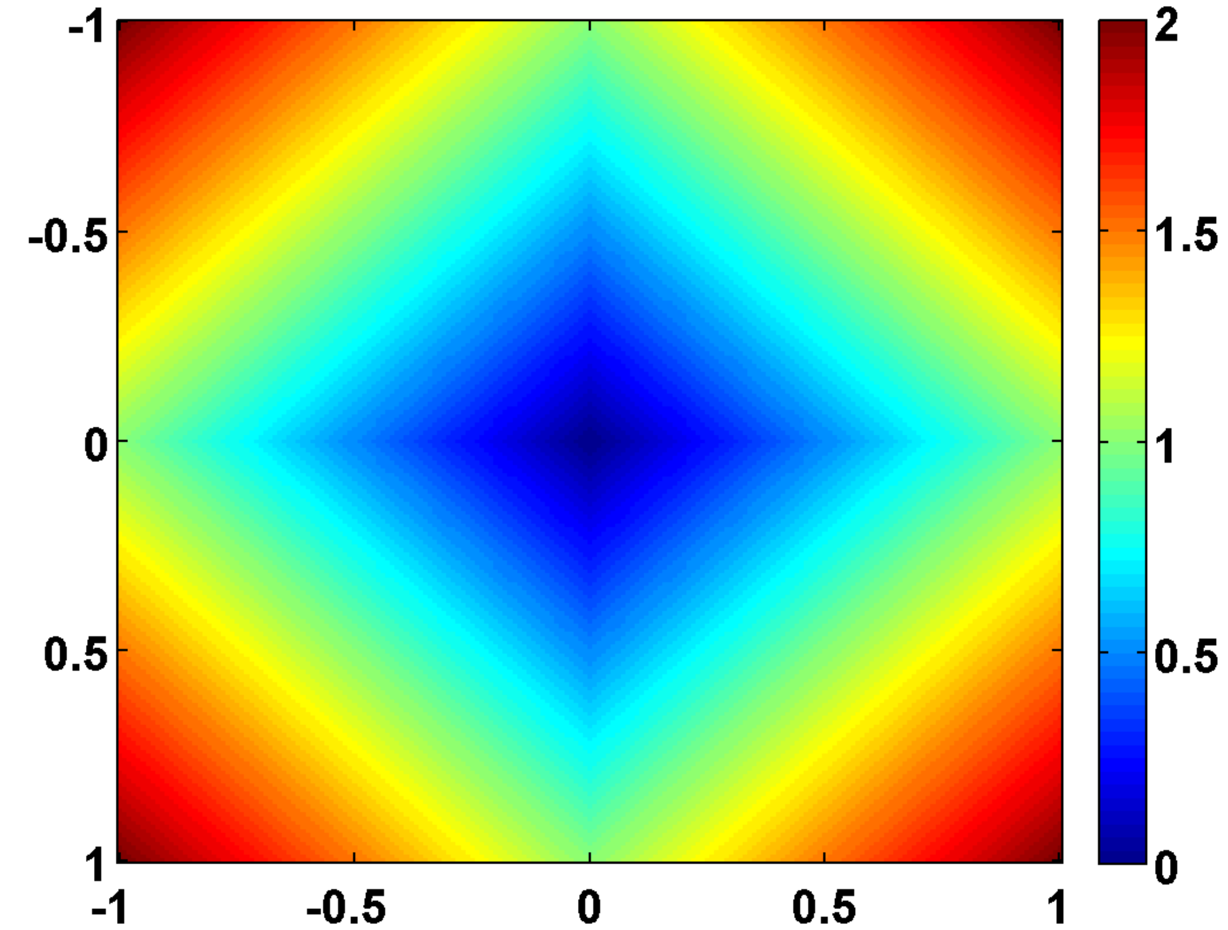
$$w = Bh.$$

Applications where l_1/l_2 Can Outperform l_1

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



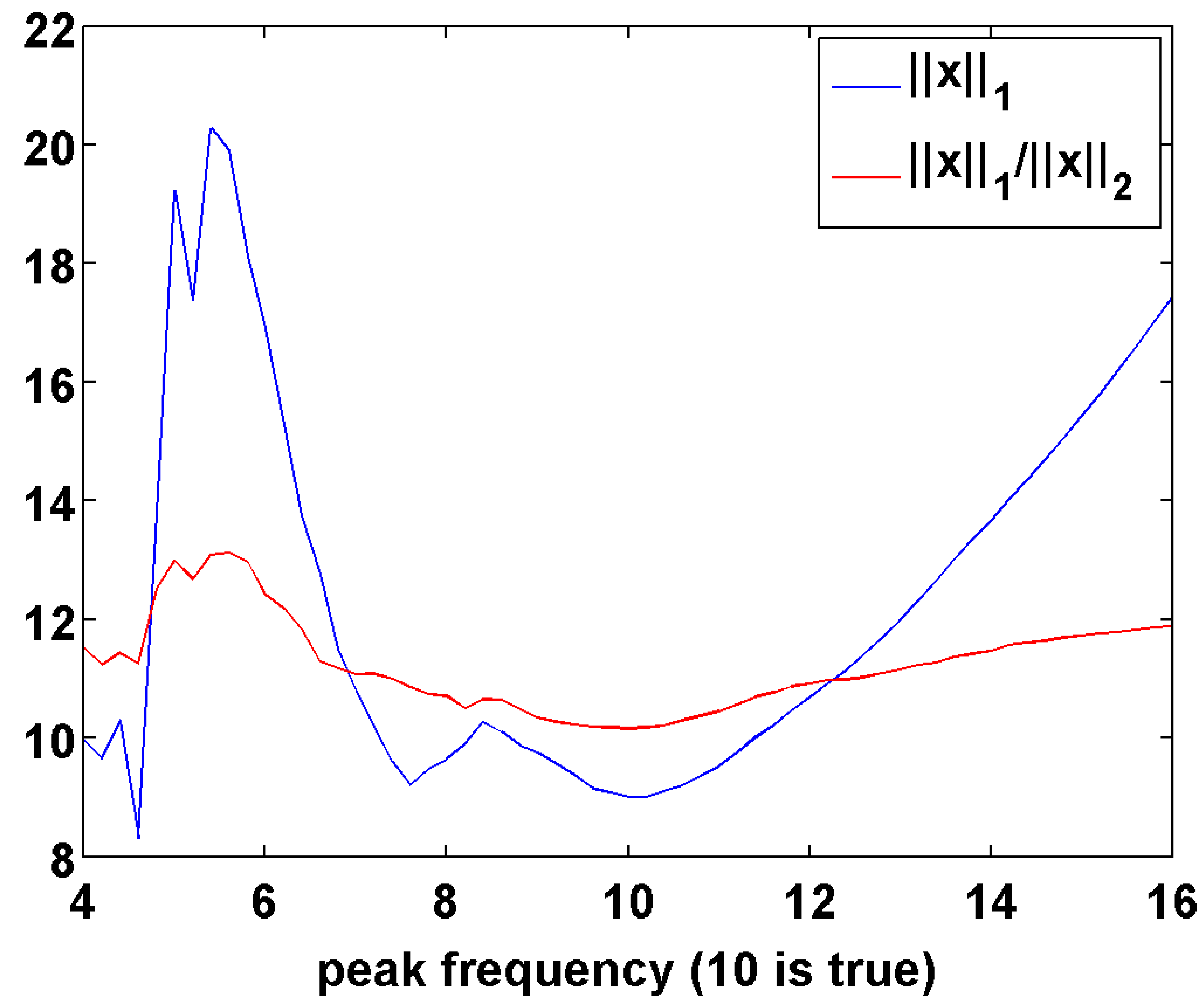
$\|x\|_1$ in two dimensions



- 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)

l_1/l_2 Can Evaluate Partially Blind Wiener Deconvolution Results

- Parameterize Ricker wavelet $w(v)$ by peak frequency v
- Use Wiener 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 $x(v)$



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 $\alpha < 2$

l_1/l_2 in the feedback system

Solve for w and x via the optimization problem:

$$\min_{[x_1,\ldots,x_n],w\in\text{span}(B)} F(x) \qquad \text{(M)}$$
$$\text{subject to } f_j = x_j * w - x_j * f_j, \quad j = 1, \ldots, n.$$

Theorem 1 (Esser-Wang-Lin-Herrmann,2015).

If $F(x) = \|x\|_{l_1}$, then (M) has a scaling ambiguity for any x and w.
If $F(x) = \|x\|_{l_1}/\|x\|_{l_2}$, and $supp(x) \cap supp(x * x) = \emptyset$, then (M) has no scaling ambiguity.

Solving the optimization problem

Original problem is non-convex non-differentiable

$$\min_{x_j, w} \sum_j \log \frac{\|x_j\|_1}{\|x_j\|_2}$$

$$\text{subject to} \begin{cases} \|f_j - x_j * w + x_j * f_j\|_2 \leq \epsilon, j = 1, .., n \\ w = Bh. \end{cases}$$

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

$$\min_{x_{j,\pm}, w} \sum_j \log \frac{\mathbf{1}^T (x_{j,+} + x_{j,-})}{\|x_{j,+} - x_{j,-}\|_2}$$

$$\text{subject to} \begin{cases} \|f_j - (x_{j,+} - x_{j,-}) * w + (x_{j,+} - x_{j,-}) * f_j\|_2 \leq \epsilon, j = 1, .., n \\ w = Bh \\ x_{j,+}^T x_{j,-} = 0. \end{cases}$$

The problem now is differentiable with Lipschitz continuous gradient.

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

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

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

by iterating

$$x^{k+1} = \arg\min_x L(x, p^k)$$

$$p_i^{k+1} = D_{\delta_i C_i}(p_i^k + \delta_i h_i(x^{k+1}))$$

In practice, approximate x^{k+1} with a quasi-Newton method such as LBFGS.

Lifted Blind Deconvolution

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

Change variable from $\begin{bmatrix} h \\ x_p \\ x_m \end{bmatrix}$ to $Z = \begin{bmatrix} h \\ x_p \\ x_m \\ 1 \end{bmatrix} \begin{bmatrix} h^T & x_p^T & x_m^T & 1 \end{bmatrix}$ and the optimization problem becomes

$$\min_Z \quad F(Z)$$
$$\text{subject to } \mathcal{A}_j(Z) = y_j, \quad j = 1, ..., n$$

Rank r Approximation

Lifting the rank of Z to r in the factorization

$$Z = \begin{bmatrix} H \\ X_p \\ X_m \\ \alpha \end{bmatrix} \begin{bmatrix} H^T & X_p^T & X_M^T & \alpha^T \end{bmatrix}$$

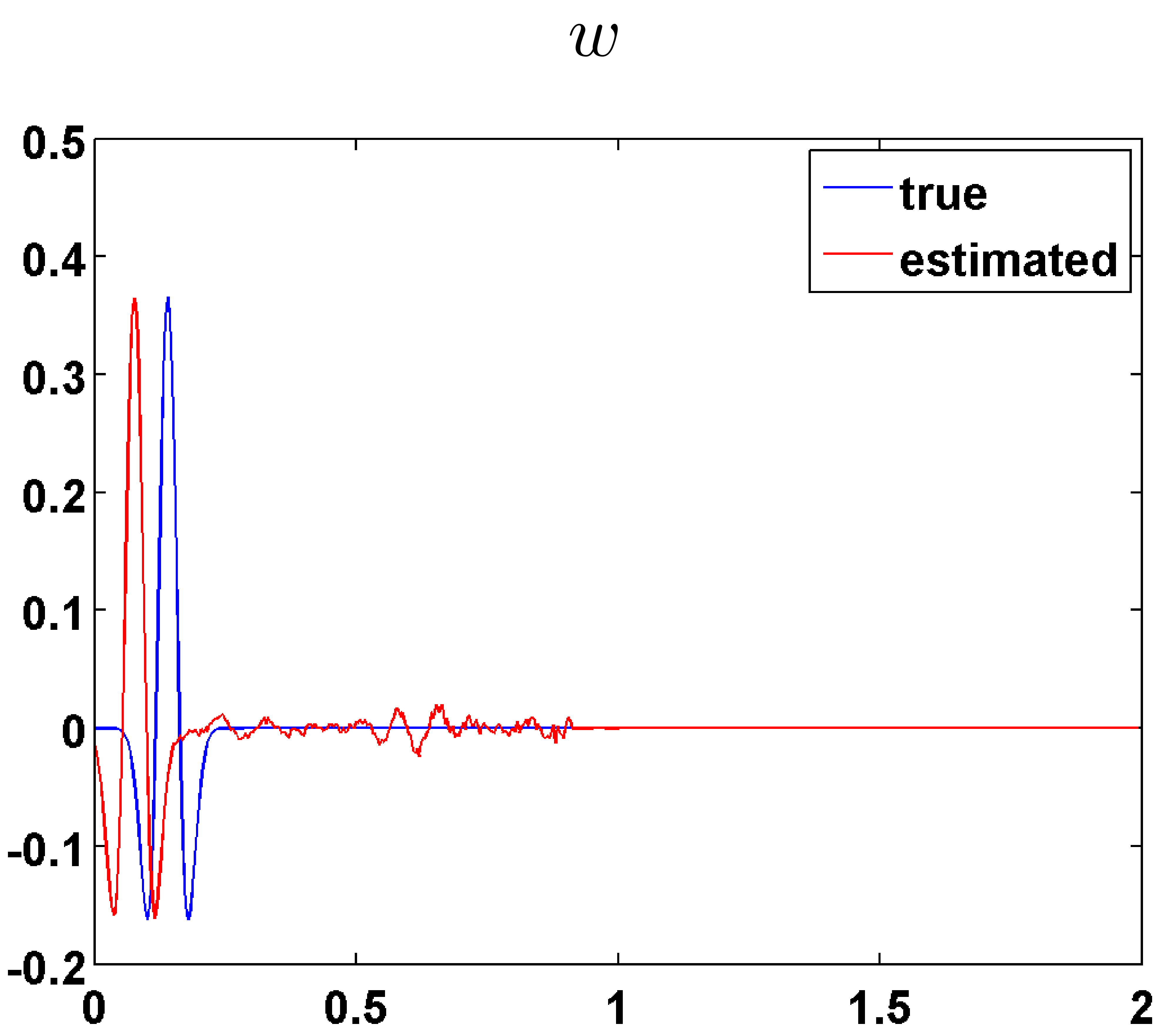
and adding low rank promoting penalty $\| \cdot \|_* - \| \cdot \|_F$

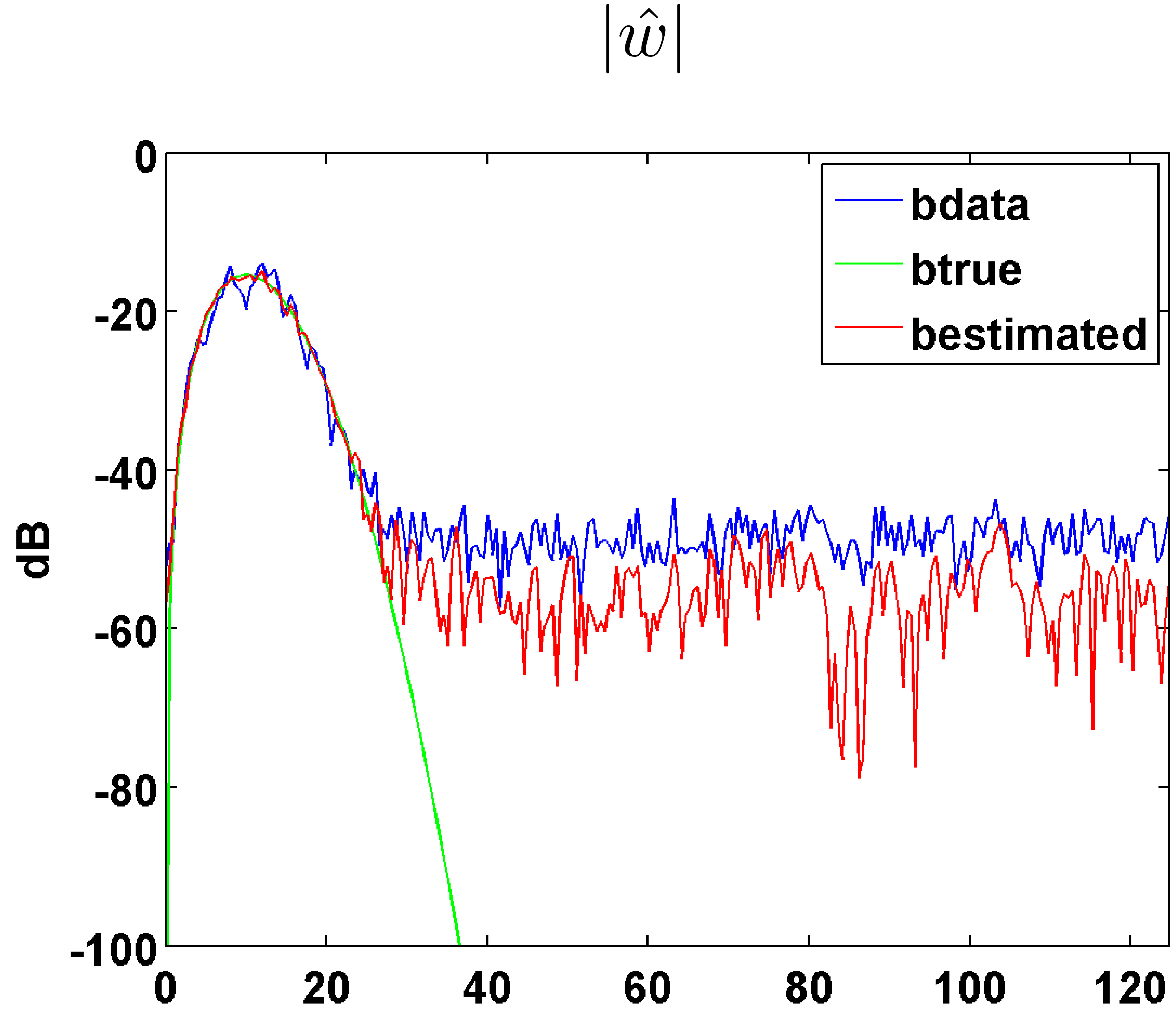
$$\min_Z \; F(Z) + \|Z\|_* - \|Z\|_F$$
$$\text{subject to } \alpha\alpha^T = 1 \text{ and } \mathcal{A}_j(Z) = y_j, \quad j = 1, ..., n$$

Additional constraints and penalties:

