A Robust Nonparametric Estimation Framework for Implicit Image Models

Himanshu Arora  
University of Illinois  
Urbana, IL61801, USA  
haroral@uiuc.edu

Maneesh Singh  
Siemens Corporate Research  
Princeton, NJ08540, USA  
msingh@scr.siemens.com

Narendra Ahuja  
University of Illinois  
Urbana, IL61801, USA  
n-ahuja@uiuc.edu

Abstract

Robust model fitting is important for computer vision tasks due to the occurrence of multiple model instances, and, unknown nature of noise. The linear errors-in-variables (EIV) model is frequently used in computer vision for model fitting tasks. This paper presents a novel formalism to solve the problem of robust model fitting using the linear EIV framework. We use Parzen windows to estimate the noise density and use a maximum likelihood approach for robust estimation of model parameters. Robustness of the algorithm results from the fact that density estimation helps us admit an a priori unknown multimodal density function and parameter estimation reduces to estimation of the density modes. We also propose a provably convergent iterative algorithm for this task. The algorithm increases the likelihood function at each iteration by solving a generalized eigenproblem. The performance of the proposed algorithm is empirically compared with Least Trimmed Squares (LTS) — a state-of-the-art robust estimation technique, and Total Least Squares (TLS) — the optimal estimator for additive white Gaussian noise. Results for model fitting on real range data are also provided.

1. Introduction

Robust model fitting is central to many computer vision tasks. Examples include tracking or registration under Euclidean, affine, or projective transformations; surface normal and curvature estimation for 3D structure detection; and, fitting intensity models for object recognition and object registration. Robust estimation implies a framework which tolerates the presence of outliers — samples not obeying the relevant model. Consider the problem of segmenting a range image with planar patches: Here, each plane satisfies a linear parametric model. For estimating parameters of each plane, samples from all other planes should be considered as outliers, i.e. they should not contribute to the error in fit. Another scenario is when the noise model for the observed samples is not known. It is not possible to come up with a cost function which is optimal for every kind of (unknown) noise model. Robust estimation seeks to provide reliable estimates in such cases — when data is contaminated with outliers in form of samples corrupted by unknown noise or when multiple structures are present in the data, some or all of which need to be detected.

Much work has been done in robust estimation in statistics, and more recently in vision. We refer the reader to [13] for a recent review of robust techniques used in computer vision. The two major classes of robust methods proposed in statistics, M-estimators, and least median of squares (LMedS), are regularly used by computer vision researchers to develop applications. M-estimators, a generalization of maximum likelihood estimators and least squares method, were first defined by Huber [6] and their asymptotic properties were studied by Yohai et al. [14], and Koenker et al. [8] in separate works. Least median of squares (LMedS) was proposed by Rousseeuw [10], wherein the sum of squared residuals in traditional least squares is replaced by median of squared residuals. The Hough Transform [7], [9], and RANSAC [4] were independently developed in computer vision community for robust estimation. For Hough Transform, entire parameter space is discretized and optimal parameters are estimated by a voting scheme due to each data sample. It can be viewed as a discrete version of M-estimation. RANSAC [4] uses the number of points with residual below a threshold as the objective function. It has similarities with both M-estimators, and LMedS. Recently, Chen et al. [2] showed that all robust techniques applied to computer vision, i.e. those imported from statistics, and those developed in computer vision literature, can be described as specific instances of the general class of M-estimators with auxiliary scale. In a separate work [1], Chen et al. explore the relationship between M-estimators and kernel density estimators, and propose a technique for robust estimation based on kernel density estimators.

