A General Framework for Robust Compressive Sensing Based Nonlinear Regression

Brian Moore and Balasubramaniam Natarajan
Department of Electrical and Computer Engineering
Kansas State University, Manhattan, Kansas 66506, USA
mooreb@k-state.edu
bala@k-state.edu

Abstract—In this paper, we present a general framework for robust nonlinear regression that leverages concepts from the field of compressive sensing to simultaneously detect outliers and determine optimally sparse representations of noisy data from arbitrary sets of basis functions. Our framework employs a two-component noise model and compressive sensing recovery techniques to exploit the inherent sparsity of outliers while (optionally) performing model order reduction over all predictive variables and basis functions. As such, our algorithm can de-emphasize the effect of predictive variables that become uncorrelated with the measurement data. This desirable property has various applications like real-time detection of faulty sensors and sensor jamming in wireless sensor networks. After developing our framework and making the connection to compressive sensing theory, we present simulations that demonstrate the superior performance of our framework with respect to classic robust regression techniques like least absolute value and iteratively reweighted least-squares.

I. INTRODUCTION

Regression is a fundamental tool in many science and engineering disciplines. As such, least-squares (LS) regression, the most common and simple form of regression, has enjoyed great popularity over the years. However, it is well known that the LS regression solution is not robust to violations of its underlying assumptions, particularly measurement outliers. In practice, measurement outliers could result from faulty, compromised, or jammed sensors. In such cases, it is desirable to elegantly detect and ignore the erroneous information conveyed by the faulty sensors.

Consider the standard regression problem

\[ y_i = f(a, x_i) + \epsilon_i, \quad i = 1, \ldots, M \]  

(1)

where \( \{y_i\}_{i=1}^M \) are the measurements, \( \{x_i\}_{i=1}^M \in \mathbb{R}^L \) are vectors of \( L \) predictive variables corresponding to their respective measurements, \( \{\epsilon_i\}_{i=1}^M \) is measurement noise, \( f \) is the assumed functional relationship on the data, and \( a \in \mathbb{R}^L \) are the regression coefficients. The LS solution to (1) is

\[ \hat{a} = \arg \min_{a} \| y - f(a, x) \|_2 \]  

(2)

where \( y = [y_1, \ldots, y_M]^T \) and \( f(a, x) \in \mathbb{R}^M \) is the vector whose \( i^{th} \) entry is \( f(a, x_i) \). When \( a \) appears linearly in \( f \), we have \( f(a, x) = \Phi a \) for some \( \Phi(x) \); thus (2) has the convenient closed-form solution \( \hat{a} = \Phi^\dagger y \), where \( \Phi^\dagger \) denotes the Moore-Penrose pseudoinverse of \( \Phi(x) \). In any case, \( \{y_i\}_{i=1}^M \) contains outliers, (2) may often produce erroneous and, therefore, undesirable fits.

A. Background

In lieu of LS regression, a number of robust techniques have been proposed. In particular, the method of Least Absolute Values (LAV) [1] replaces the \( \ell_2 \) norm in (2) with an \( \ell_1 \) norm

\[ \hat{a} = \arg \min_{a} \| y - f(a, x) \|_1 \]  

(3)

The \( \ell_1 \) norm effectively de-emphasizes the impact of outliers on the selection of \( \hat{a} \). In practice, the solution to (3) can be efficiently computed using linear programming techniques like simplex methods [2] or interior point methods [3].