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

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

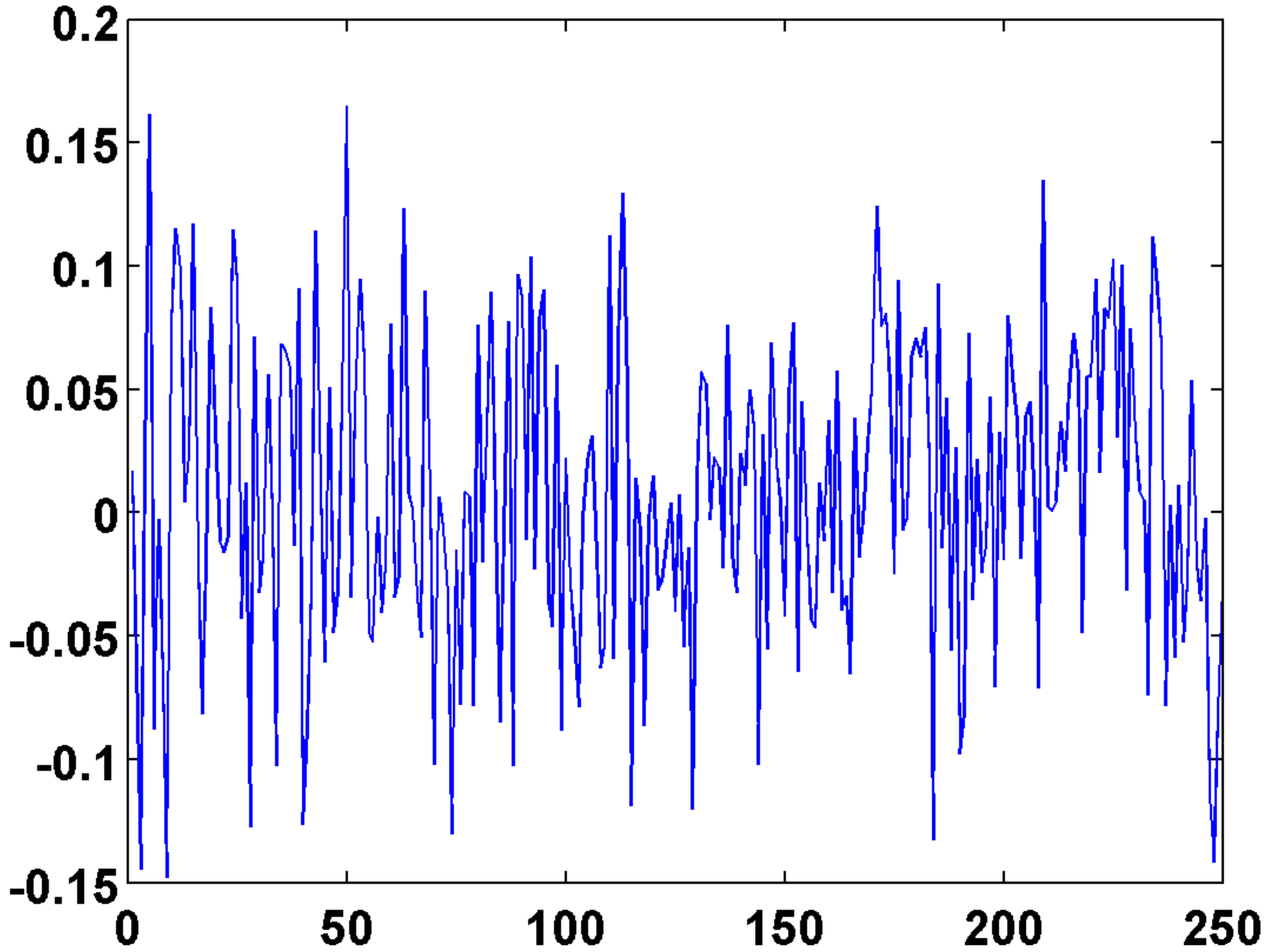




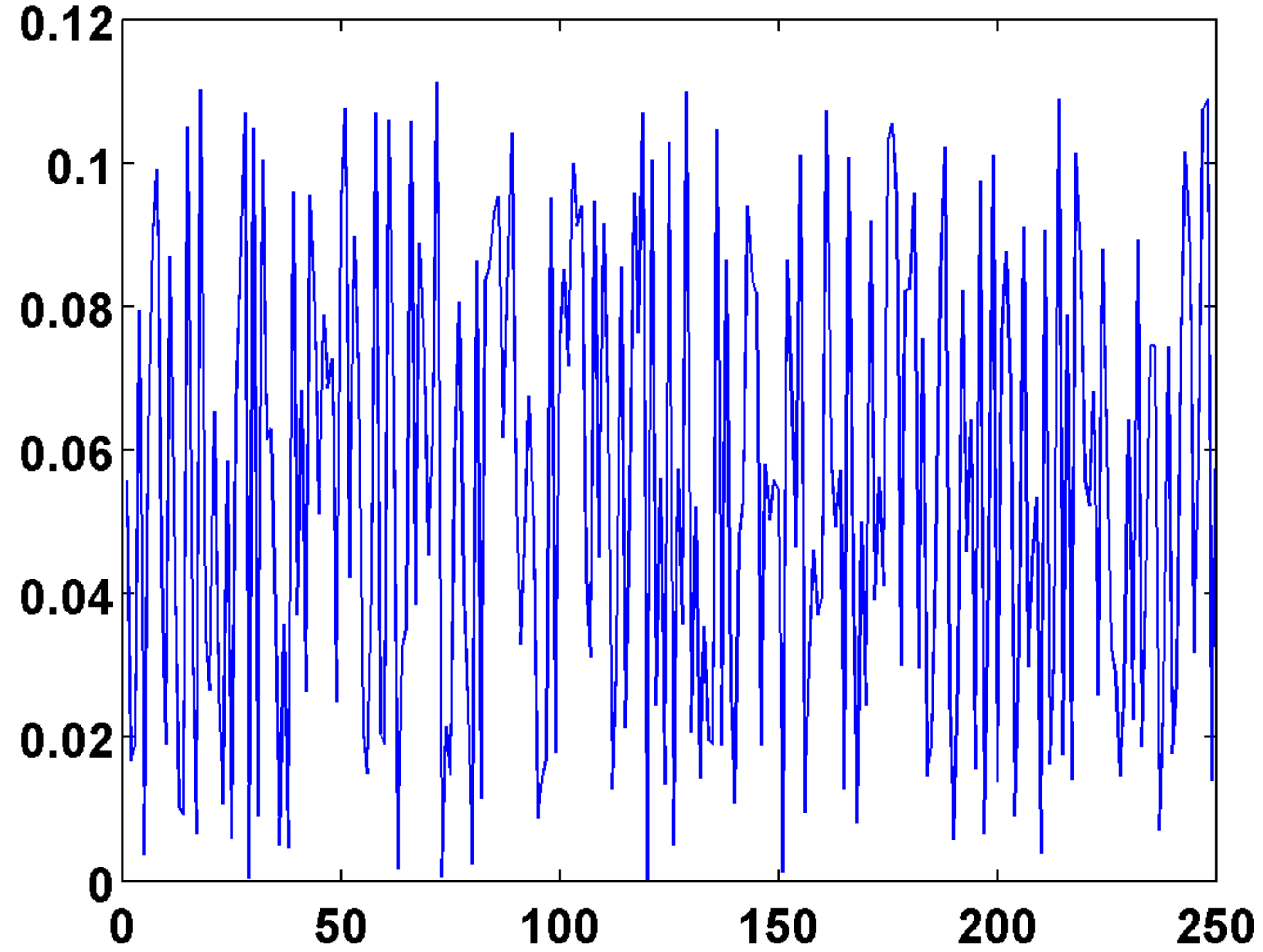
(included $\|\Gamma H\|_F^2$ to promote impulsive wavelet)