Many parameter estimation problems in computer vision can be formulated with the linear errors-in-variable model (EIV) [15], where the observations are assumed to be corrupted by additive noise. Further, it is often desirable to use implicit functional form. For instance, consider the problem of range image segmentation mentioned earlier. We can extract the 3D world coordinates, \((x_i, y_i, z_i)\) from the range data. If the range measurements, \(r_i\) are noisy, \((x_i, y_i, z_i)\)
Substituting the values of \( \hat{\theta}, \hat{\alpha} \) for \( \theta, \alpha \), the equations become:

\[
\begin{align*}
\sum_{i=1}^{n} (x_i^T \hat{\theta} - \hat{\alpha}) x_i + \lambda \hat{\theta} &= 0 \\
\sum_{i=1}^{n} x_i^T (\hat{x}_i - \hat{\alpha}) &= 0
\end{align*}
\]

Similarly, the solution, \( \hat{\theta}, \hat{\alpha} \), to the problem specified by (2) subject to the constraints in Definition 2.1, should satisfy the following equations:

\[
\begin{align*}
x_{io} : & \quad x_i - \hat{x}_{io} = \lambda_i \hat{\theta} & i = 1, ..., n \\
\theta : & \quad \sum_{i=1}^{n} x_i^T \hat{\theta} + \gamma \hat{\theta} = 0 \\
\alpha : & \quad \sum_{i=1}^{n} \lambda_i = 0
\end{align*}
\]

These two equations can be used in (8) to get,

\[
\sum_{i=1}^{n} (x_i^T \hat{\theta} - \hat{\alpha}) x_i + (\gamma - \sum_{i=1}^{n} \lambda_i^2) \hat{\theta} = 0
\]

Substituting the values of \( \lambda_i \) in (9) also gives,

\[
\sum_{i=1}^{n} (x_i^T \hat{\theta} - \hat{\alpha}) = 0
\]
Further, (5) and (6) can be used to show that \( \lambda = \sum_{i=1}^{n}(x_i^T \hat{\theta} - \hat{\alpha})^2 \). Similarly, (7)-(9) imply \( \gamma = \sum_{i=1}^{n} \lambda_i^2 = \sum_{i=1}^{n}(x_i^T \hat{\theta} - \hat{\alpha})^2 \). Thus, solutions to (5)-(6) and (10)-(11) are the same. Hence both problems are equivalent. \( \square \)

Now, let us examine the problem in (4) more closely. Defining \( \beta = (\theta, \alpha)^T \), \( A = \sum_{i=1}^{n} \left( x_i x_i^T - x_i^T \right) \) and \( B = \left( I_p \ 0 \ 
\ 
\ 0 \ 0 \right) \), where \( I_p \) is the \( p \times p \) identity matrix, we can rewrite (4) as

\[
\hat{\beta} = \arg\min_{\beta} \beta^T A \beta; \quad \beta^T B \beta = 1
\]

Solving for \( \beta \) leads to \( A \beta = \lambda B \beta \), where \( \lambda \) is minimum eigenvalue for the generalized eigenproblem. Thus, the solution is the generalized (minimum) eigenvector of \( A \) with respect to \( B \). Consequently, The Maximum likelihood estimation of the linear EIV model parameters in case of Gaussian noise reduces to a generalized eigenproblem.

The assumption, made above, of Gaussian noise is not always desirable. The model of noise is often unknown, and the estimator proposed above may not be optimal in general. In particular, for heavy tailed distributions (e.g. log-normal distribution as will be discussed in section 4), the approach above might have a really bad performance. Also, the structure that we need to detect (in this case the model that we need to fit) might only be valid locally. For instance, while detecting multiple planer segments in a range image, the model parameters are valid only on (local) segments of data. Since the segmentation is not a priori available, robust estimation becomes important for the discovery of any local models. It tolerates the presence of data samples that do not obey the model that is to be estimated. In the next Section, we propose a principled approach to carry out robust estimation.

### 3. Robust EIV Estimation Using Parzen Windows

If we know the noise model for the linear EIV problem, then we can use the maximum likelihood approach to estimate the EIV model parameters. However, quite often, we do not have access to such a model. In that case, one can take recourse to estimating the noise density and then applying the maximum likelihood framework. In this section, we present such an approach. We first formulate the problem in terms of a noise density estimate using Parzen windows and subsequently, we propose a solution to the said problem.