Another common approach to robust regression is Iteratively Reweighted Least-Squares (IRWLS) [4], which minimizes the \( \ell_p \) norm, \( \| y - f(a, x) \|_p \), using an iterative procedure where the \((n + 1)^{th}\) estimate of \( a \) is the solution to

\[ \hat{a}^{(n+1)} = \arg \min_{a} \sum_{i=1}^M \omega_i^{(n)} |y_i - f(a, x_i)|^2 \]  

(4)

where the weights, \( \omega_i \), are computed at each iteration via \( \omega_i^{(n)}(a^{(n)}) = |y_i - f(a^{(n)}, x_i)|^{p-2} \) to yield a first order approximation to \( \| y - f(a, x) \|_p \). As the algorithm progresses, the magnitude of the weights vary in such a way to de-emphasize large residuals (i.e., outliers) and promote a robust fit. When \( a \) appears linearly in \( f \), (4) has a closed-form solution [5] at each iteration, and thus it is tractable for many problems of interest.

The method of Least Trimmed Squares (LTS) [6] is another popular approach to robust regression. LTS replaces the traditional LS cost function from (2) with a trimmed version

\[ \hat{a} = \arg \min_{a} \sum_{i \in K} |y_i - f(a, x_i)|^2 \]  

(5)

where \( K \) contains the indices of the \( M - M_0 \) smallest residuals. The LTS approach requires prior knowledge of \( M_0 \), the expected number of measurement outliers. Because we do not assume prior knowledge of the number of measurement outliers in this paper, we will not consider LTS in our simulation study. Another closely related method is Least Median of Squares (LMS) [7], which seeks to minimize only the median residual error. By construction, LMS is robust.
up to 50% outlier contamination. However, exact algorithms require \(O(M^{L+2} \log M)\) complexity [6] and thus are not computationally feasible for many practical problems. Yet other parametric regression methods like the Gaussian mixtures model [8] seek to de-emphasize outliers by imposing various heavy-tailed distributions on the measurement noise. For a thorough survey of robust regression, see [6], [9], [10].

In this paper, we develop a robust compressive sensing based nonlinear regression approach with outlier detection. We then incorporate model order reduction and prove that our approach is well founded by inheriting a recovery theorem from compressive sensing. Finally, we present a simulation study that demonstrates the superiority of our approach to LAV and IRWLS in finding sparse representations of measurement data, a capability which is essential to robust processing of data contaminated by faulty/uncorrelated sensors.

II. COMPRESSIVE SENSING BASED NONLINEAR REGRESSION MODELS

First, consider the nonlinear regression problem

\[
y_i = \sum_{j=1}^{N} a_j \phi_j(x_i) + \epsilon_i, \quad i = 1, \ldots, M
\]

(6)

where the goal is to determine the coefficients \(\{a_i\}_{i=1}^{N}\) which form the optimal linear combination of arbitrary basis functions \(\{\phi_i\}_{i=1}^{N}\) to fit samples \(\{x_i\}_{i=1}^{M}\) given measurement noise \(\{\epsilon_i\}_{i=1}^{M}\). Note that (6) is linear in the regression coefficients, but it is nonlinear in \(x\). We can express (6) in matrix form as

\[
y = \Phi a + \epsilon
\]

(7)

where \(y \in \mathbb{R}^M\), \(a \in \mathbb{R}^N\), \(\epsilon \in \mathbb{R}^M\), and \(\Phi \in \mathbb{R}^{M \times N}\) is of the form \(\Phi_{i,j} = \phi_j(x_i)\). Furthermore, suppose that we now have \(L\) predictive variables, \(x^1, \ldots, x^L\). In this case, we can extend (7) by writing

\[
y = \Phi_L a_L + \epsilon
\]

(8)

where we now have \(\Phi_L = [1, \Phi_1, \ldots, \Phi_L] \in \mathbb{R}^{M \times (NL+1)}\), \(a_L = [a_0, a_1^T, \ldots, a_L^T]^T \in \mathbb{R}^{NL+1}\), and where \(\Phi_i\) and \(a_i\) are defined by (7) for the corresponding predictive variable \(x^i\), \(i = 1, \ldots, L\). Note that we have prepended an \(M\)-vector of \(1\)’s to \(\Phi_L\) and a coefficient \(a_0\) to \(a_L\) to avoid ambiguity in distributing constant measurement offset between the predictive variables.