Random Initial Guess

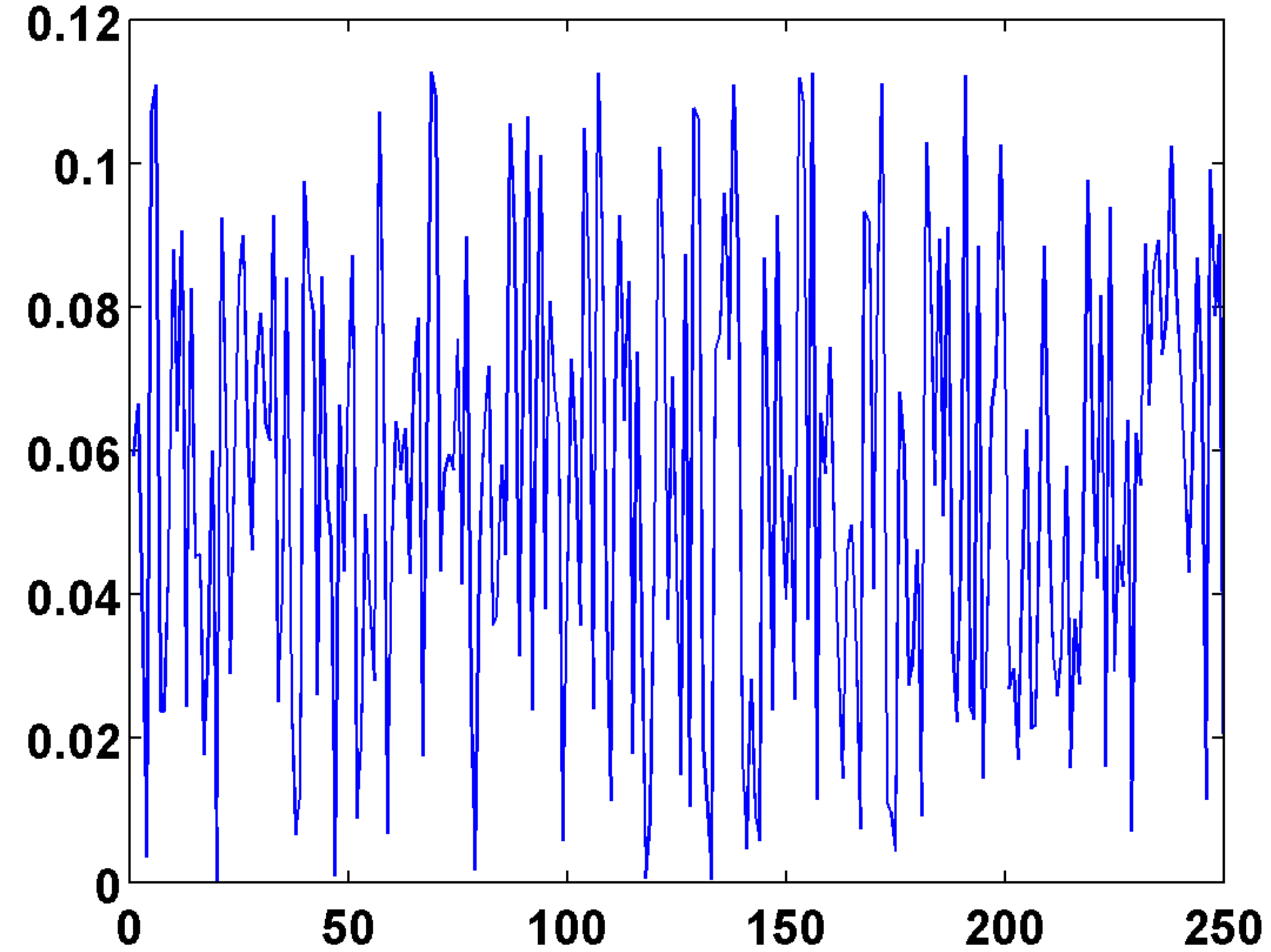
Initial h



Initial x_p

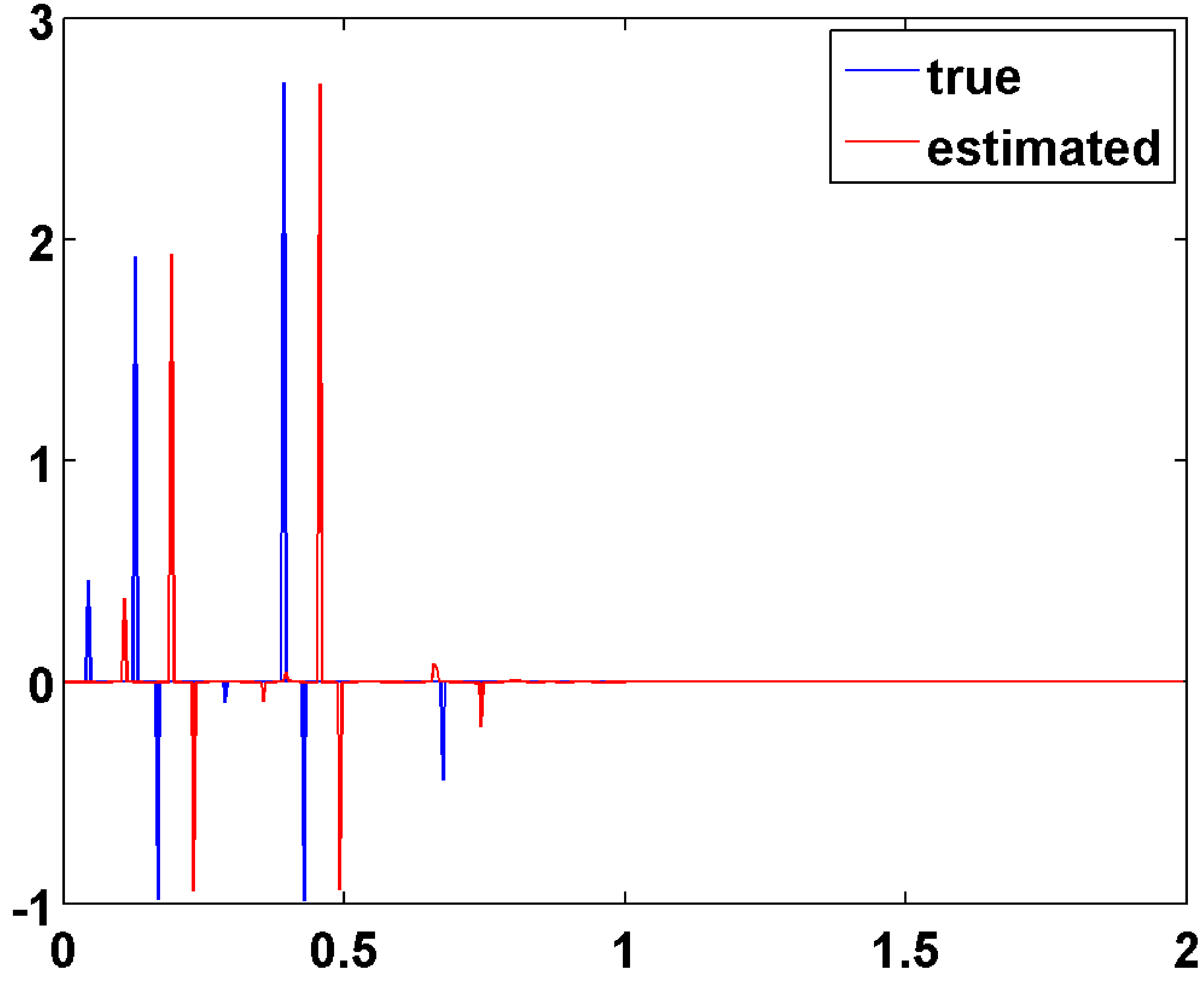


Initial x_m

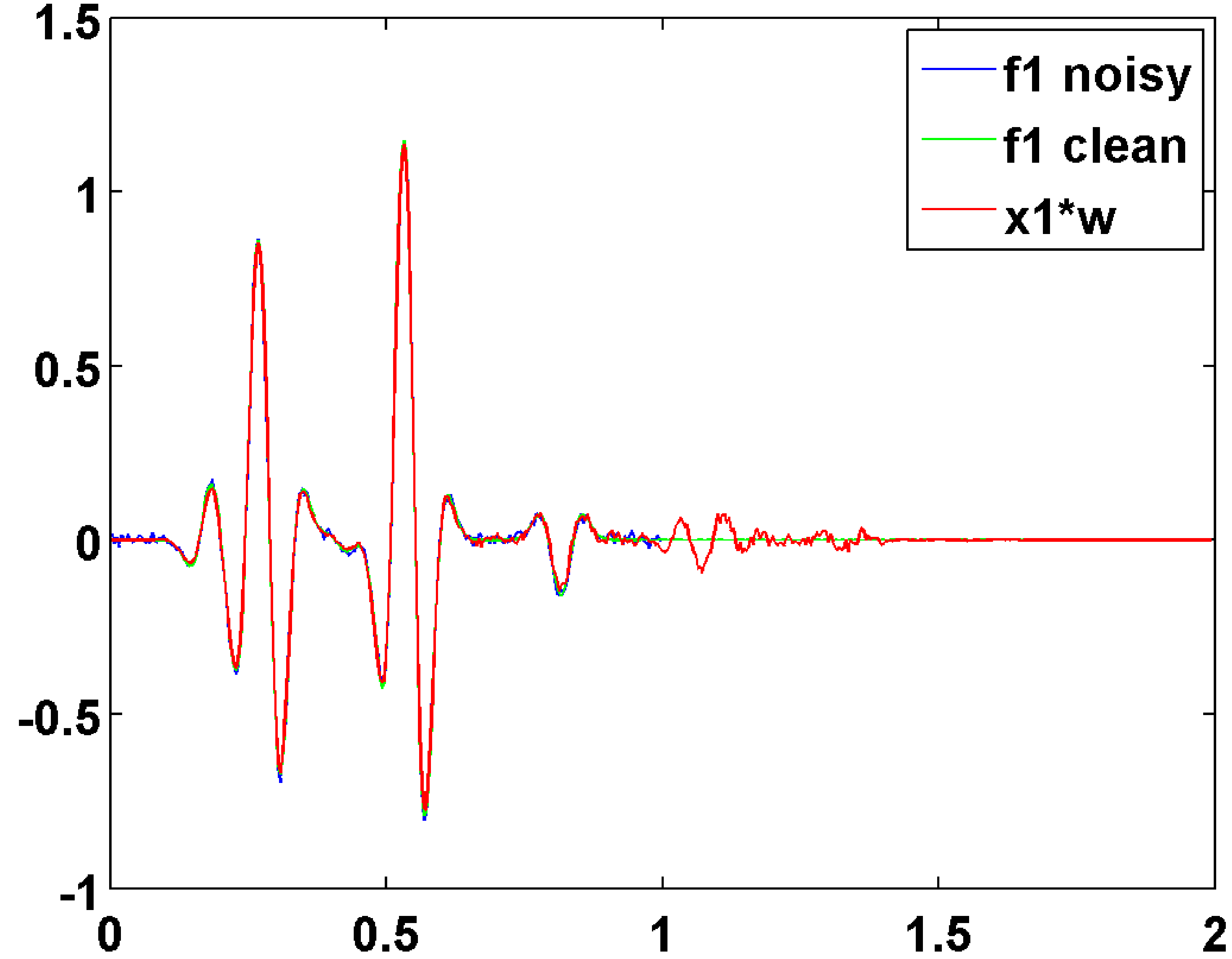


Recovered Sparse Signal for $n = 5$, $r = 1$, SNR $= 23.6$

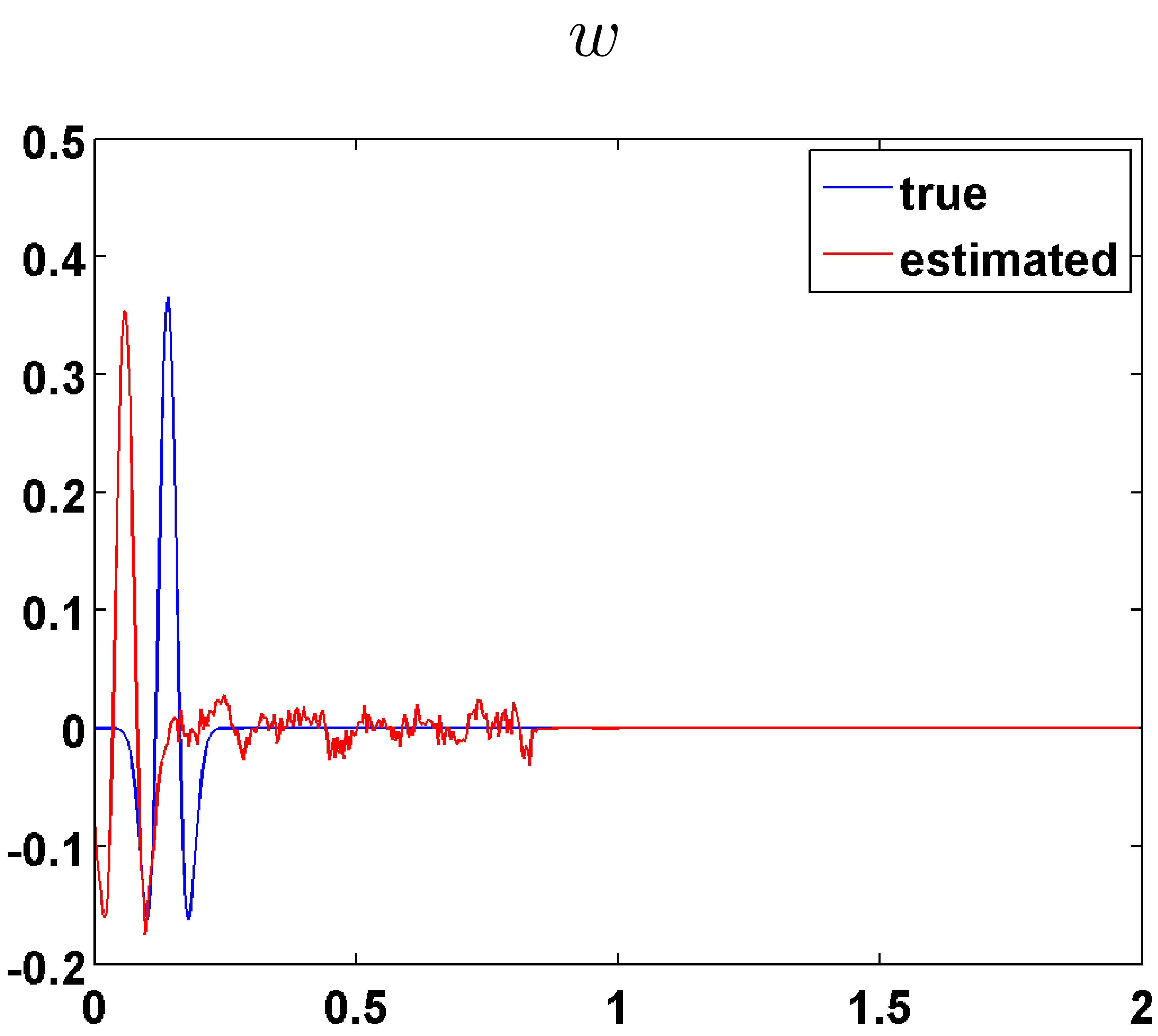
x_1

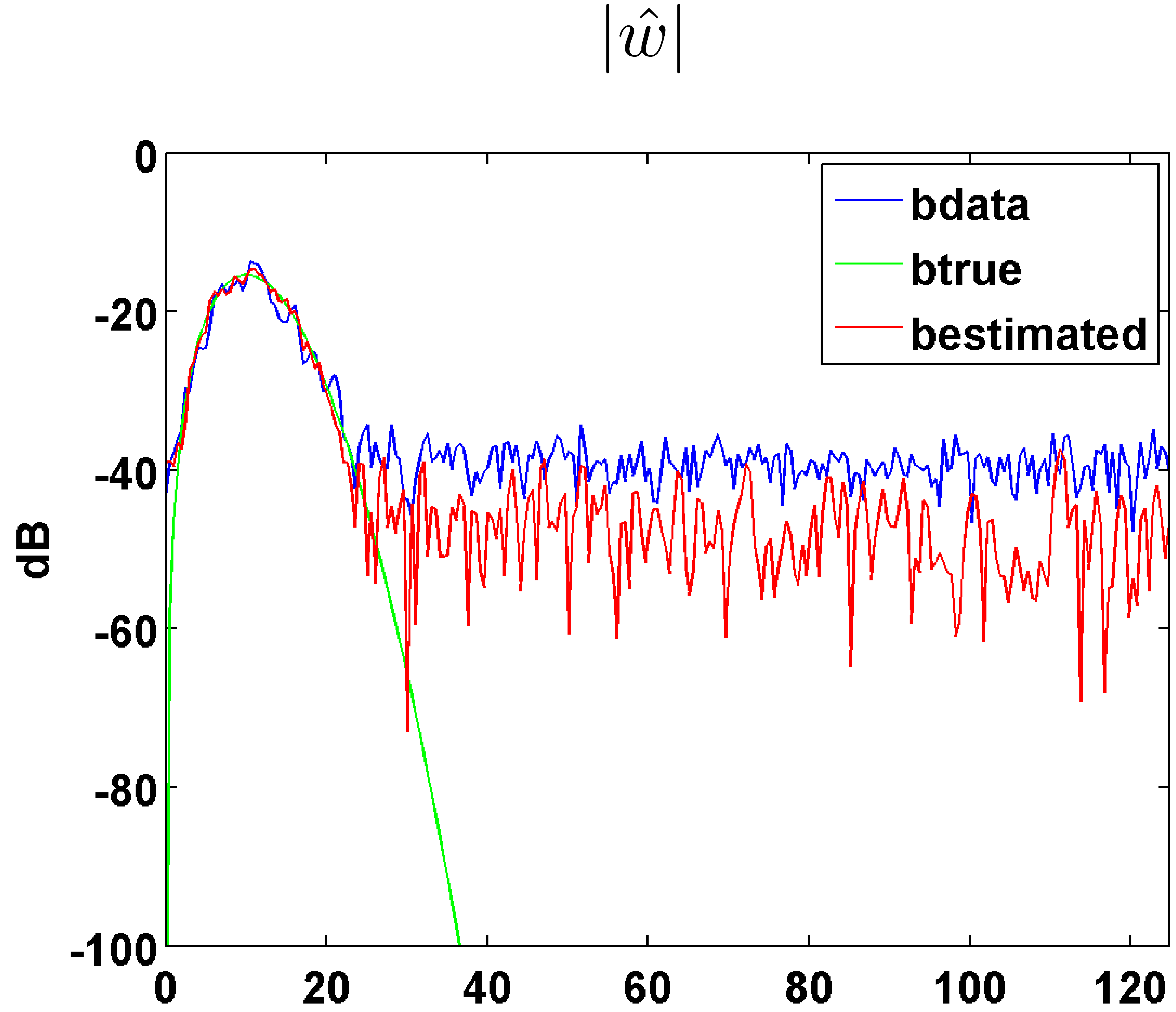


f_1



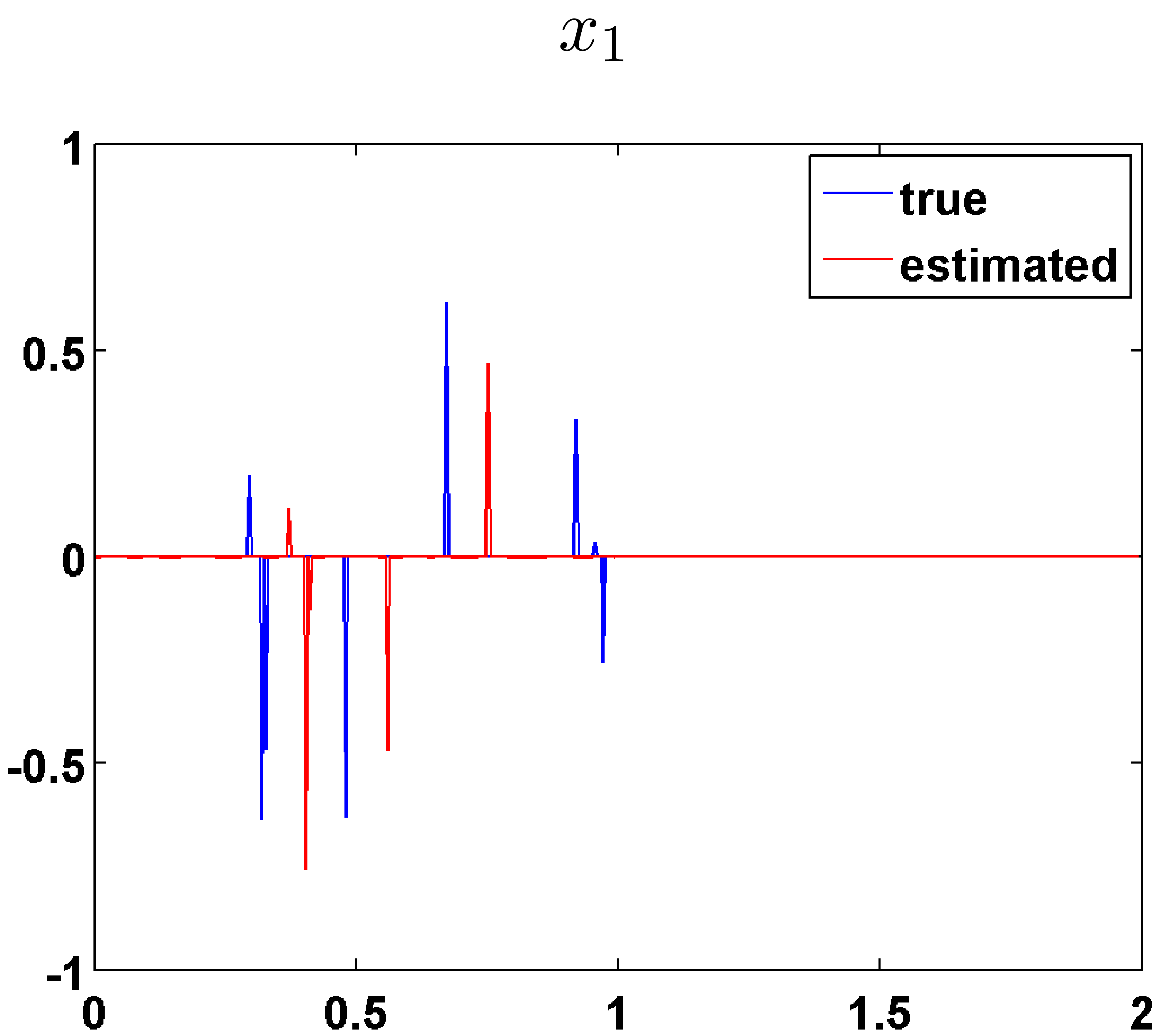
Recovered Wavelet for $n = 5$, $r = 1$, SNR $= 13.5$

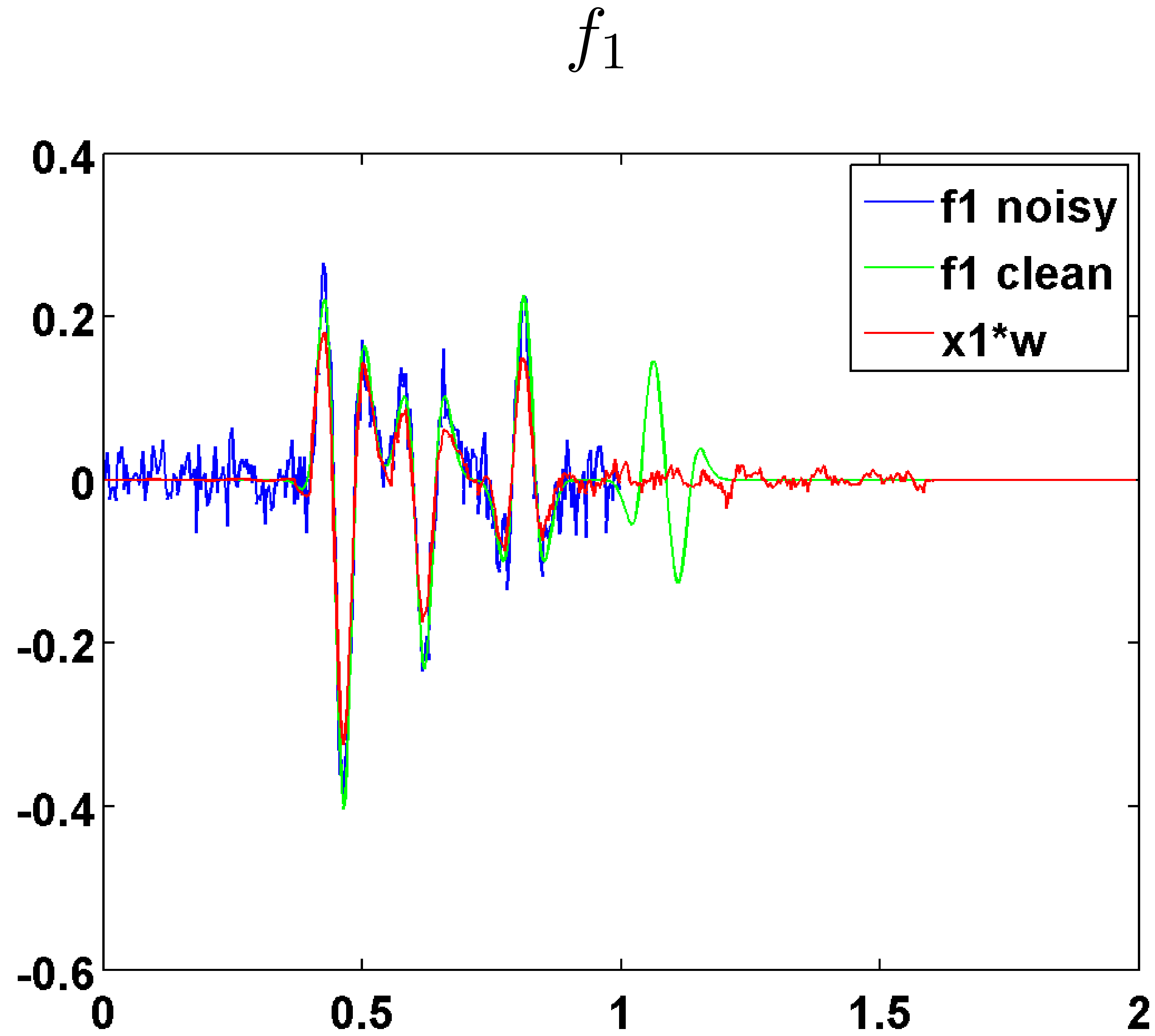




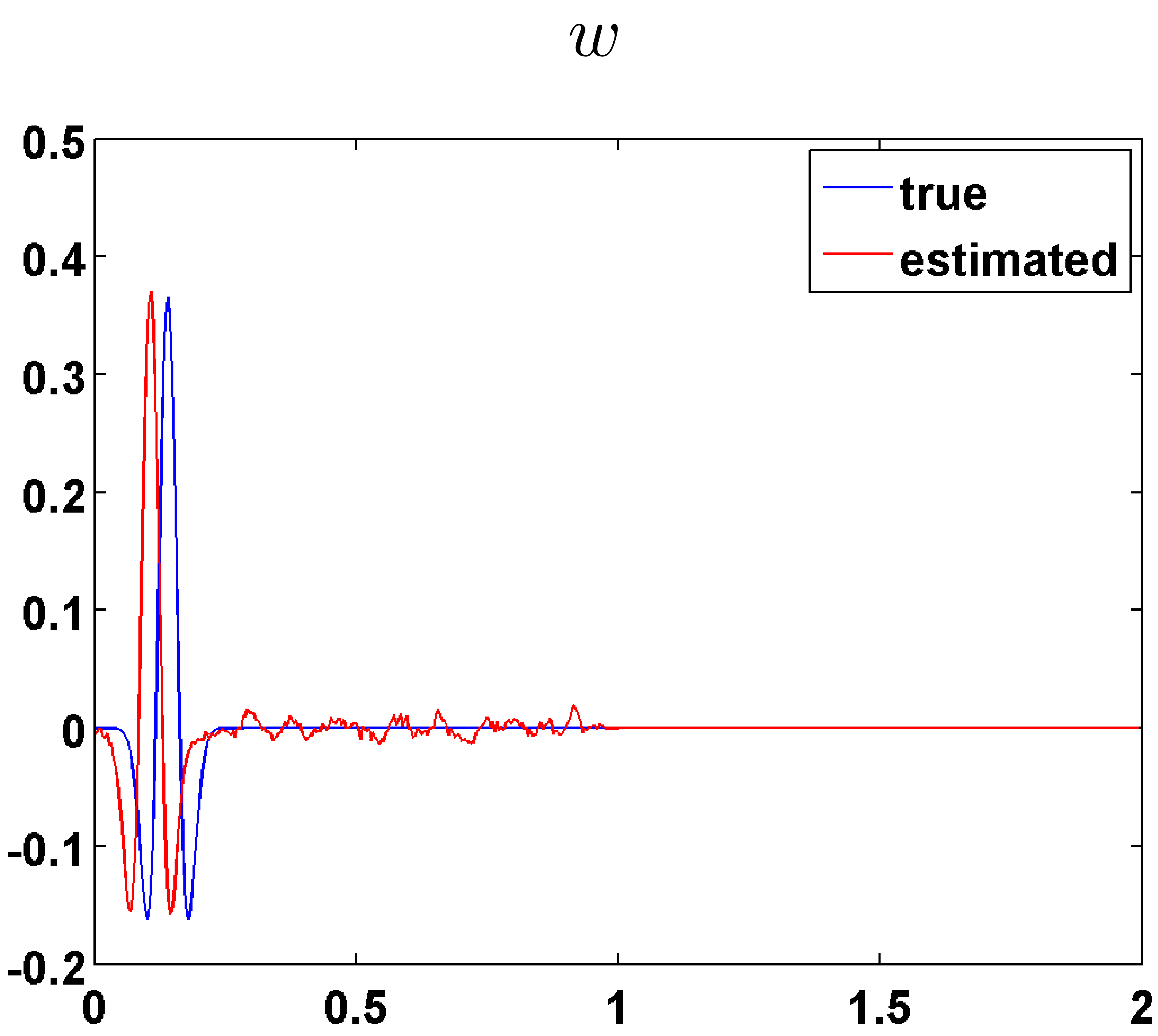
(included $\|\Gamma H\|_F^2$ to promote impulsive wavelet)

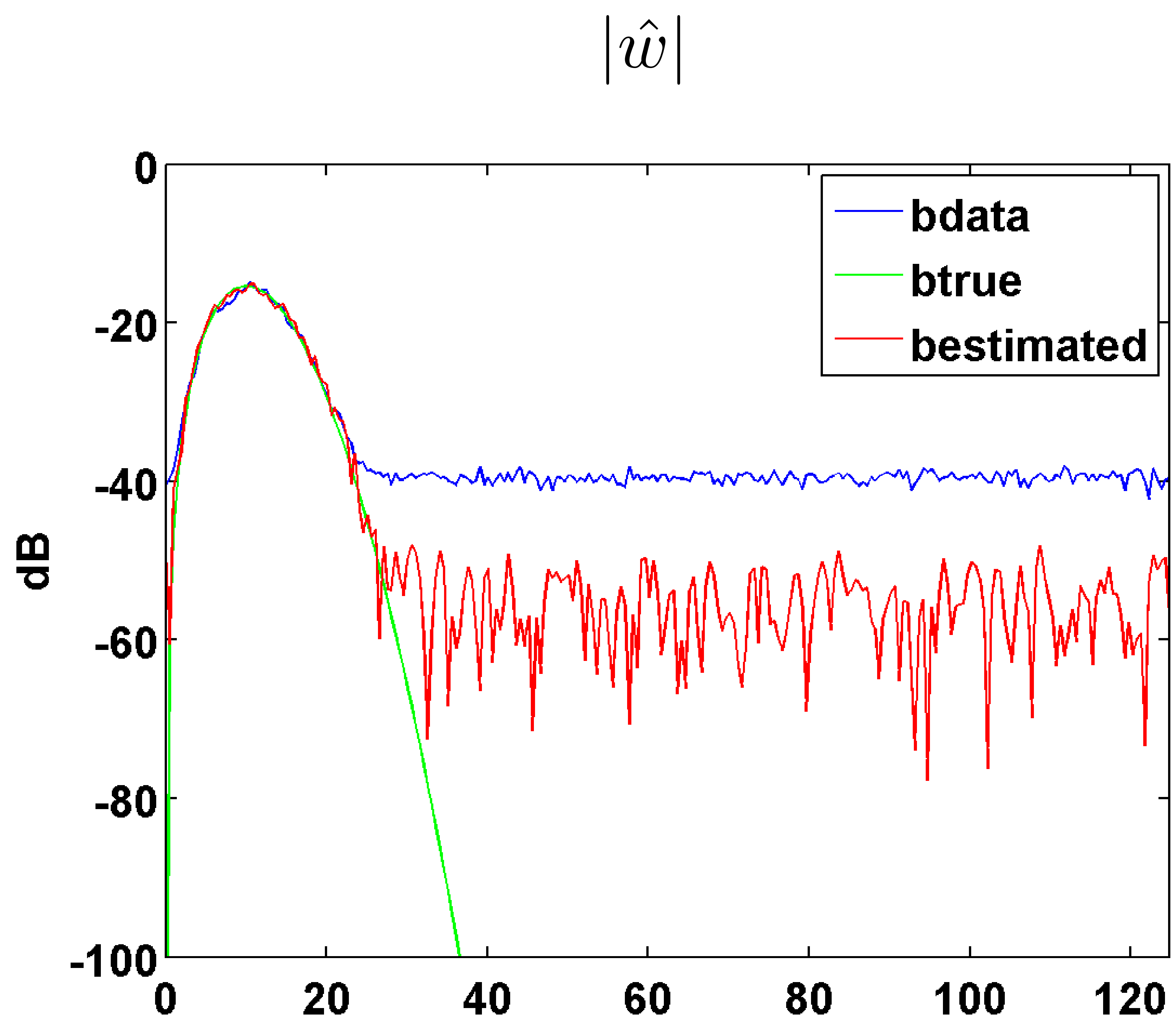
Recovered Sparse Signal for $n = 5$, $r = 1$, SNR $= 13.5$