Parzen windows or kernel density estimators are a popular non parametric density estimation technique in pattern recognition and computer vision [3]. The Parzen window estimate of the pdf from a given set of data samples can be defined as follows:

*Definition 3.1 (Kernel Density Estimator). Let the observed samples \( y_i \in \mathbb{R}^p, i = 1, ..., n \) be generated independently from an underlying probability distribution function \( f(x), f : \mathbb{R}^p \rightarrow \mathbb{R}^+ \). Then the kernel density estimate for \( f \) is defined as,\n
\[
\hat{f}(y) = \frac{1}{n \det(H)} \sum_{i=1}^{n} K(H^{-1}(y - y_i))
\]

where, \( H \) is a nonsingular bandwidth matrix, and \( K : \mathbb{R}^p \rightarrow \mathbb{R}^+ \) is the kernel function with zero mean, unit area, and identity covariance matrix.

The kernel function \( K(\cdot) \) used above is often assumed to be rotationally symmetric. We find it convenient to define the profile of this rotationally symmetric kernel as a univariate kernel function \( \kappa : \mathbb{R} \rightarrow \mathbb{R}^+ \), where \( K(y) = c_k \kappa(-||y||^2) \), \( c_k \) being a normalization constant.

The kernel density estimate of an arbitrary set of data samples can be computed as shown above. However, the above density estimate does not factor in any prior knowledge that one may have of the data. For example, the data might be generated using a parametric model. For such a case, we proposed a zero bias-in-mean kernel estimator in our earlier work [12]. We used this estimator for robust (parameter) estimation where the image is specified using an explicit parametric formulation. In this paper, we adapt the aforementioned approach to define a robust kernel maximum likelihood estimation framework for the EIV model. We draw the reader’s attention to the fact that the EIV model is an implicit function formulation unlike our previous work.

Now we explain our approach in terms of noise density estimation: Let us assume that the noise free values \( x_{io} \) and the parameters \( [\theta, \alpha] \) are known such that the constraints in Definition 2.1 are satisfied. Then, the noise can be estimated as \( \epsilon_i = x_i - x_{io}, \) with \( x_{io}^T \theta = \alpha, i = 1, ..., n \), and \( ||\theta|| = 1 \). The noise is nothing but the deviation of the observation from the model described by the parameters. The key question is to decide the metric that is to be chosen on these deviations to estimate the model parameters. If the noise density was known, one could easily formulate such a metric using the maximum likelihood framework. However, since the noise density is not known, we take the next best approach — we use Parzen windows to estimate the noise density using Definition 3.1.

For a set of observed data samples \( \{x_{io}\}_{i=1}^{n} \), the kernel density estimate of noise given the noise free samples \( \{x_{io}\}_{i=1}^{n} \) and parameters \([\theta, \alpha]\) can be written as

\[
\hat{f}(\epsilon|\theta, \alpha, x_{io}) = \frac{1}{n} \sum_{i=1}^{n} K(H^{-1}(\epsilon - (x_i - x_{io})))
\]

under the constraint \( x_{io}^T \theta = \alpha, ||\theta|| = 1 \). Let us define space \( S = \{\Theta = [\theta, \alpha, \{x_{io}\}_{i=1}^{n}] \mid ||\theta|| = 1, x_{io}^T \theta = \alpha \forall i = 1, ..., n\} \).
Then the model parameters $\Theta \in S$, and assuming the noise to be zero-mean, the maximum likelihood estimate of model parameters is given by

$$\Theta_{ML} = \arg \max_{\theta \in S} \hat{f}(0|\Theta) \quad (15)$$

Note that the above definition is not restrictive, since any shift in the $\epsilon$-space can be accounted for by a shift in $x_{io}$ and $\alpha$. In absence of a disambiguating prior, we assume a zero-mean noise process.

In general, there might be multiple structures in the data, all of which we might need to discover (akin to the Hough transform). Thus, the estimated density function $\hat{f}(0|\Theta)$ might be multimodal. In such a case, one seeks all local minima of the density function. Consequently, we define the parameter estimates as follows,

$$\Theta_{kml} = \arg \max_{\theta \in S} \hat{f}(\epsilon = 0|\Theta) \quad (16)$$

where $\arg \max$ denotes a local maximum. It can be shown that the estimator above is a redescending M-estimator [12].