A. Two Component Noise Model

Given the general model in (8), we next seek to identify measurement outliers by imposing a two component model on the noise \(\epsilon\). Indeed, consider the decomposition [11]

\[
\epsilon = \tilde{\epsilon} + \omega
\]

(9)

where our interpretation of \(\tilde{\epsilon}\) and \(\omega\) are as follows: when \(y_i\) contains an outlier, \(\omega_i\) is such that \(y_i = \omega_i\) appears to be contaminated by normal measurement noise \(\tilde{\epsilon}_i\); when \(y_i\) does not contain an outlier, \(\omega_i \equiv 0\) and \(y_i\) appears to be contaminated by normal measurement noise \(\tilde{\epsilon}_i\). More generally, we propose an outlier noise model with respect to an arbitrary basis \(\Psi \in \mathbb{R}^{M \times M}\). Under this general framework, we can express the measurement noise as

\[
\epsilon = \tilde{\epsilon} + \Psi \omega
\]

(10)

Equation (10) has many practical applications. For example, when \(\Psi = W_M\), the \(M \times M\) DFT matrix, the model in (10) can be used to remove outlier frequencies (e.g., \(60Hz\) noise) from the measurement data. Alternatively, if \(\Psi\) is a wavelet basis, (10) can remove outliers that are localized in both space and frequency. In any case, (10) exploits the inherent sparsity of outliers w.r.t. \(\Psi\) to isolate their noise contribution in \(\omega\).

B. Robust Outlier Detection

Combining our two component noise model with (8), we propose solving the following convex optimization problem to perform robust nonlinear regression with outlier detection

\[
\hat{a}_L, \hat{\omega} = \arg\min_{a_L, \omega} ||\omega||_1 \text{ s.t. } ||y - \Phi_L a_L - \Psi \omega||_2^2 \leq \zeta
\]

(11)

Equation (11) is known in optimization literature [12] as a basis pursuit denoising (BPDN). It can be cast as a second order cone program (SOCP) [13], a special case of quadratic programming, and solved efficiently using interior point methods [3]. When knowledge of the measurement noise is available, \(\zeta = E[||\epsilon||_2^2]\) is an appropriate choice [14]. Note that, in general, the \(\ell_p\) norm \((0 \leq p \leq 1)\) makes sense in the cost function of (11); however, in the sequel, we restrict ourselves to \(p = 1\) to preserve convexity. Equivalently, (11) can be recast as

\[
\hat{a}_L, \hat{\omega} = \arg\min_{a_L, \omega} \frac{1}{2}||y - \Phi_L a_L - \Psi \omega||_2^2 + \lambda ||\omega||_1
\]

(12)

where convex optimization theory [13] guarantees that (11) and (12) yield the same solution for some \(\lambda = f(\zeta)\). In optimization literature, (12) is known as the LASSO [15] and is well known for its applications in sparsity-based optimization.

Equation (11) allocates space in \(\omega\) for outlier measurements while simultaneously searching for the optimal regression coefficients for \(y-\omega\), the resulting outlier-free measurements. Moreover, the \(\ell_1\) cost function promotes sparsity in \(\omega\), which agrees with our intuition that outliers are inherently sparse. Equation (11) can be considered an extension of the LAV method, which is obtained by letting \(\zeta \to 0\). Our formulation has the desirable flexibility to distinguish outliers from standard measurement noise that arises in \(\tilde{\epsilon}\).