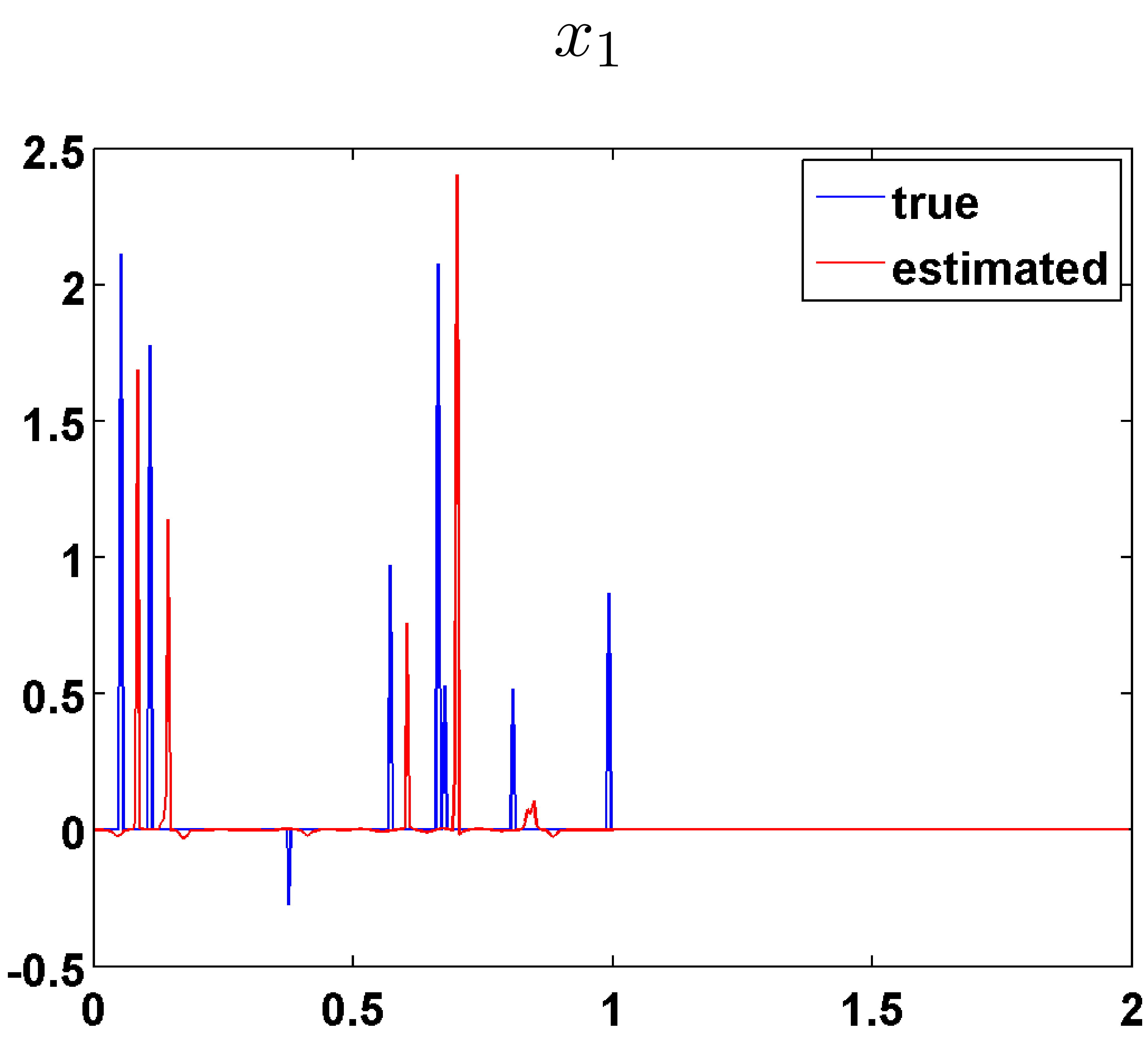


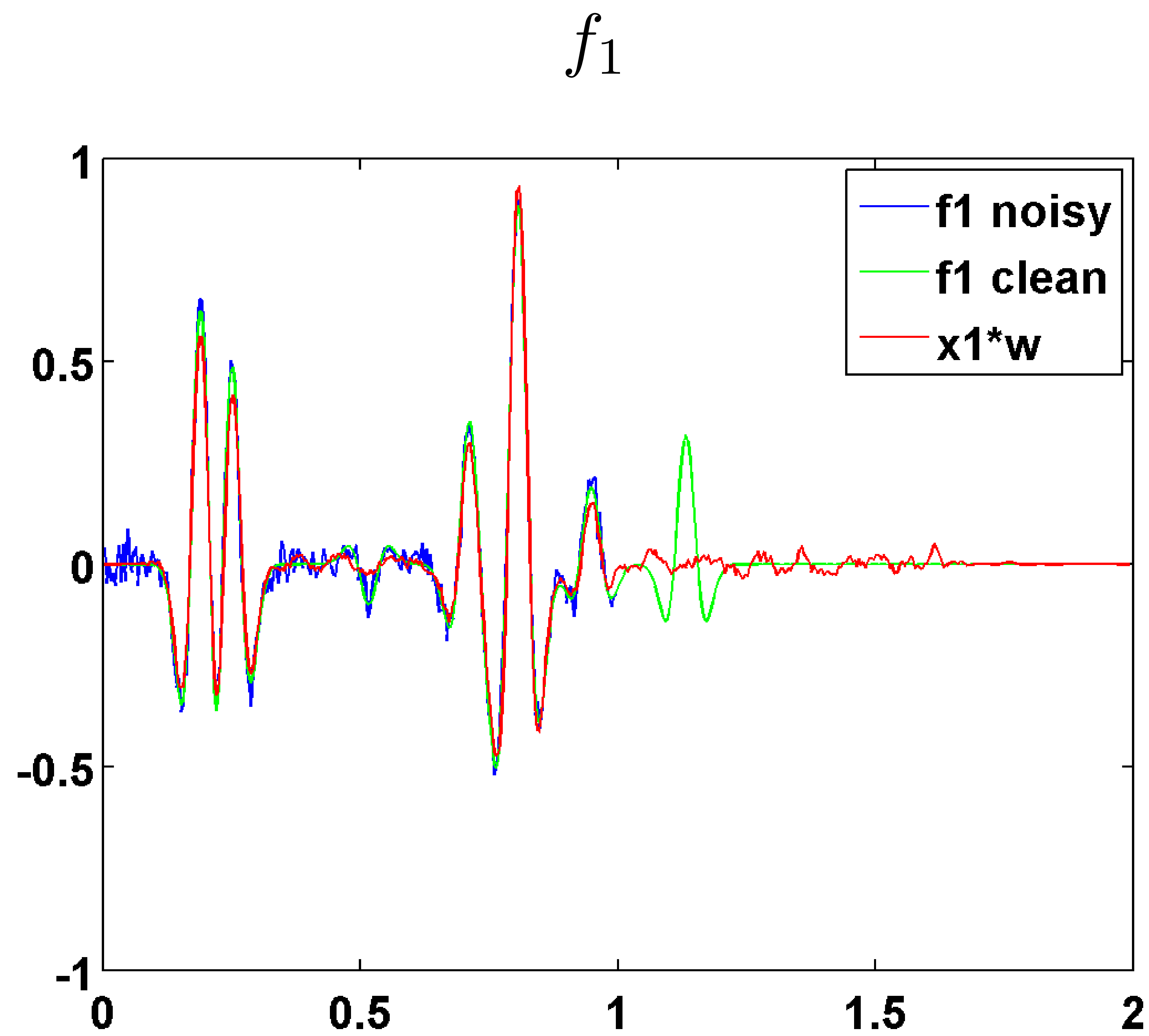
Recovered Wavelet for $n = 50$, $r = 1$, SNR $= 13.5$



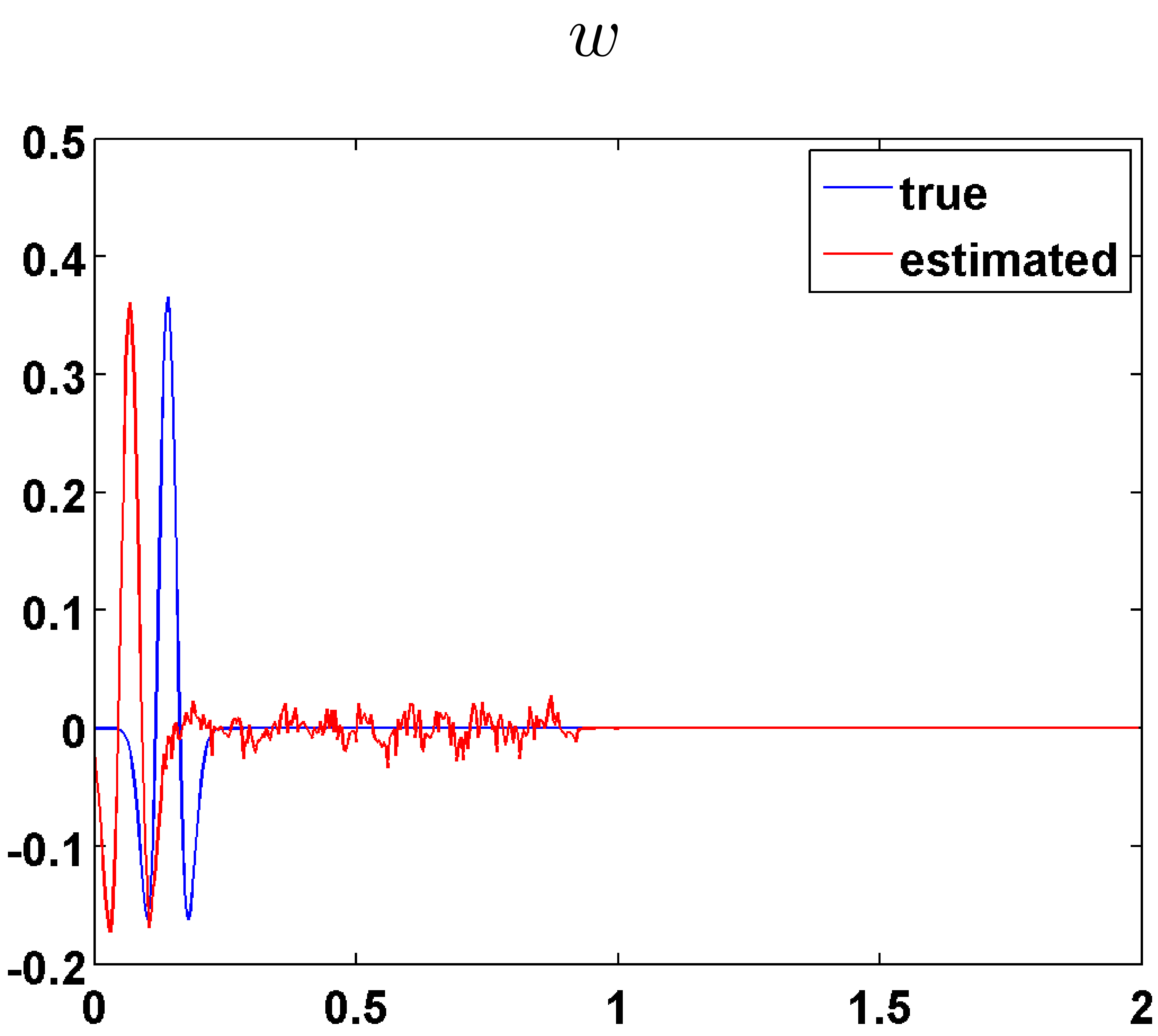


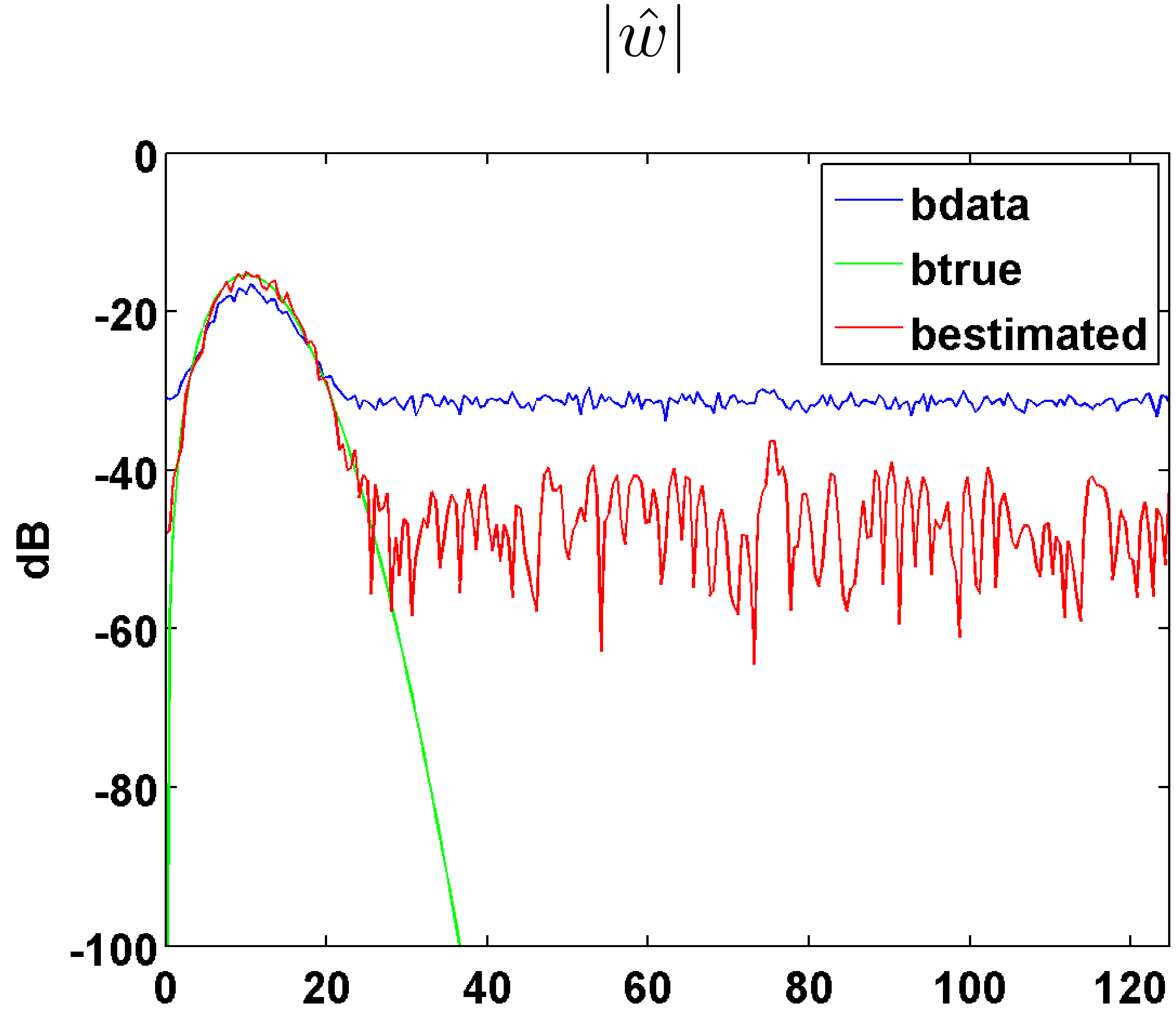
Recovered Sparse Signal for $n = 50$, $r = 1$, SNR $= 13.5$



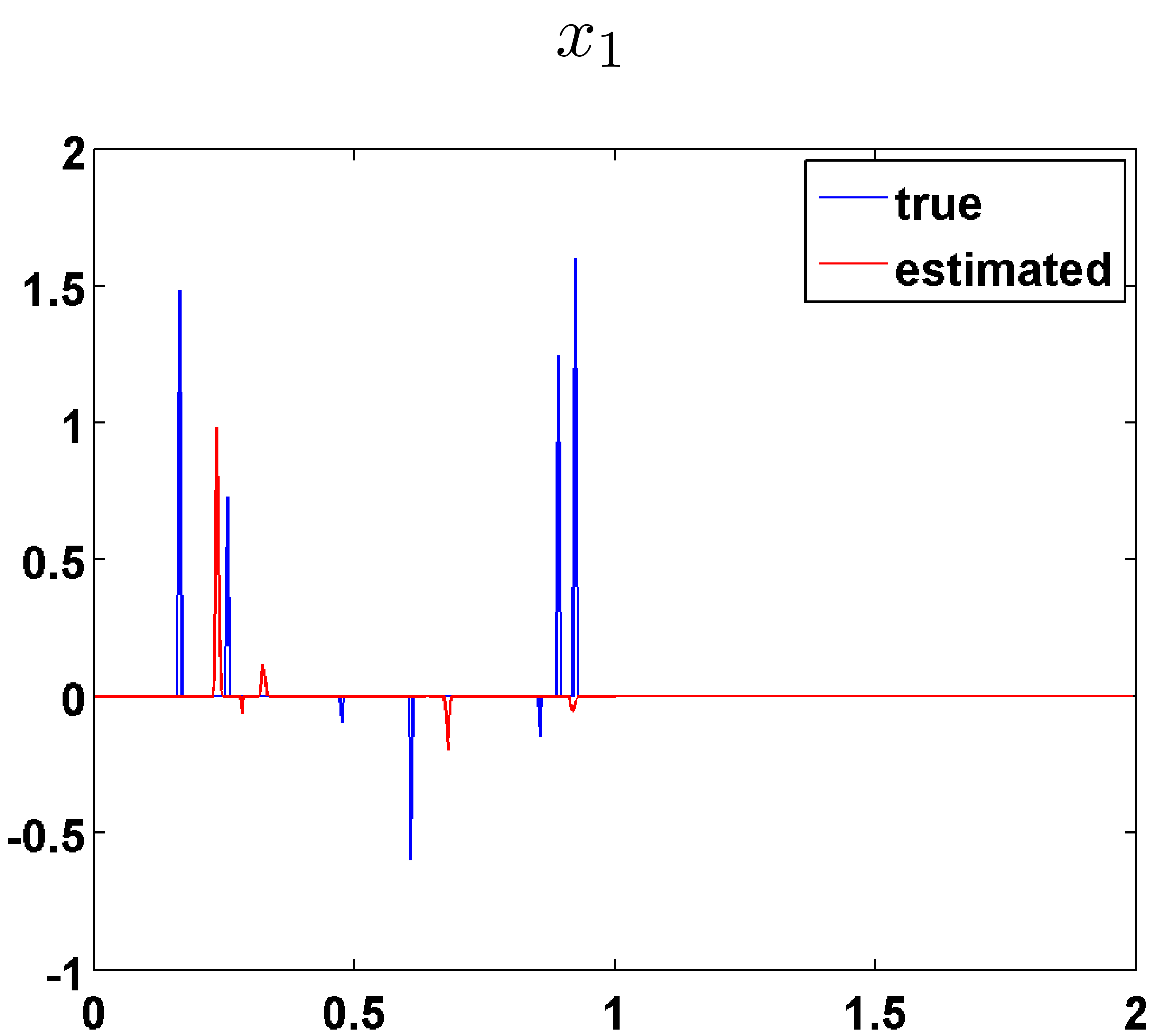


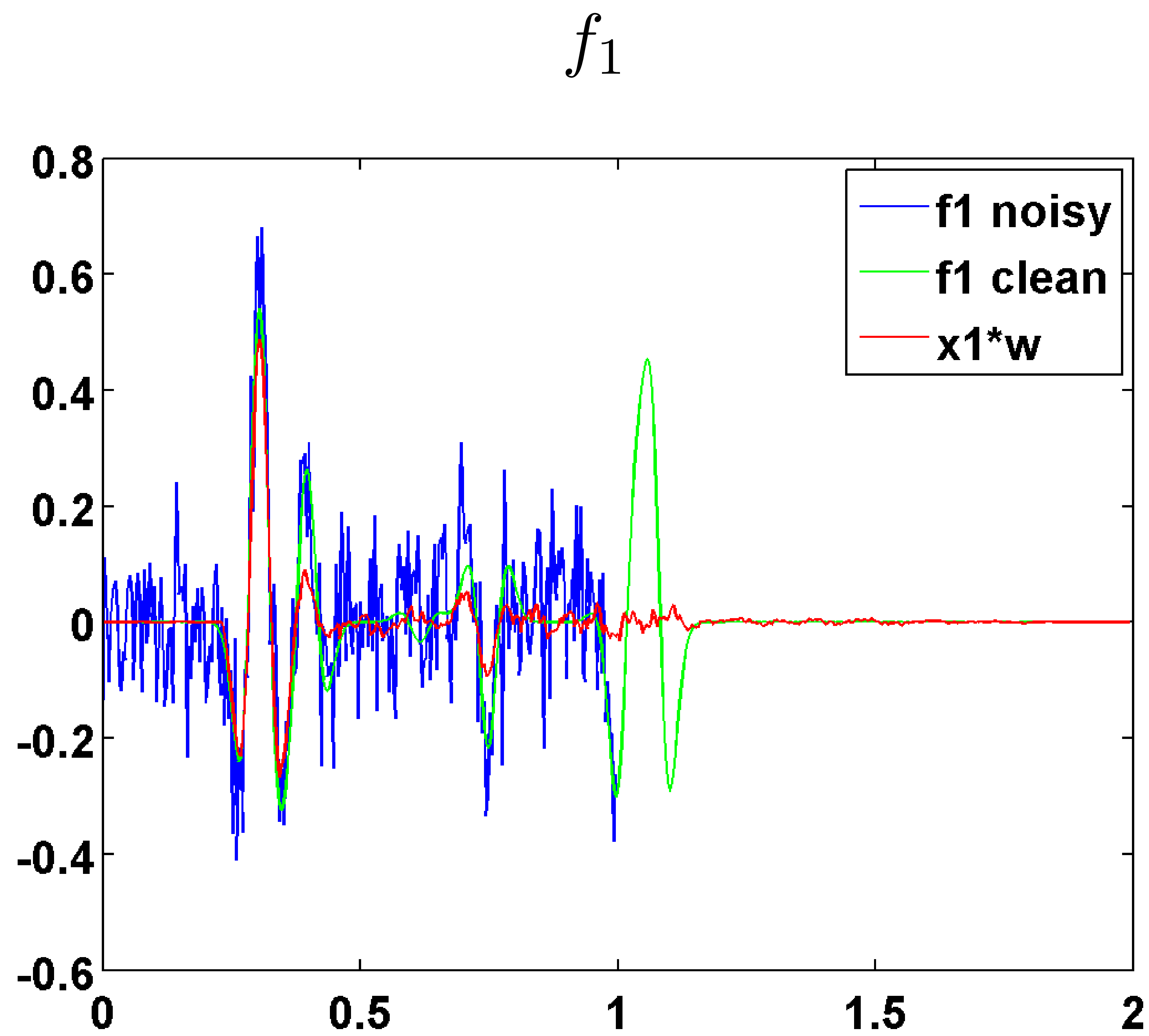
Recovered Wavelet for $n = 50$, $r = 1$, SNR $= 5.25$