$\Theta_{kml}$ is a solution to a constrained nonlinear program. The local maximum of $\hat{f}(\cdot)$ can be sought in general by gradient ascent. We now propose an iterative algorithm to seek the constraint that the profile of the kernel, $\kappa$, be convex bounded function, the algorithm is guaranteed to increase the objective function at each iteration and converges to a local maximum.

Since, $\Theta_{kml}$ is constrained to lie in the space $S$, we can define the objective function $q : S \rightarrow \mathbb{R}^+$ as

$$q(\Theta) = \frac{1}{n} \sum_{i=1}^{n} \kappa(-\|H^{-1}(x_i - x_{io})\|^2) \quad (17)$$

Now, let the derivative of the profile $\kappa$ be $\kappa' = g$. Assuming that the initial estimate of $\Theta$ is $\Theta^{(0)}$, and using the convexity of the profile, we see that

$$q(\Theta) - q(\Theta^{(0)}) \geq \frac{1}{n} \sum_{i=1}^{n} g(-\|H^{-1}(x_i - x_{io})\|^2)$$

$$\left(\|H^{-1}(x_i - x_{io})\|^2 - \|H^{-1}(x_i - x_{io})\|^2\right) \quad (18)$$

Defining the weights $w_i = g(-\|H^{-1}(x_i - x_{io})\|^2)$, we seek the next iterate $\Theta^{(1)}$ as the maximizer of right hand side of (18), i.e.,

$$\Theta^{(1)} = \arg \max_{\theta \in S} \sum_{i=1}^{n} w_i(\|H^{-1}(x_i - x_{io})\|^2) \quad (19)$$

The problem in (19) is similar to the ML estimation for i.i.d Gaussian noise samples with identity covariance matrix, as discussed in Section 2. It can be reduced to the following minimization on the space $\tilde{S} = \{[\tilde{\theta}, \tilde{\alpha}, \{x_{io}\}^n_{i=1}] | \tilde{x}_{io}^T \tilde{\theta} = \tilde{\alpha}, \tilde{\theta}^T H^{-2} \tilde{\theta} = 1\}$

$$\tilde{\Theta}^{(1)} = \arg \min_{\theta \in \tilde{S}} \sum_{i=1}^{n} w_i \|H^{-1}x_i - \tilde{x}_{io}\|^2 \quad (20)$$

where $\theta = H^{-1} \tilde{\theta}, \alpha = \tilde{\alpha}, x_{io} = H\tilde{x}_{io}$

$$\tilde{\Theta}^{(1)} = [\tilde{\theta}^{(1)}, \tilde{\alpha}^{(1)}, \{\tilde{x}_{io}^{(1)}\}^n_{i=1}]$$. By Theorem 2.2, we can write

$$[\tilde{\theta}^{(1)}, \tilde{\alpha}^{(1)}] = \arg \min_{\theta \in \tilde{S}} \sum_{i=1}^{n} w_i (\|H^{-1}x_i\|^2 - \|H^{-1}x_i\|^2 - \|H^{-1}x_i\|^2 + 1)$$

$$= \left(\begin{array}{cc} H^{-2} & 0 \\ 0 & 0 \end{array}\right)$$

from the discussion in Section 2, we get

$$[\tilde{\theta}^{(1)}, \tilde{\alpha}^{(1)}]$$ as the generalized eigenvector of $A$ with respect to $B$ corresponding to the minimum eigenvalue. $\Theta^{(1)}$ can thus be estimated using (20).

The above mentioned process is repeated iteratively with new estimates to yield a sequence $\{\Theta^{(n)}\}_{n=1}^{\infty}$ of parameter estimates, and a sequence $\{q(\Theta^{(n)})\}_{n=1}^{\infty}$ of function values. Clearly, for $\Theta = \Theta^{(1)}$, the right hand side of (18) is positive, implying that $q(\Theta^{(1)}) \geq q(\Theta^{(0)})$. The sequence $\{q(\Theta^{(n)})\}_{n=1}^{\infty}$ is thus increasing and bounded above (since $\kappa$ is bounded), implying that it is convergent.