C. Incorporating Model Order Reduction

Having successfully developed a robust convex outlier detection method, we will next slightly alter our algorithm to elegantly incorporate model order reduction capability. Indeed, consider the modified version of (11)

\[
\hat{a}_L, \hat{\omega} = \arg\min_{a_L, \omega} \left[ \begin{array}{l} \omega \\ a_L \end{array} \right] \text{ s.t. } ||y - \Phi_L a_L - \Psi \omega||_2^2 \leq \zeta
\]

(13)

Note that the cost function of (13) now contains \(\omega\) and \(a_L\). This subtle change not only retains outlier sparsity promotion
in $\omega$, but it also promotes regression coefficient sparsity in $a_L$. From the definition of $a_L$, we see that (13) now effectively seeks regression coefficients which yield a close fit using a minimal number of basis functions $\phi(x)$ across all predictive variables. Thus (13) can be viewed as imposing model order reduction on the robust regression. Our simulation study verifies that (13) achieves superior results for regression problems with sparse $a_L$.

D. Connection to Compressive Sensing

We will now develop the link between our proposed regression algorithms and standard compressive sensing (CS) problems in order to inherit their corresponding recovery guarantees. CS [16], [17], [18] is a sparse signal recovery technique that addresses problems of form (7) for $\Phi$ with certain properties. One of the main results of CS theory is a condition under which algorithms of the form

$$\hat{a} = \arg \min_{a \in \mathbb{R}^N} \|a\|_1 \text{ s.t. } \|y - \Phi a\|_2^2 \leq \zeta$$

(14)
can reliably recover arbitrary $K$-sparse vectors $a \in \Sigma_K$, where $\Sigma_K = \{a \in \mathbb{R}^N : \|a\|_0 \leq K\}$ is the set of all $N$-length vectors with no more than $K$ nonzero entries. Specifically, we have the following theorem [19], which we state without proof.

**Theorem 1:** (Recovery Guarantee) Let $a_K$ be the result of hard thresholding the $K$ largest entries of $a$. Then if

$$K < \frac{1 + \mu^{-1}(\Phi)}{4}$$

(15)
and $\zeta \triangleq E[\|e\|_2^2]$, the solution, $\hat{a}$, to (14) obeys

$$\|\hat{a} - a_K\|_2 \leq 2\zeta \sqrt{\frac{1}{a - \mu(\Phi)(4K - 1)}}$$

(16)

Theorem 1 bounds the recovery error of (14) based on sparsity, noise level, and coherence of $\Phi$, which is defined as

$$\mu(\Phi) \triangleq \sqrt{N} \max_{i \neq j} |\langle \phi_i, \phi_j \rangle|$$

(17)
The Welch bound [20] reveals that $\mu(\Phi) \in \left[\sqrt{\frac{M-N}{N(M-N)}}, 1\right]$ for arbitrary $\Phi$. Theorem 1 suggests that small $\mu(\Phi)$ relaxes the sparsity constraint and allows (14) to accurately reconstruct $a$ with more nonzero entries.

The relevance of Theorem 1 to our proposed model order reduction method lies in the similarity between (13) and (14). Indeed, consider the following equivalent form of (13)

$$\hat{a}_L, \hat{\omega} = \arg \min_{a_L, \omega} \|a_L\|_1 \text{ s.t. } \|y - \Phi \hat{a}\|_2^2 \leq \zeta$$

(18)
where $\hat{a} \triangleq [\hat{a}_L, \hat{\omega}]$ and $\hat{\Phi} \triangleq [\Psi, \Phi_L]$. Comparison of (18) with (14) reveals that we can directly apply the recovery guarantee of Theorem 1 to our model order reduction formulation. Specifically, $\mu(\hat{\Phi})$ determines the sparsity level (i.e., number of outliers and basis functions) that our algorithm can reliably detect. In any case, Theorem 1 and (18) provide a foundation that guarantees our robust approaches in (11) and (13) are well founded.