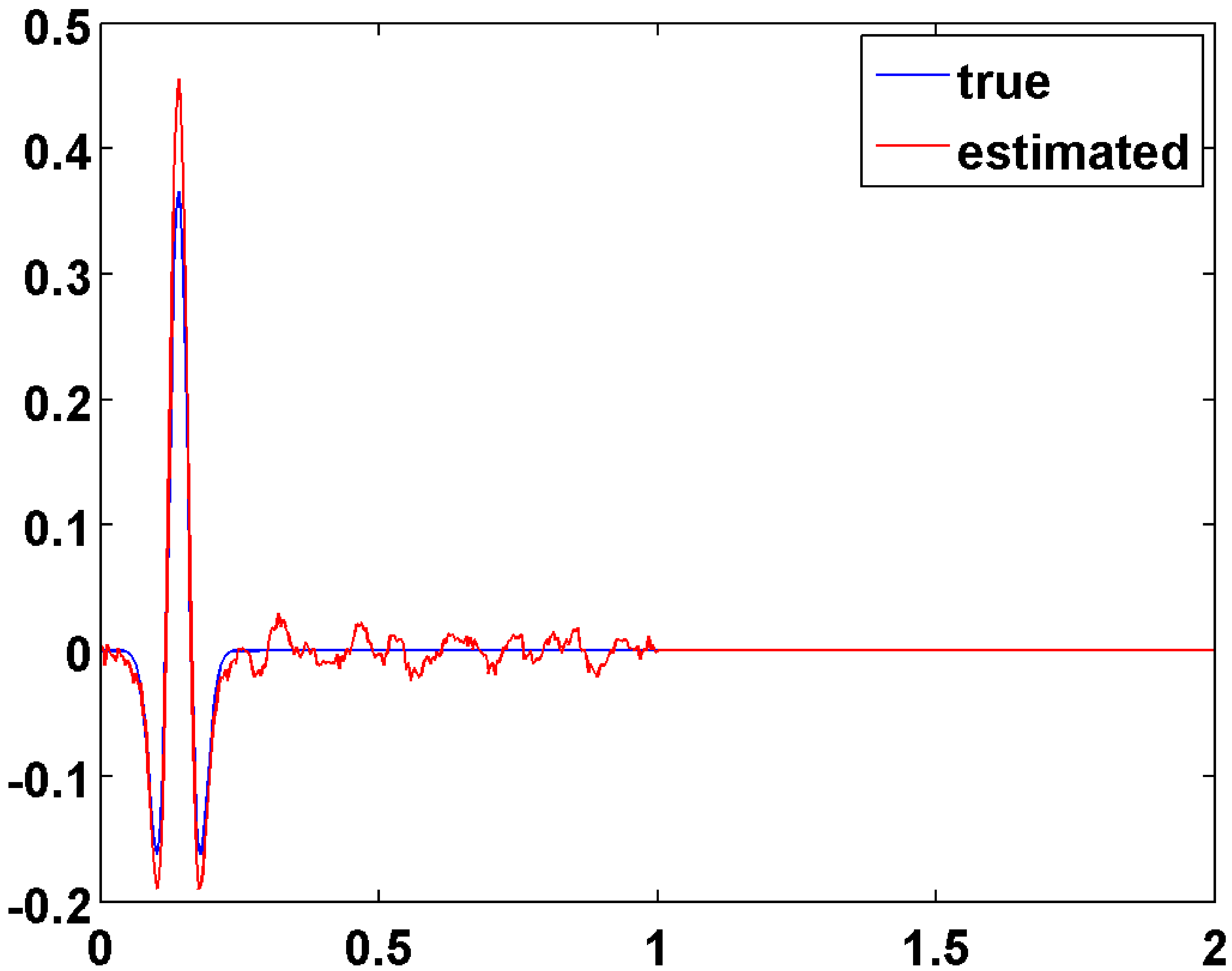
Recovered Sparse Signal for $n = 50$, $r = 1$, SNR $= 5.25$



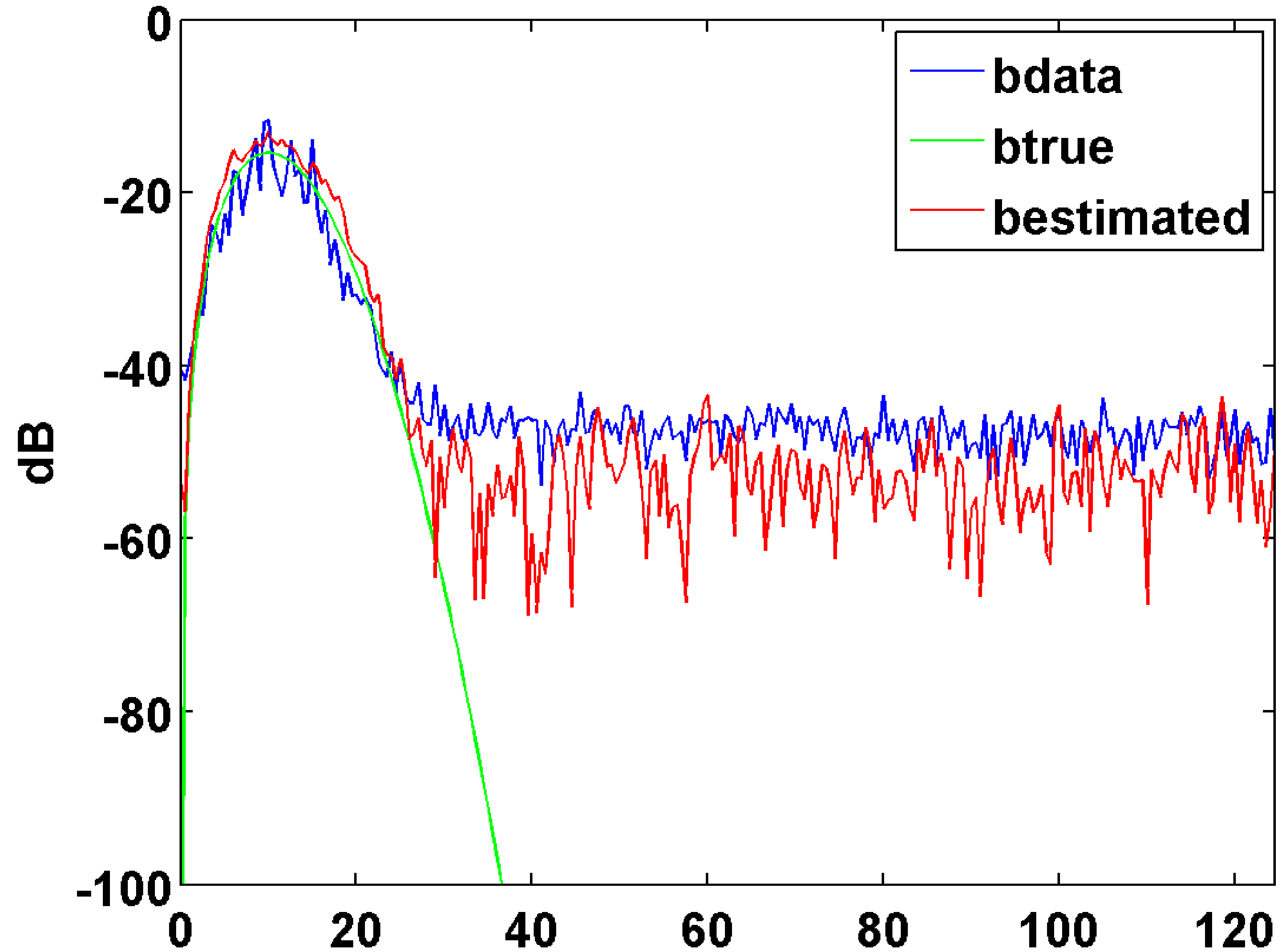


Feedback system: Recovered Wavelet for $n = 5$, $r = 1$, SNR $= 20.8$

w

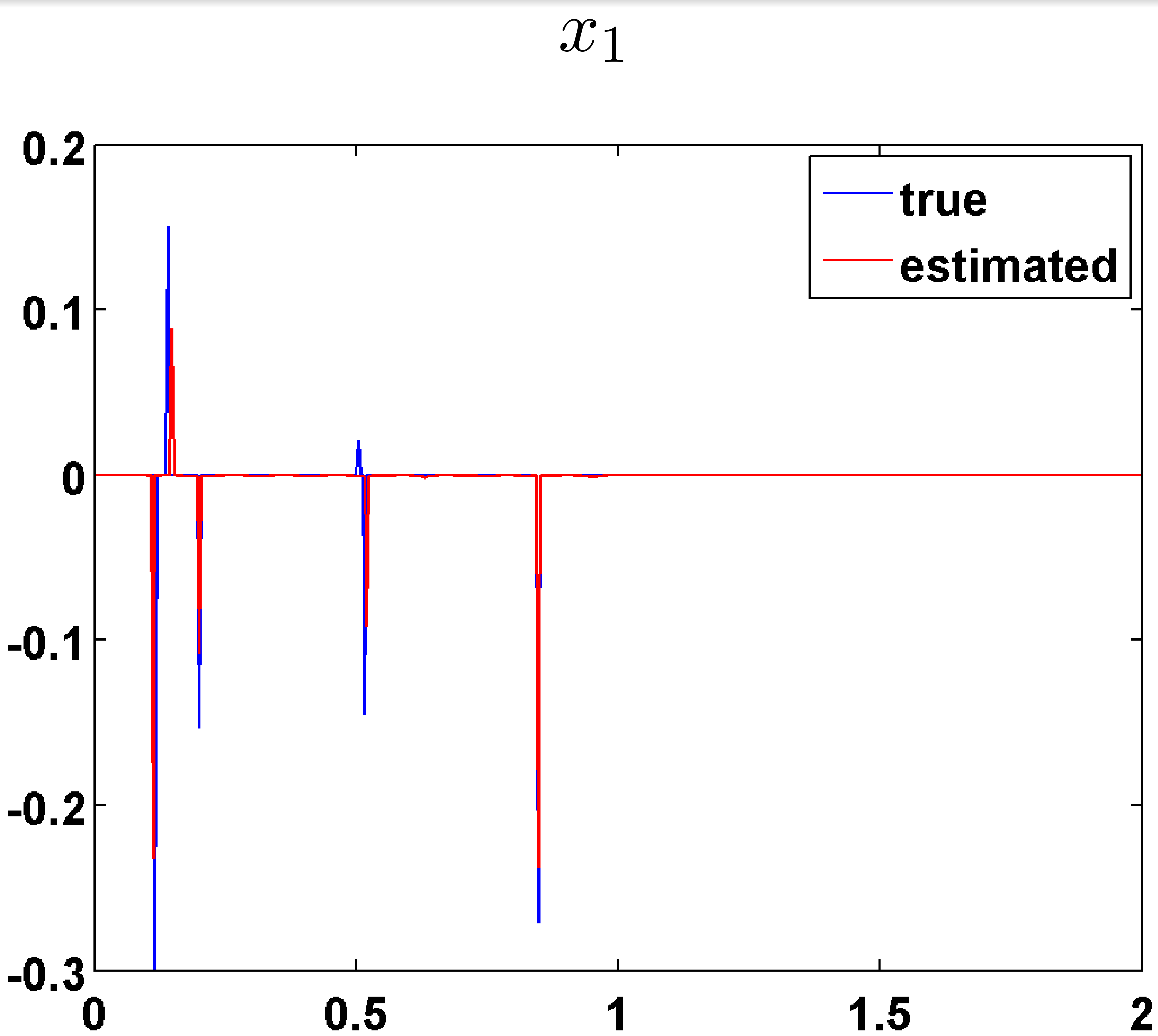


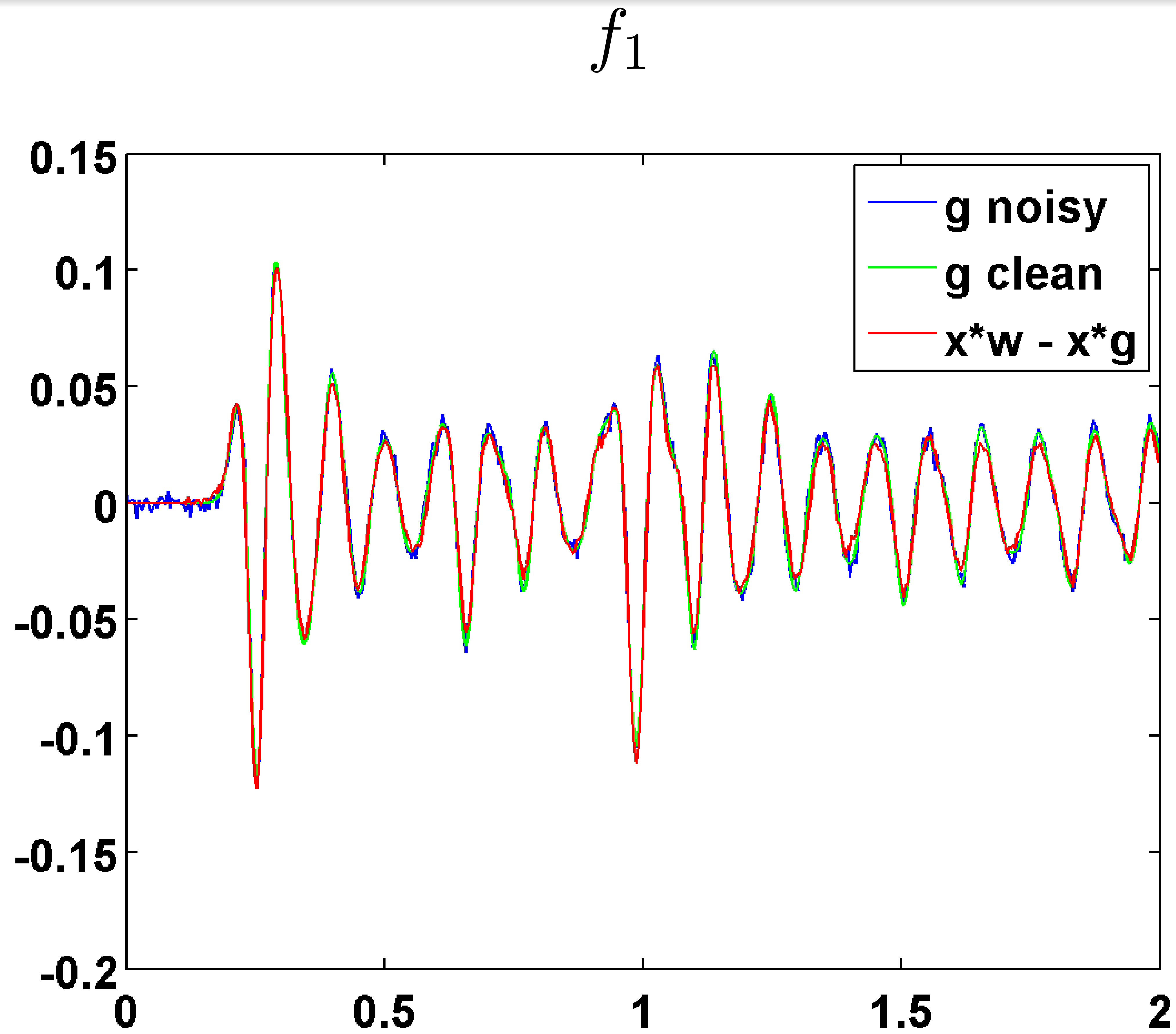
$|\hat{w}|$



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

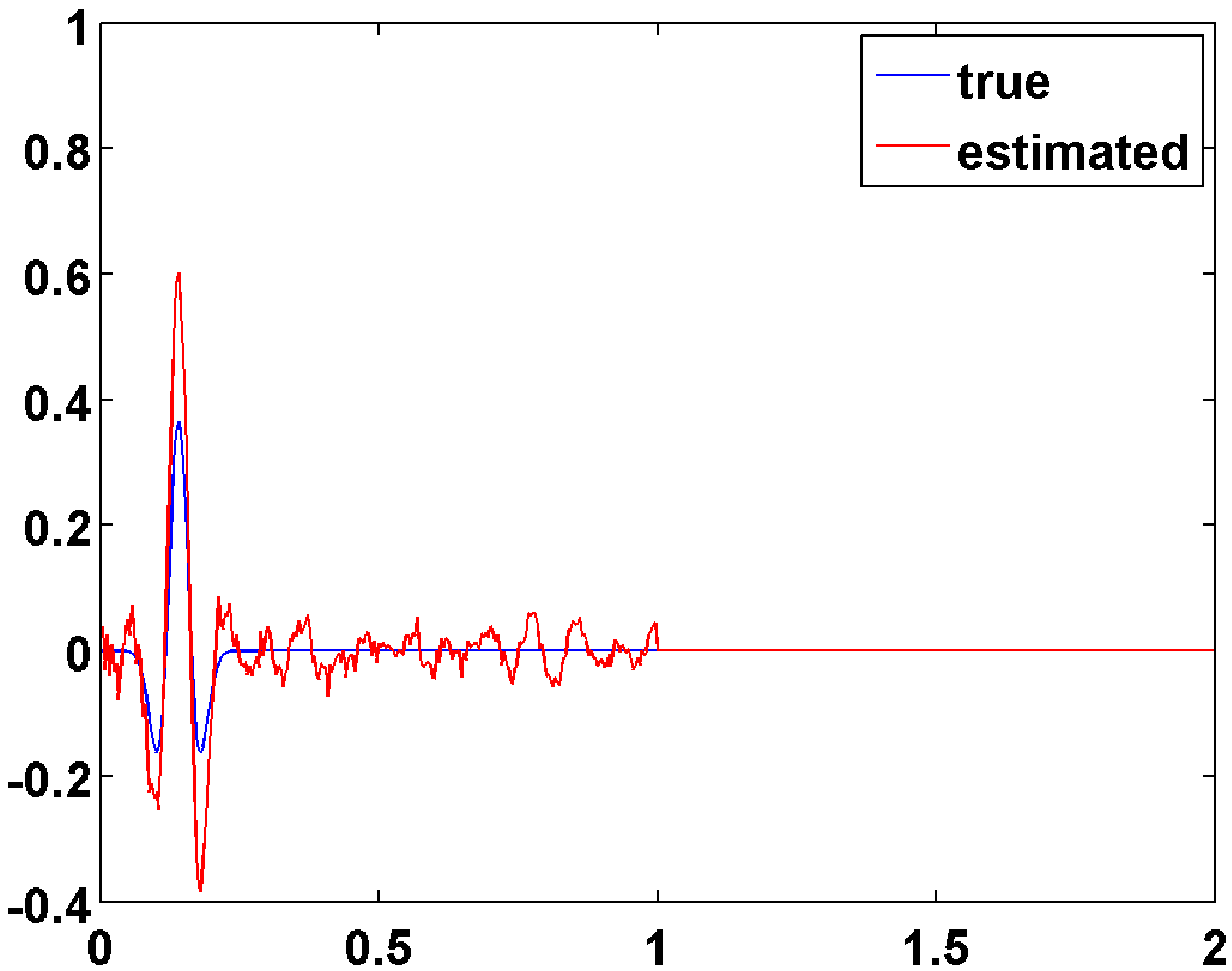
Feedback system: Recovered Sparse Signal for $n = 5$, $r = 1$, SNR $= 20.8$



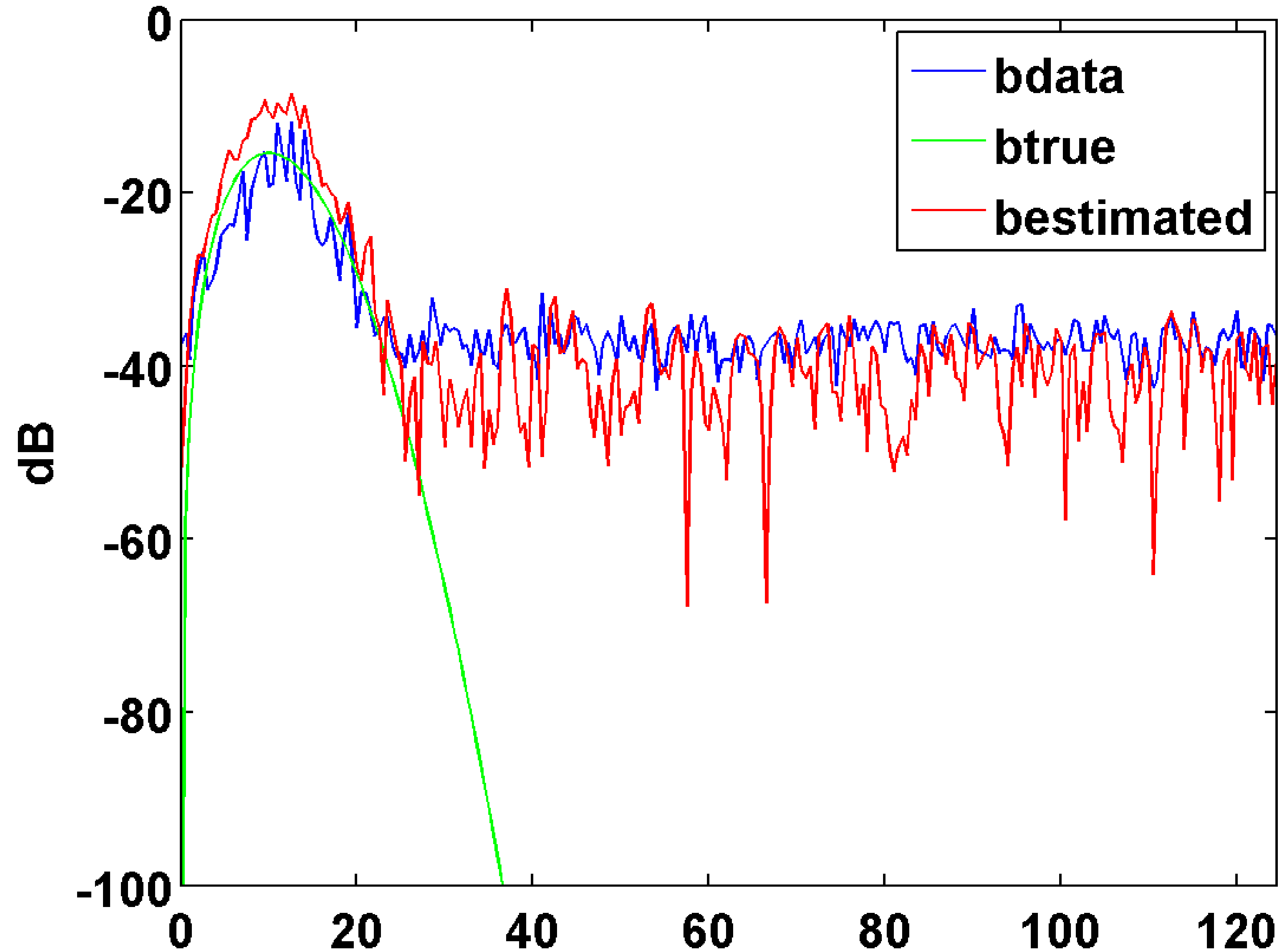


Feedback system: Recovered Wavelet for $n = 5$, $r = 1$, SNR $= 9.82$

w



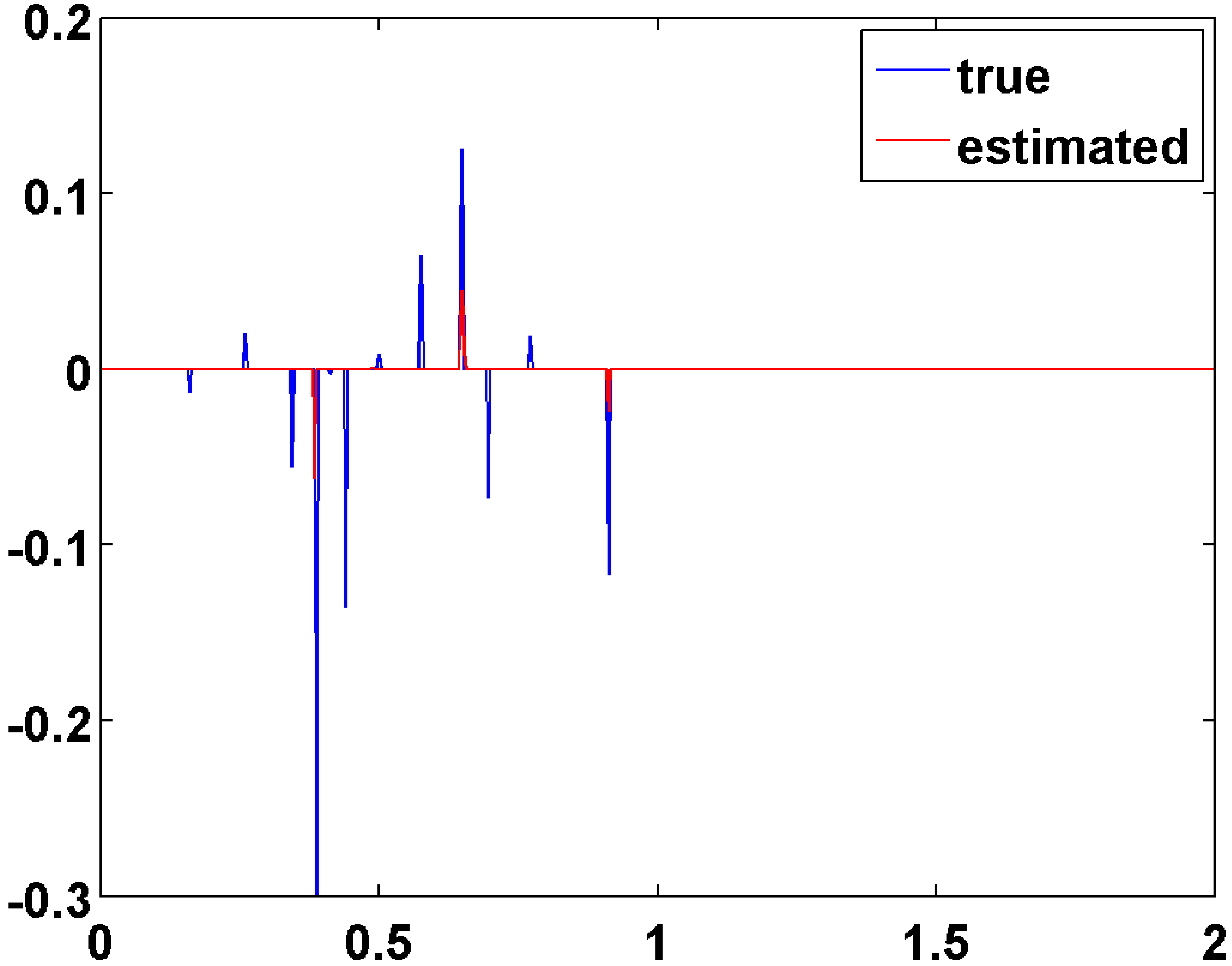
$|\hat{w}|$



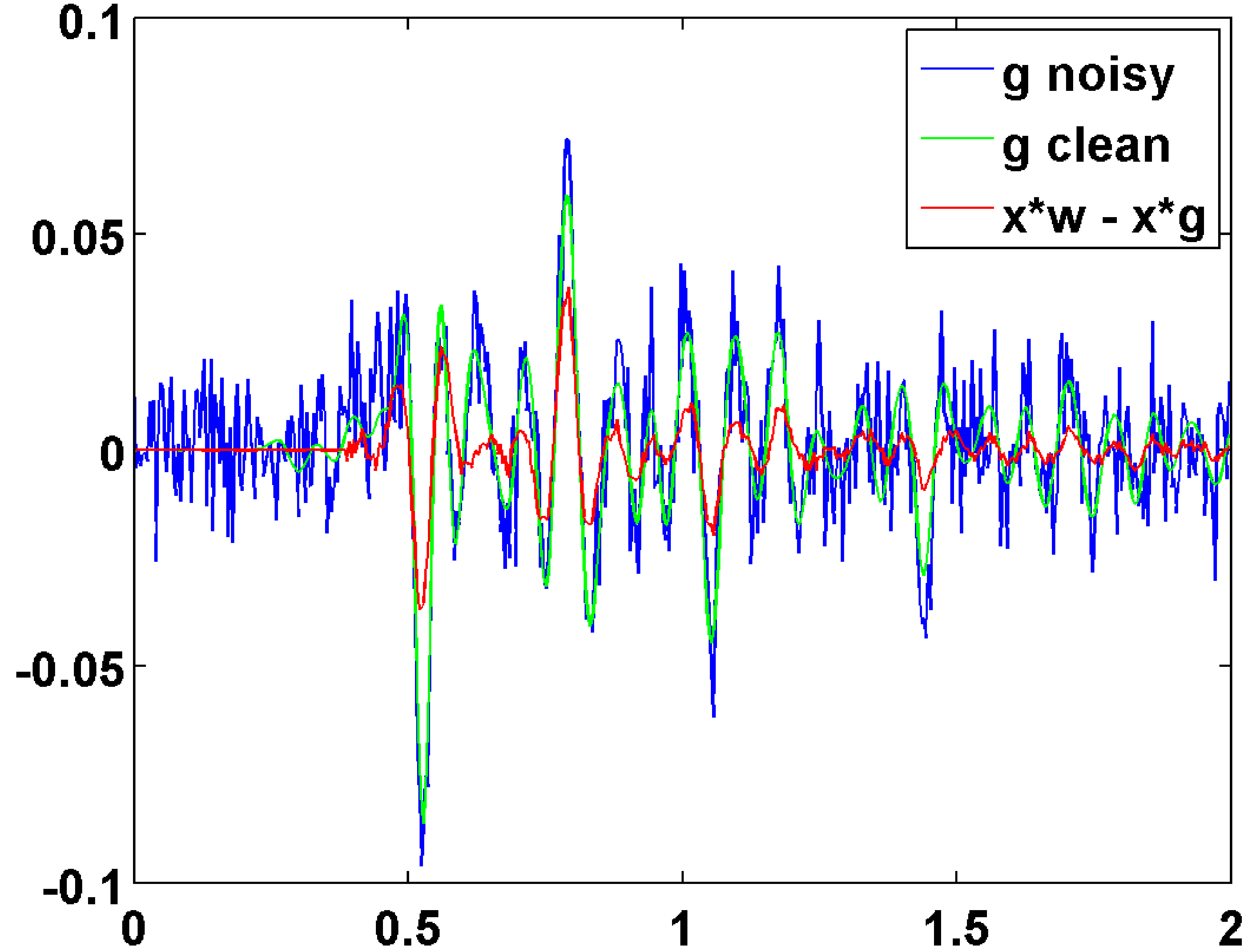
(included $\|\Gamma H\|_F^2$ to promote impulsive wavelet)

Feedback system: Recovered Sparse Signal for $n = 5$, $r = 1$, SNR $= 9.82$

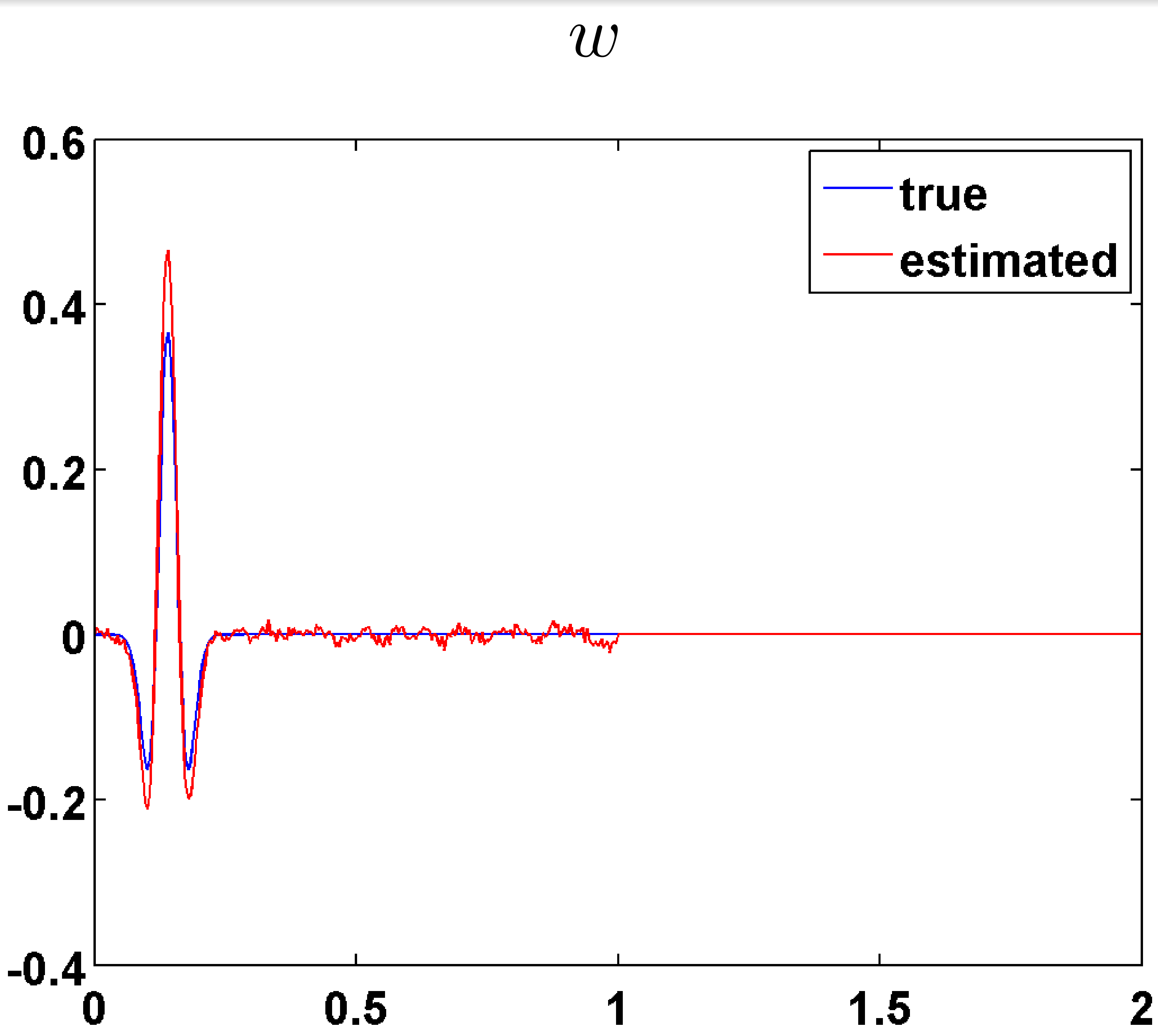
x_1

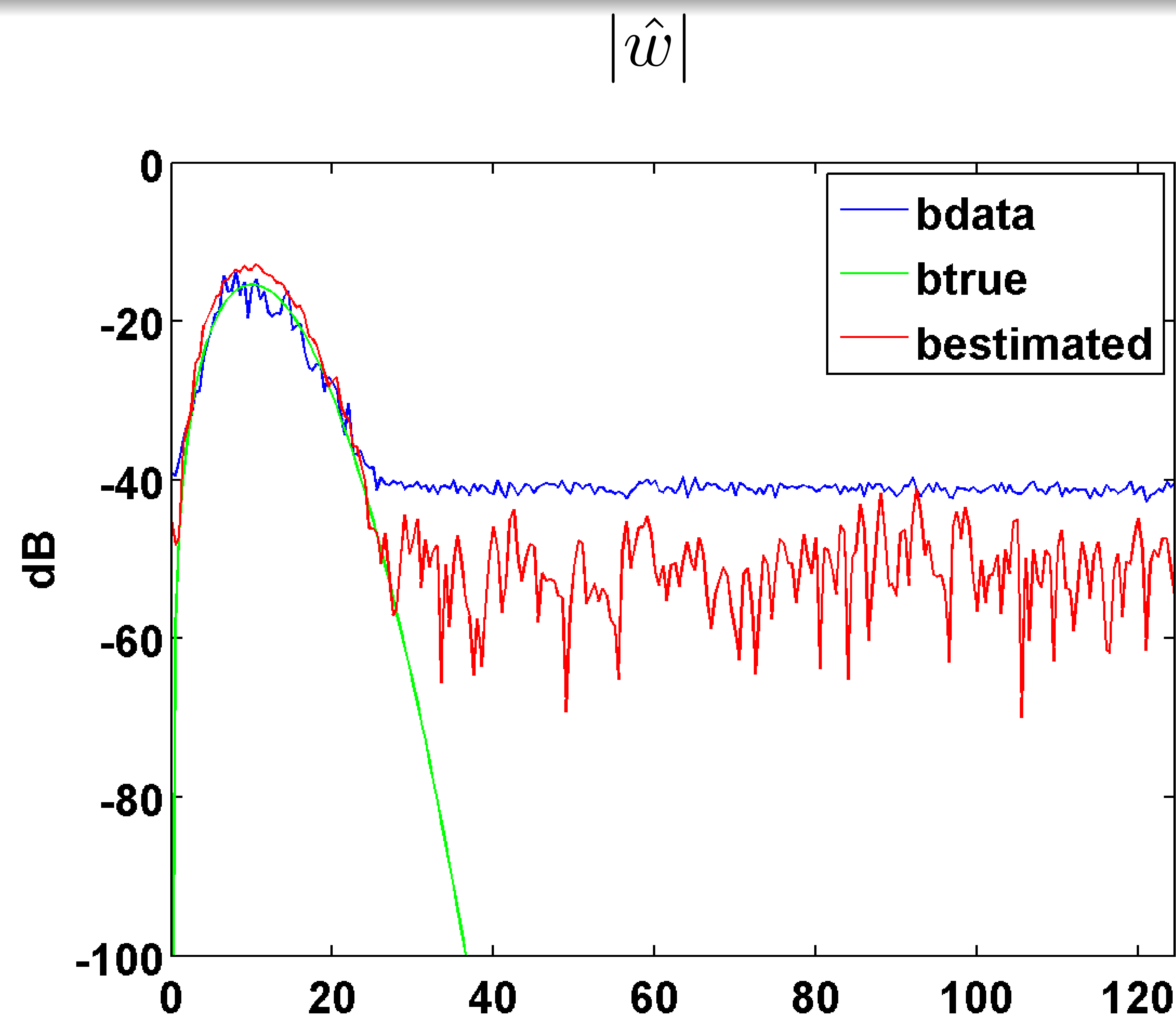


f_1

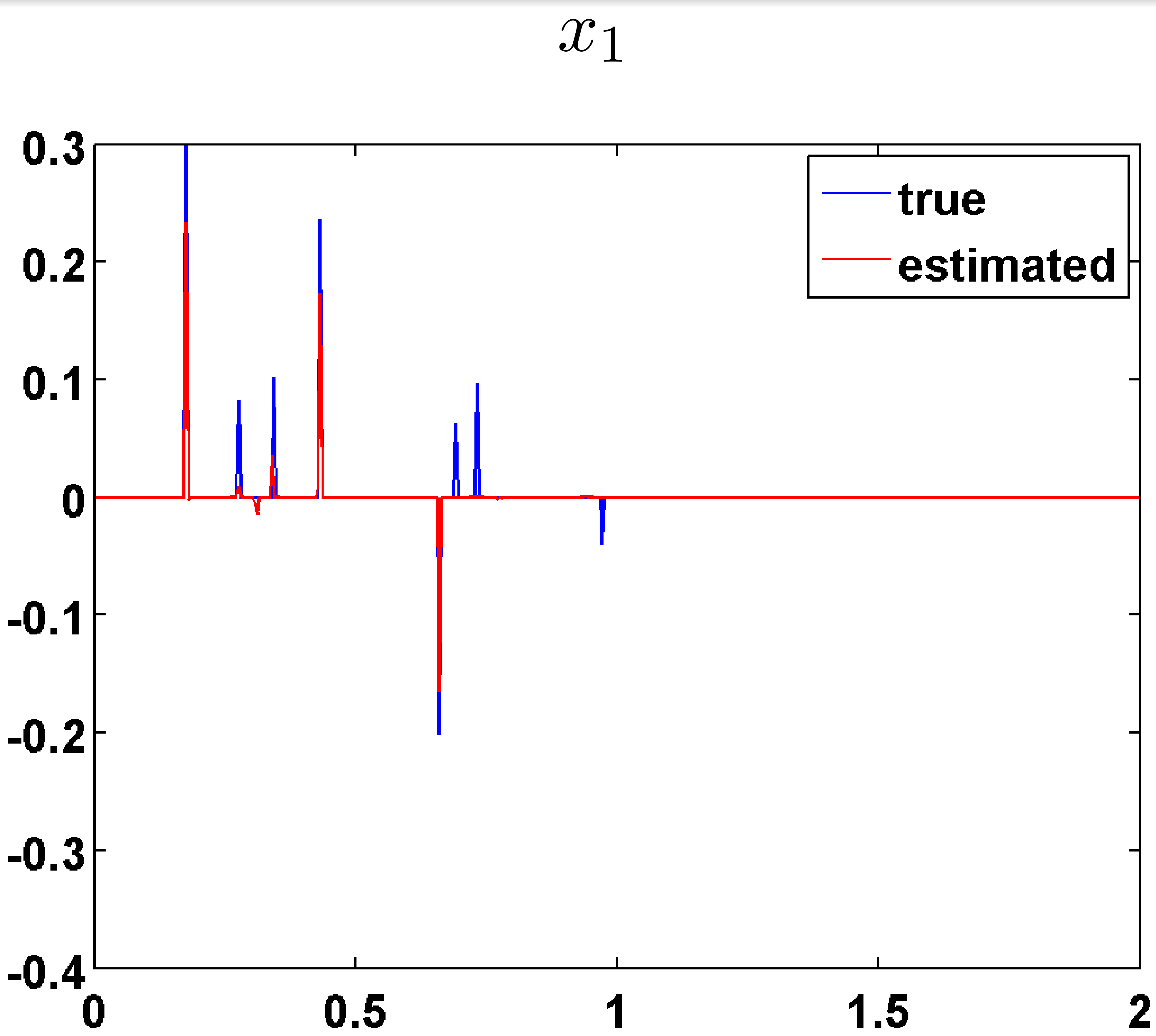


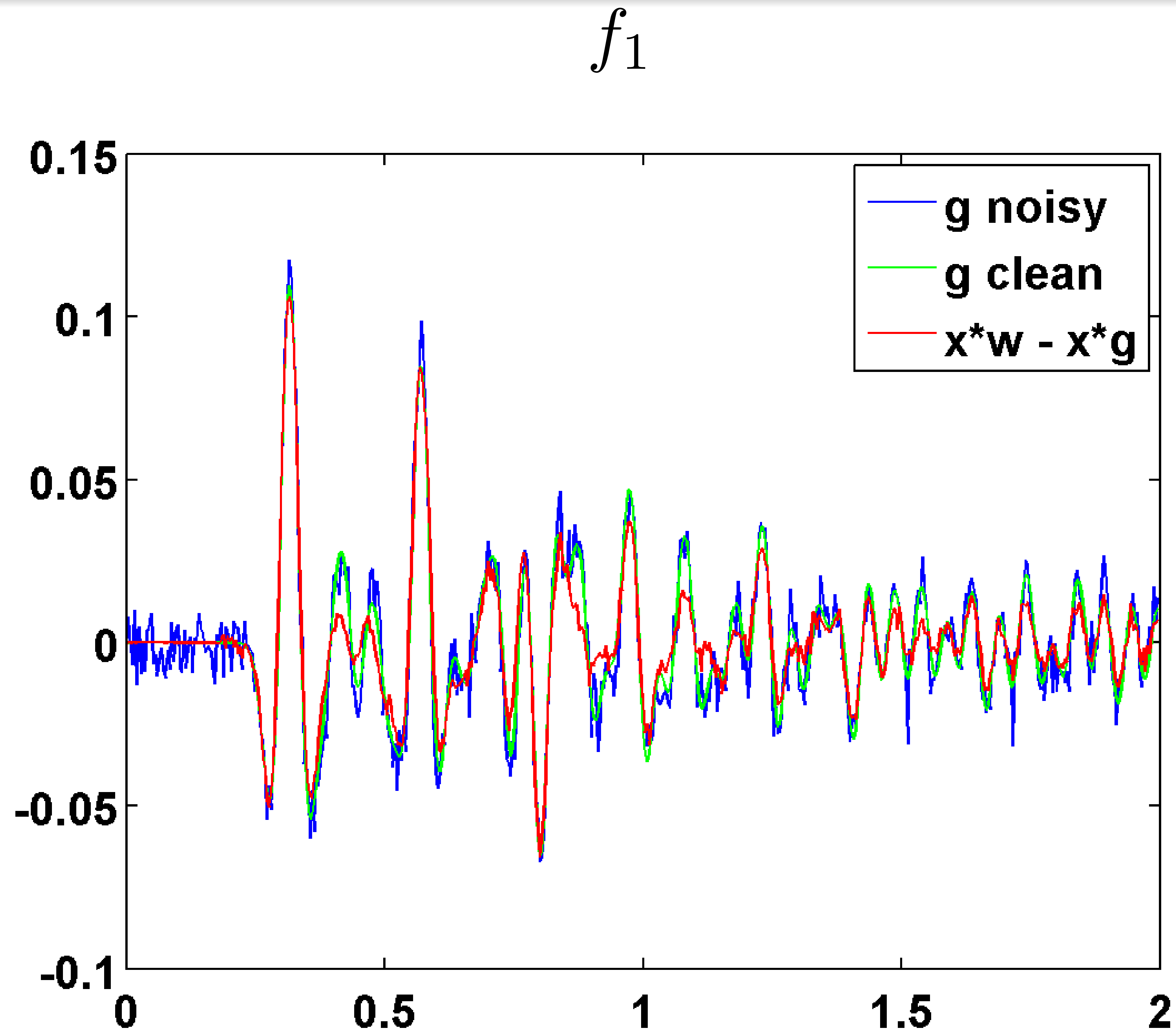
Feedback system: Recovered Wavelet for $n = 50$, $r = 1$, SNR $= 14.2$