III. SIMULATION STUDY

To study the practical results of our proposed robust outlier detection and model order reduction regression techniques, we considered the regression problem

$$y_i = a_0 + \sum_{j=1}^{12} a_j \phi_j(x_i) + \epsilon_i + \omega_i \quad i = 1, \ldots, M$$

(19)
where we fix $M = 500$. Note that (19) is a nonlinear regression problem with 1 predictive variable and 12 basis functions. To obtain arbitrary noisy data sets, we first generated 50 unique 1-sparse coefficient vectors $a_r = [a_{r0}, \ldots, a_{r12}], r = 1, \ldots, 50$. For each $r$, we randomly selected an integer $K \in [0, 12]$ and set $a_k \sim \pm U(2, 5)$ if $k = K$ and $a_k \equiv 0$ if $k \neq K$. We then let $\epsilon_i \sim N(0, \sigma^2) \forall i$, where $\sigma = 0.1$. To contaminate our data with $\delta \times 100\%$ outliers, we then selected $\omega_i$ according to $\omega_i \sim \pm U(10, 30)$ with probability $\delta$ and $\omega_i \equiv 0$ with probability $1 - \delta \ \forall i$. Finally, we generated 500 predictive variable samples according to $x_i \sim U(0, 8)$.

Using the above data, we then computed our measurement vector $y = [y_1, \ldots, y_{500}]$ from (19) after computing $\Phi$ from (8) using the basis functions

$$\phi_i(x) = \cos \left(\frac{15\pi x}{8}\right)$$

and $\phi_j(x) = \sin \left(\frac{15\pi x}{8}\right)$

$$\phi_4(x) = \cos \left(\frac{\pi x}{2}\right)$$

$$\phi_5(x) = \sin \left(\frac{\pi x}{2}\right)$$

$$\phi_6(x) = 4(x-4)^{-1}$$

$$\phi_7(x) = \frac{1}{3}(x-4)$$

$$\phi_8(x) = e^{-\frac{(x-9)^2}{2}}$$

With $y$ and $\Phi$ in hand, we then applied our robust regression methods from (11) and (13) with $\Phi_L = \Phi$, $\Psi = I_{500}$, and $\zeta = E[\|e\|_2^2] = \sigma \sqrt{M} = \sqrt{5}$ to all 50 data sets. In the sequel, we will refer to (11) as the compressive sensing method (CS), and we will refer to (13) as the compressive sensing method with model order reduction (CS w/ MOR). To evaluate the performance of our methods, we performed each regression two additional times using the LAV and IRWLS algorithms.

Figs. 1 and 2 show the results of our simulation for $\delta = 0.05$ and $\delta = 0.3$, respectively. The left plot displays the regression coefficient error, defined as $\|a_L - a_r\|_2^2$; the middle plot displays the mean residual, defined as the square of the mean value of $\Phi(a_L - a_r)$; the right plot displays the number of significant regression coefficients, defined as the number of entries of $a_L$ with magnitude greater than $\sigma = 0.1$. Although all four algorithms achieved comparable mean residuals, the coefficient error and significant coefficient plots show that CS w/ MOR was the only algorithm that successfully detected the single significant coefficient. All other methods utilized significant components of 7-10 basis functions but achieved
no further residual reduction. Similarly, by recognizing the inherent sparsity of $\alpha_r$, CS w/ MOR was able to achieve four orders of magnitude decrease in coefficient error with respect to the other algorithms. Comparison of Figs. 1 and 2 also shows that all four algorithms are robust to outliers in the sense that their performance was largely unaffected by the 25% increase in outlier contamination. As discussed earlier, LAV is a special case of our CS approach. Thus it is not surprising to see the striking similarity between the mean residual characteristics of each.

IV. CONCLUSION

In this paper, we presented a general framework for compressive sensing based nonlinear regression. First, we designed a convex optimization problem that can simultaneously detect measurement outliers and compute the optimal regression coefficients for the outlier-less data. Next, we extended our approach to incorporate model order reduction and leveraged compressive sensing theory to provide explicit performance guarantees based on the structure of the candidate problem. Finally, we conducted a simulation study that experimentally confirmed the performance of our approach and demonstrated its superior model order reduction capabilities.

REFERENCES