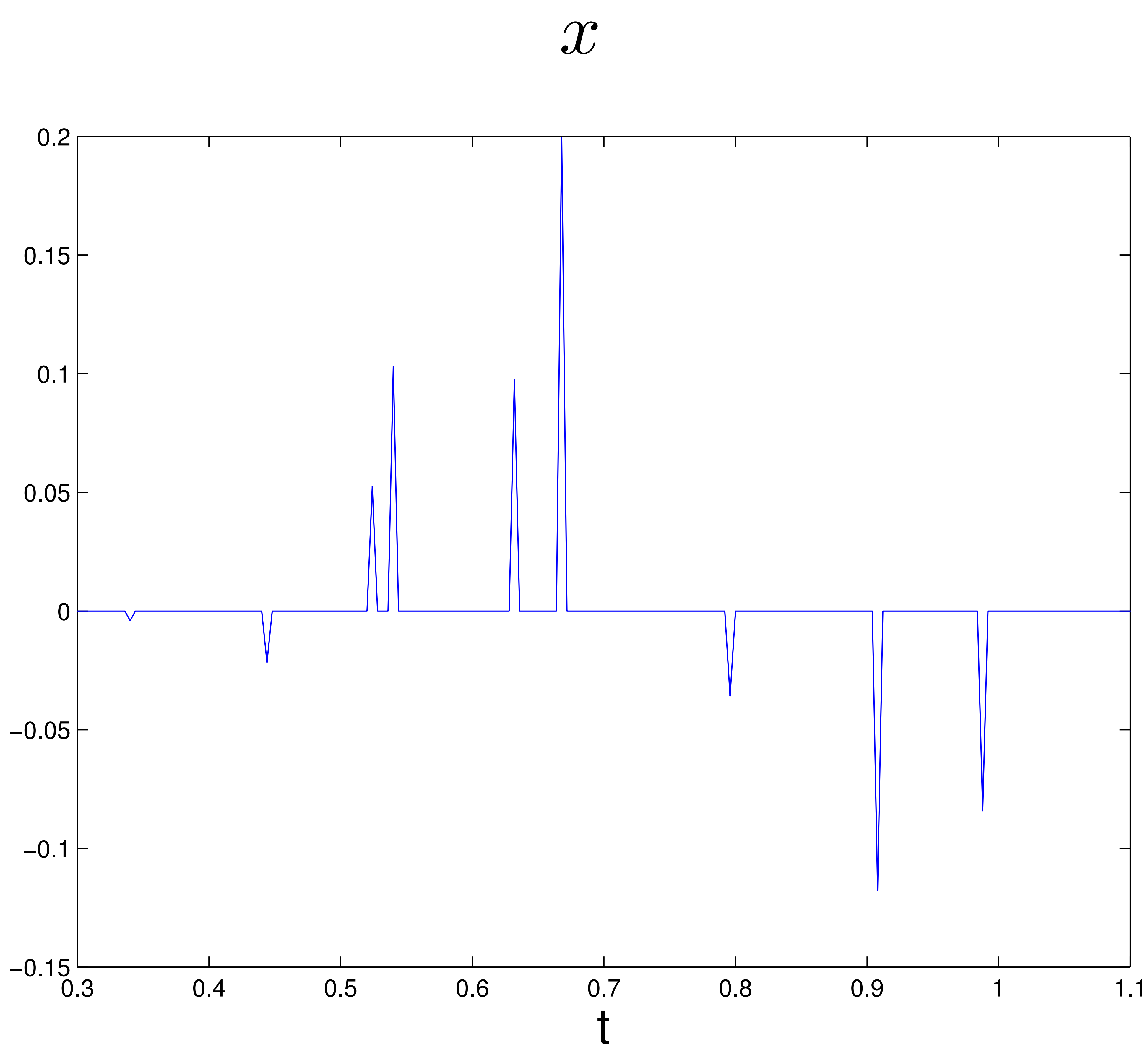


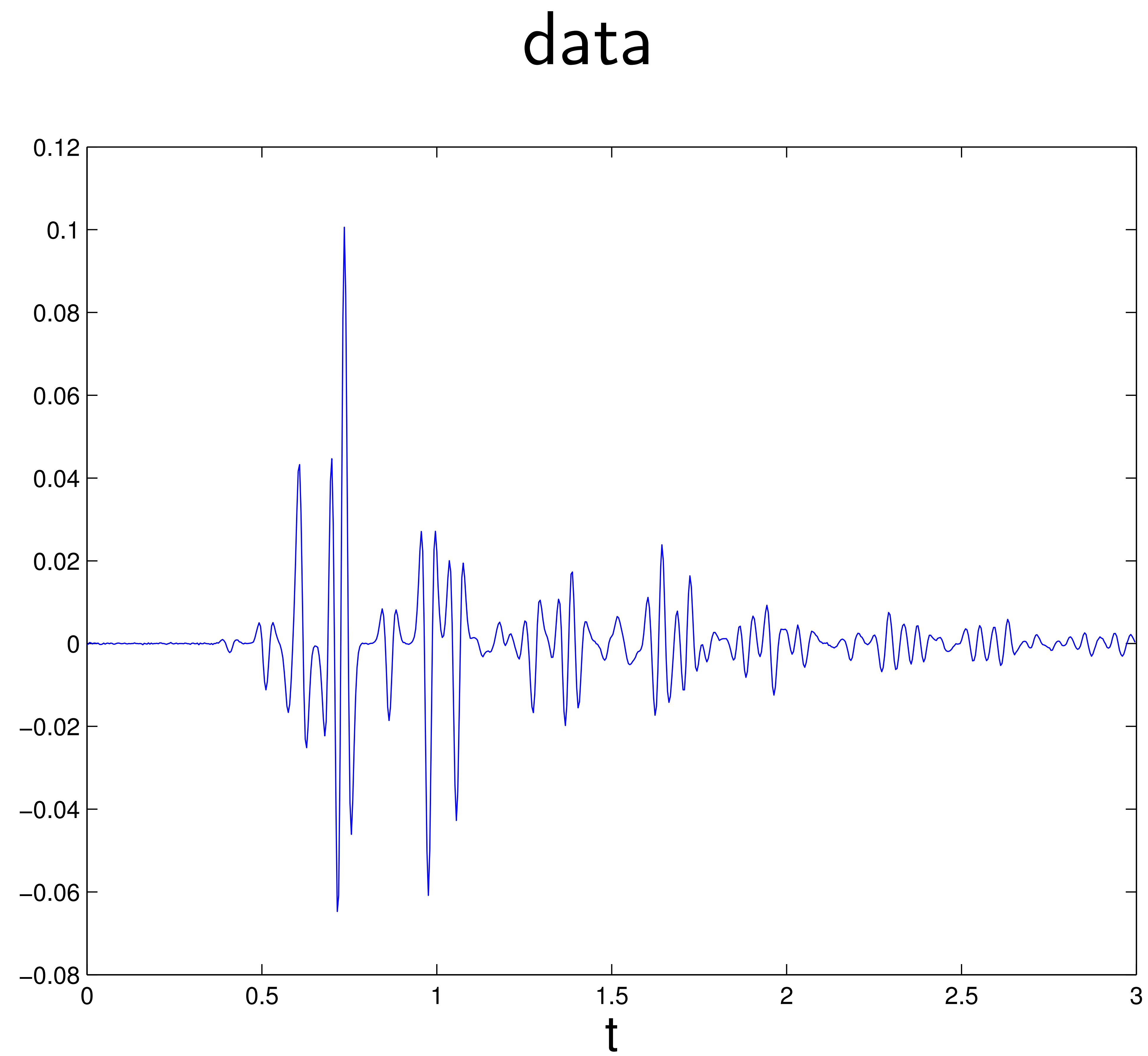
Feedback system: Recovered Sparse Signal for $n = 50$, $r = 1$, SNR $= 14.2$



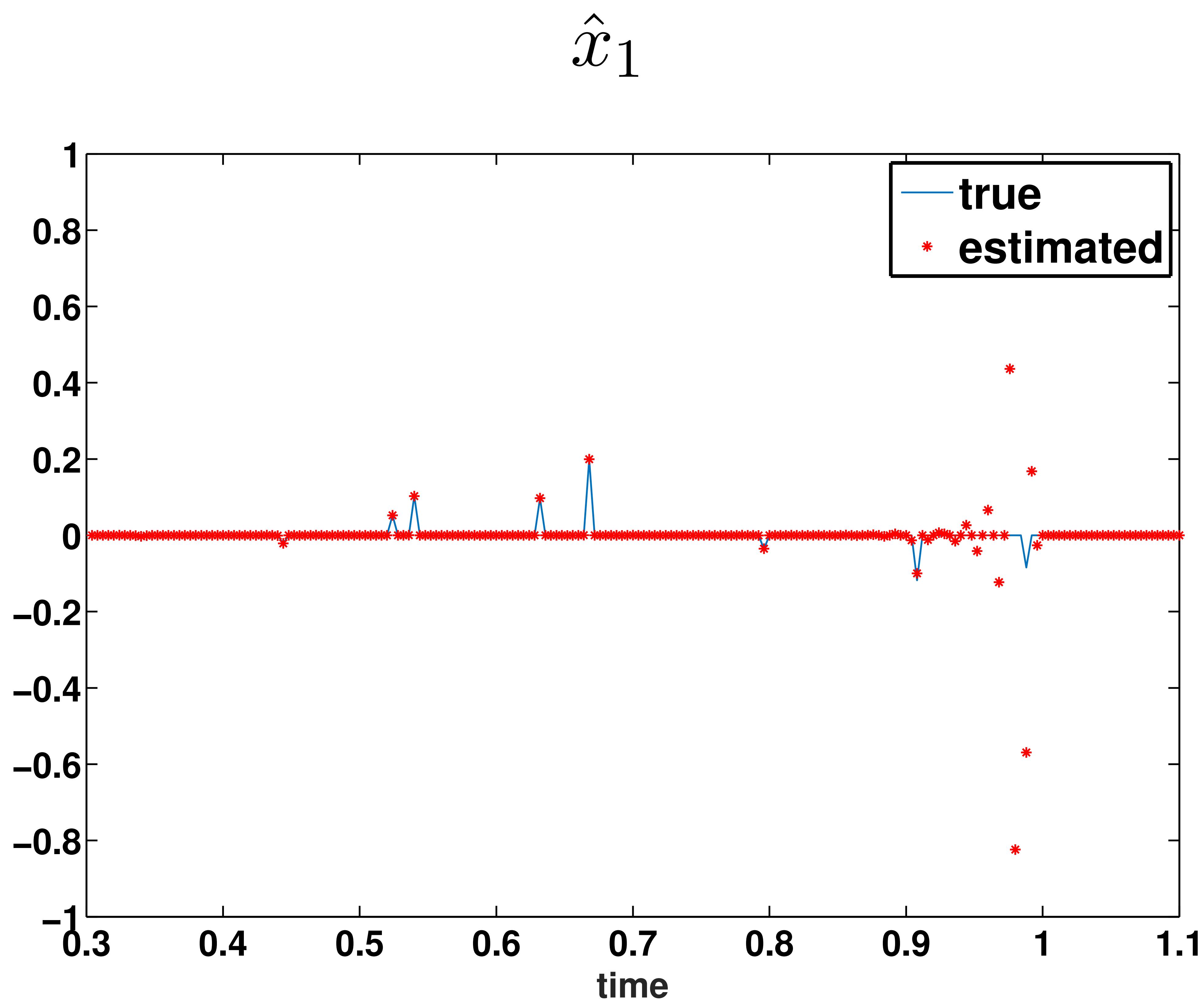


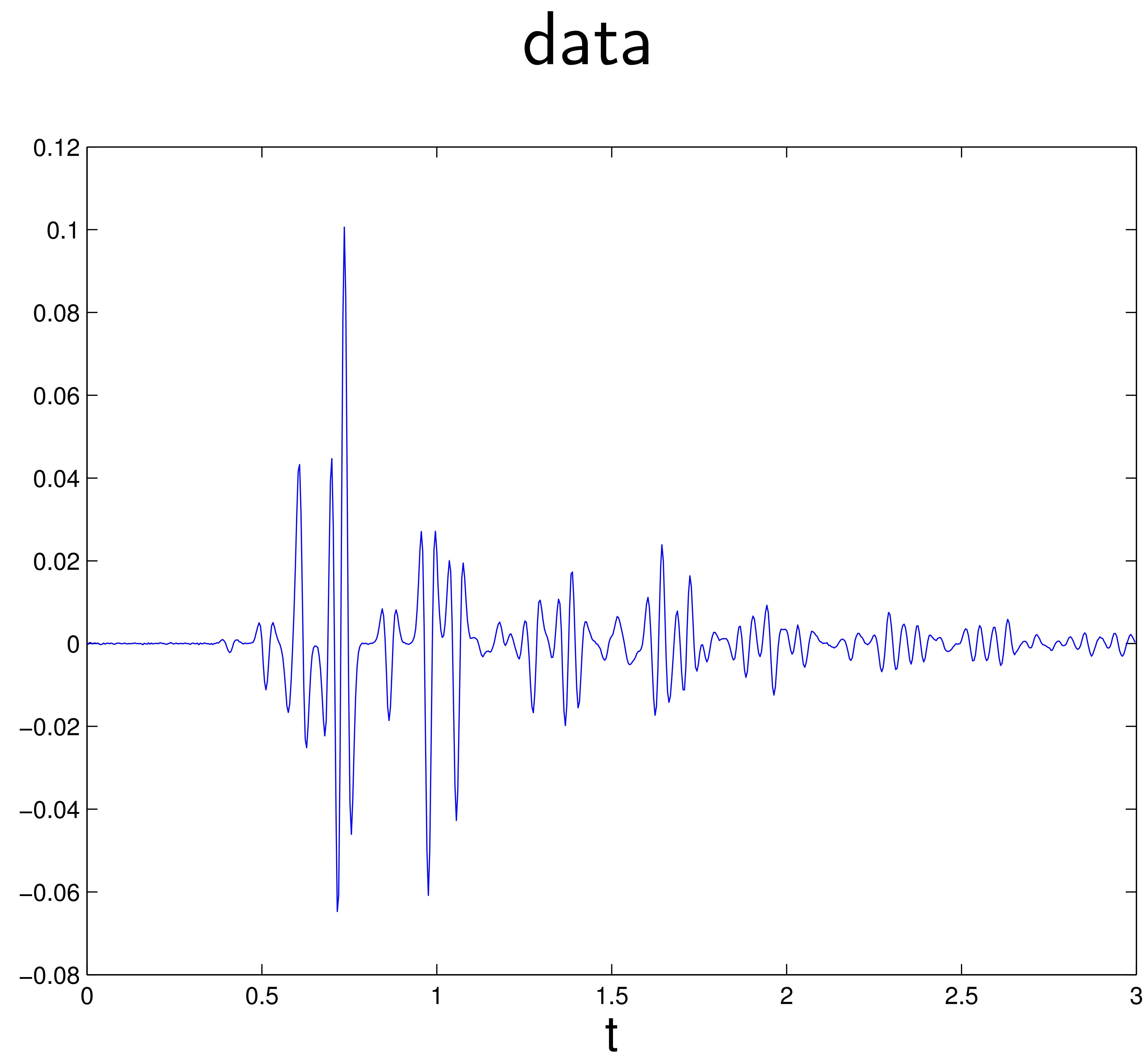
Feedback system: Recovered Sparse Signal for $n = 1$, $r = 1$, SNR $= 50$



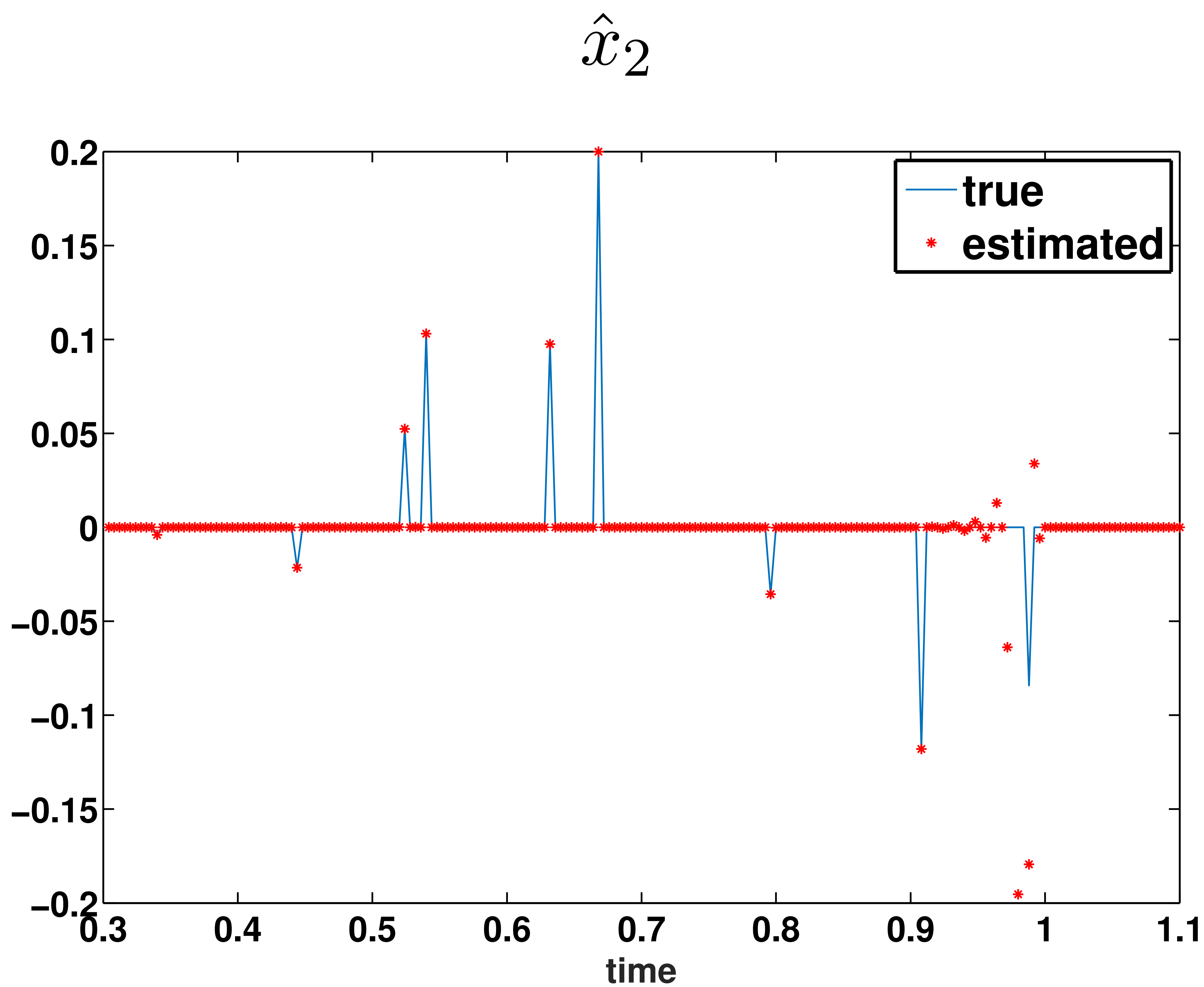


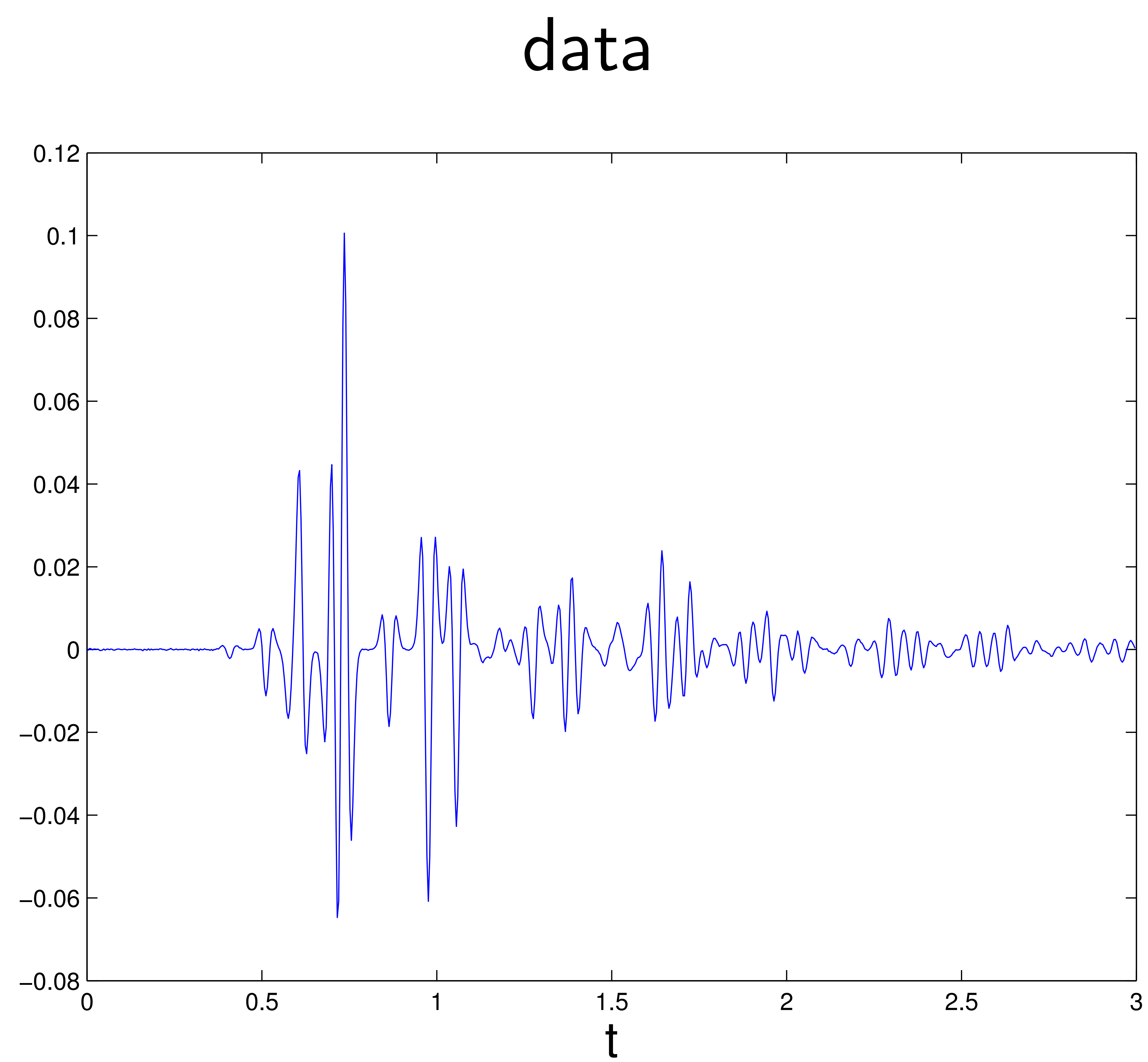
Feedback system: Recovered Sparse Signal for $n = 1$, $r = 1$, SNR $= 50$



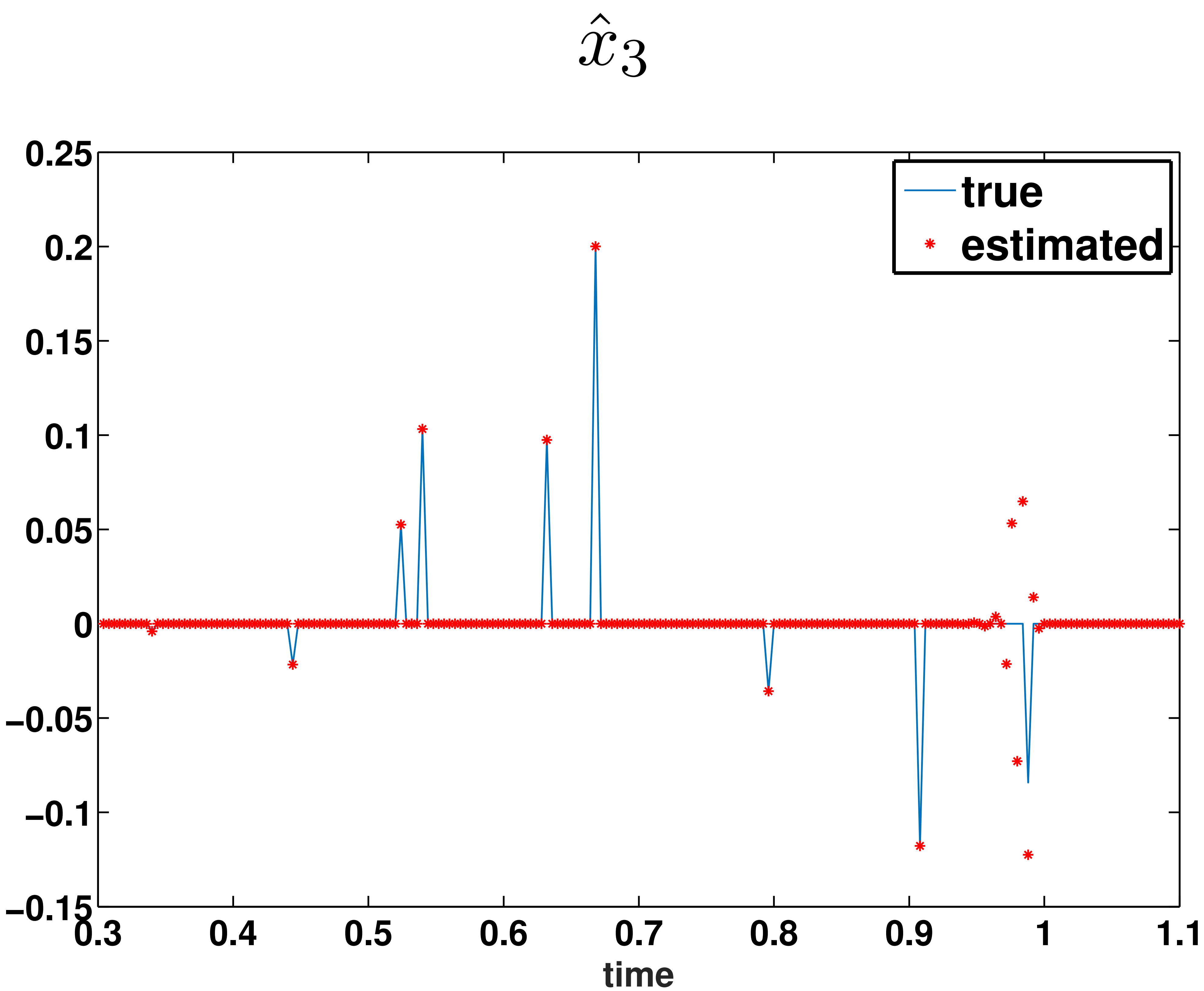


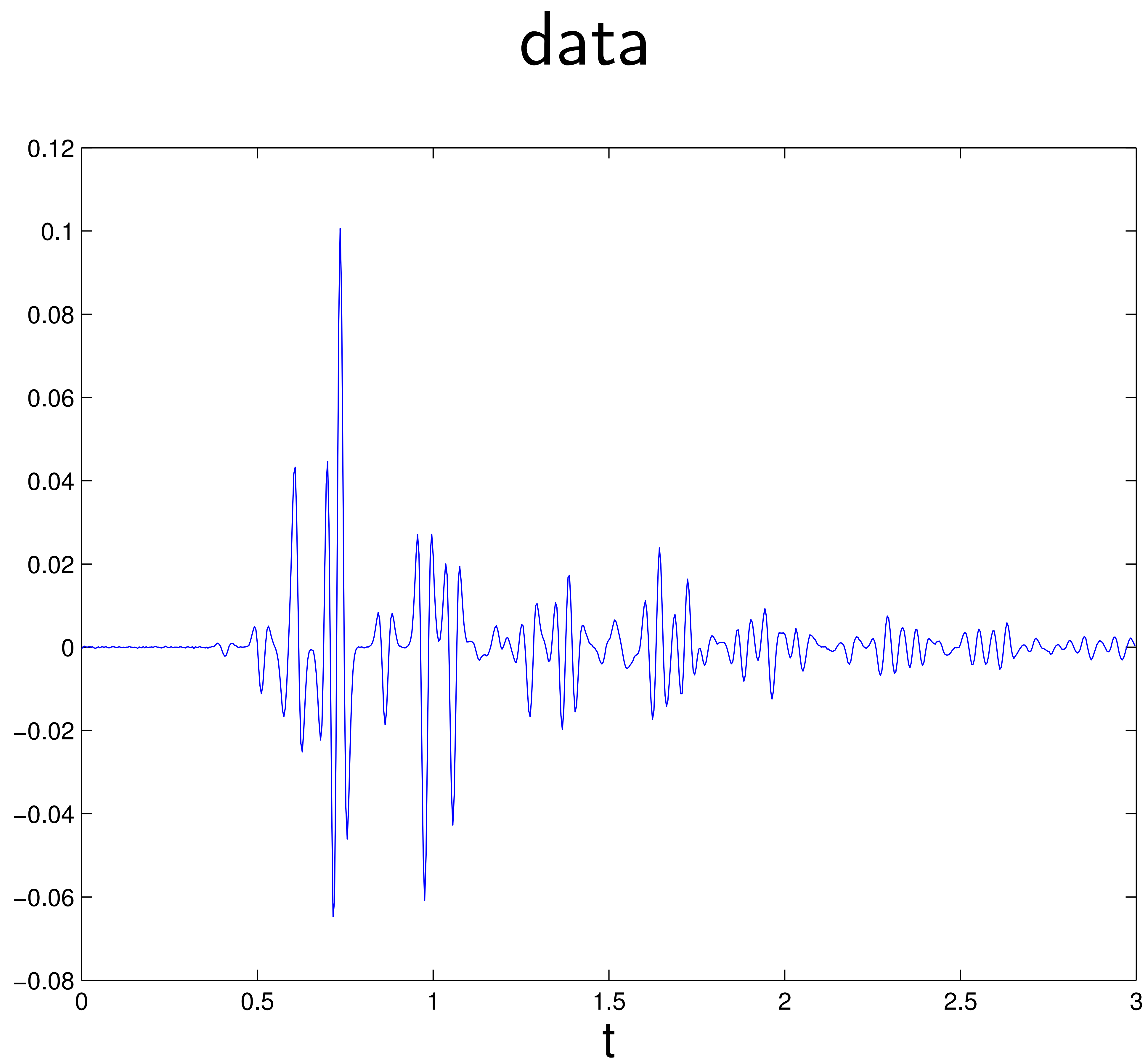
Feedback system: Recovered Sparse Signal for $n = 1$, $r = 2$, SNR $= 50$



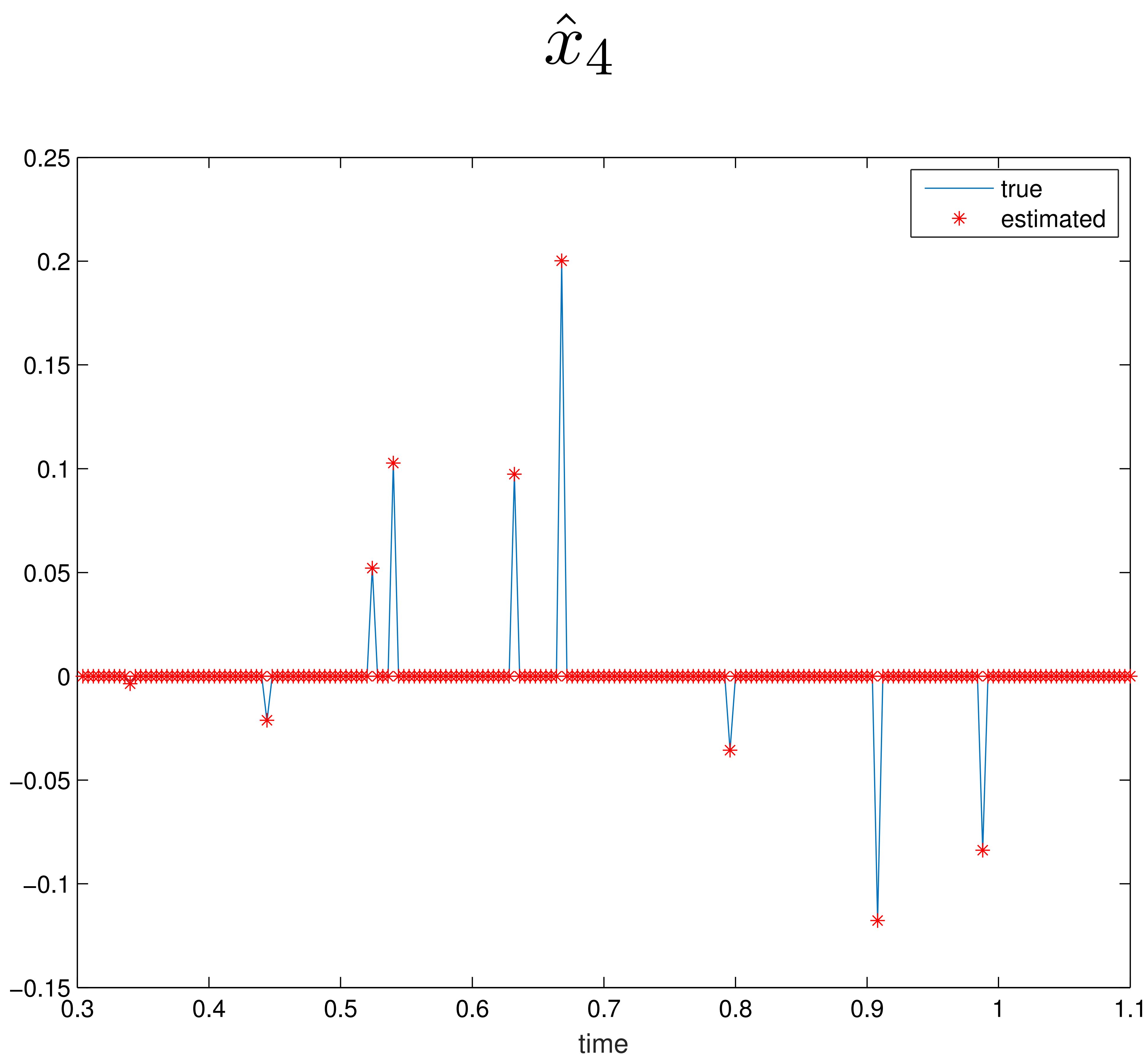


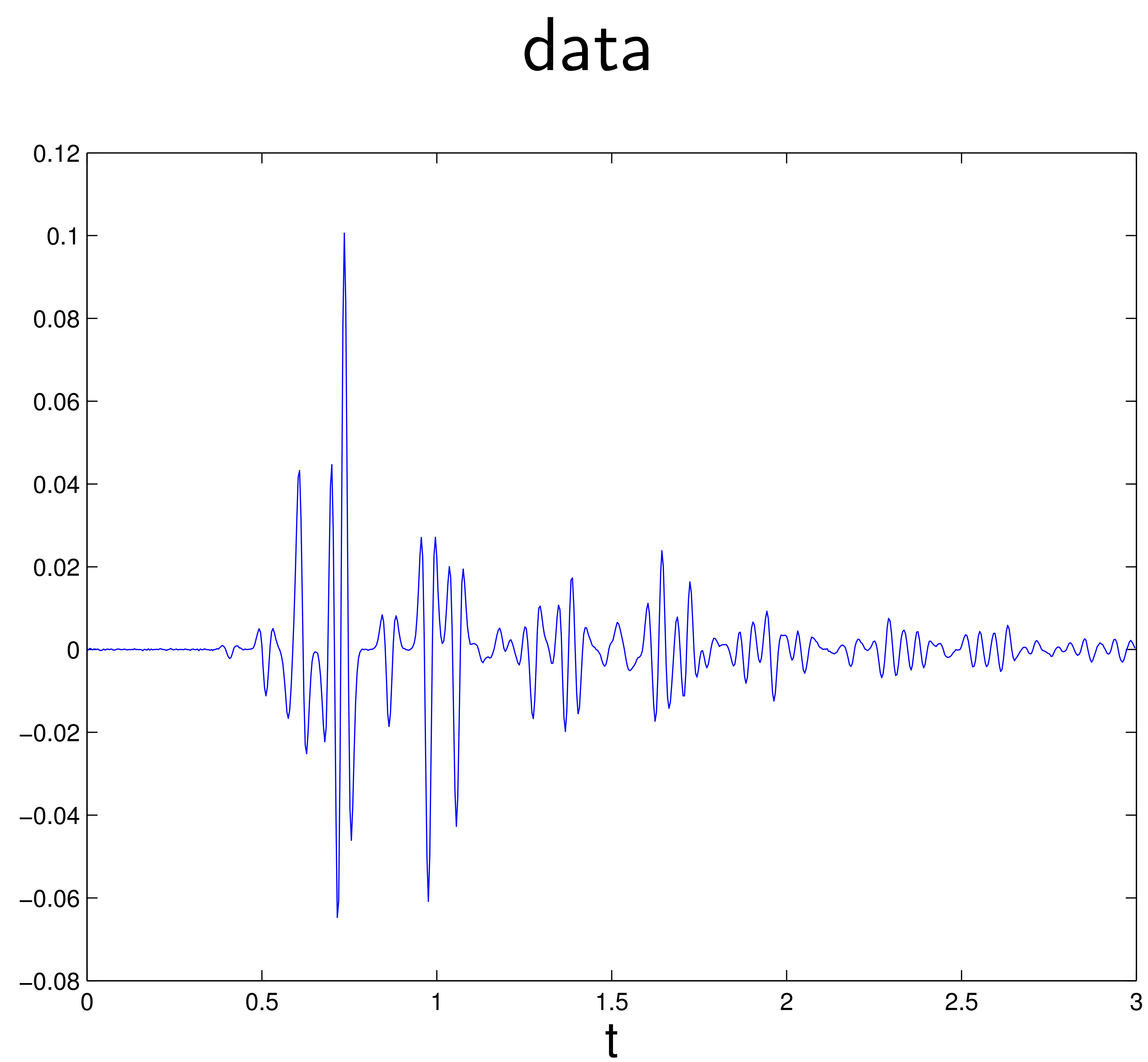
Feedback system: Recovered Sparse Signal for $n = 1$, $r = 3$, SNR $= 50$





Feedback system: Recovered Sparse Signal for $n = 1$, $r = 4$, SNR $= 50$





Related References

[Cosse, Shank, and Demanet, 2015] (used lifting technique for FWI)

[Long, Solna, and Xin, 2013] (used full lifting to solve l_1/l_2).

[Repetti, Pham, and Duval, 2015] (another l_1/l_2 based solver for blind deconvolution)

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

Future Work:

- implementation of the algorithm on 2D model.
- Incorporate into multilevel EPSI algorithm at the coarsest level, where the EPSI deconvolution problems are smaller but more difficult
- Show EPSI model with l_1/l_2 removes shift ambiguity for sparse signals