4. Experiments and Results

First, we empirically compare the performance of the algorithm proposed in Section 3 with (a) the total least squares (TLS) solution, and (b) Least Trimmed Squares(LTS) [11]. TLS, as discussed in Section 2, is the optimal estimator for additive, white Gaussian noise (AWGN) and comparison with TLS shows the comparable performance of our algorithm to the optimal solution in case of often-used AWGN model. To test the robustness, we compare our algorithm with Least Trimmed Squares(LTS) which is a state-of-the-art method for robust regression.

We use line fitting in 2D space as the testbed for our experiment. The true samples $(x_{io}, y_{io})$ satisfy $ay_{io} + bx_{io} + e = 0$ where $b = c$, and $a = -1$. The true values are set as $x_{io} = \frac{i}{100} - 1$, and $y_{io} = x_{io} + 1, i = 0, ..., 100$. The data samples $(x_i, y_i)$ are generated by adding uncorrelated noise samples to $(x_{io}, y_{io})$. The noise samples are generated from the Gaussian distribution (a standard noise model) and two-sided log-normal distribution (to simulate outliers) with several different variance values.

Figure 1 shows two sample realizations. At each value of variance, we generated 1000 realizations and computed the
The above steps were applied to the data depicted in Figure 1: Points generated according to the model \((x_i, y_i) = (x_{io}, y_{io}) + \epsilon_i\) and \(\{\epsilon_i\} \in \mathbb{R}^2\) (a). \(\epsilon_i\) is gaussian with mean 0 and variance 0.09. (b) \(\epsilon_i\) is log-normal with \(M = -4\) and \(S = 1.5\).

![Figure 1](image)

Table 1: TLS, LTS, and KML estimates of \((b, c)\) for \((x_i, y_i) = (x_{io}, y_{io}) + \epsilon_i\) and \(\{\epsilon_i\} \in \mathbb{R}^2\) are i.i.d. Gaussian with mean 0 and variance \(\sigma^2 I_2\). Ground-truth values are \((b, c) = (1, 1)\). Mean and deviation of the estimated values for 1000 experiments are presented in the top and bottom four rows, respectively.

<table>
<thead>
<tr>
<th>(\sigma)</th>
<th>TLS</th>
<th>LTS</th>
<th>KML</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.03</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
</tr>
<tr>
<td>0.06</td>
<td>1.002</td>
<td>1.000</td>
<td>0.991</td>
</tr>
<tr>
<td>0.09</td>
<td>1.001</td>
<td>0.998</td>
<td>0.979</td>
</tr>
<tr>
<td>0.12</td>
<td>1.001</td>
<td>1.000</td>
<td>0.963</td>
</tr>
</tbody>
</table>

| Standard Deviation |
|---|---|---|---|
| 0.03 | 0.007 | 0.004 | 0.008 | 0.004 | 0.007 | 0.004 |
| 0.06 | 0.015 | 0.007 | 0.015 | 0.009 | 0.015 | 0.007 |
| 0.09 | 0.020 | 0.013 | 0.023 | 0.013 | 0.020 | 0.013 |
| 0.12 | 0.027 | 0.016 | 0.032 | 0.020 | 0.029 | 0.016 |

Table 2: TLS, LTS, and KML estimates of \((b, c)\) for \((x_i, y_i) = (x_{io}, y_{io}) + \epsilon_i\) and \(\{\epsilon_i\} \in \mathbb{R}^2\) are i.i.d. log-normal with parameter \(\mu_0 = -4\) and \(S^2 = \sigma^2 I_2\). Ground-truth values are \((b, c) = (1, 1)\). Mean and deviation of the estimated values for 1000 experiments are presented in the top and bottom four rows, respectively.

<table>
<thead>
<tr>
<th>(\sigma)</th>
<th>TLS</th>
<th>LTS</th>
<th>KML</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1.000</td>
<td>1.000</td>
<td>0.990</td>
</tr>
<tr>
<td>1.0</td>
<td>1.002</td>
<td>1.000</td>
<td>0.976</td>
</tr>
<tr>
<td>1.5</td>
<td>1.149</td>
<td>0.996</td>
<td>0.953</td>
</tr>
<tr>
<td>2.0</td>
<td>5.314</td>
<td>1.497</td>
<td>0.919</td>
</tr>
</tbody>
</table>

![Figure 2](image)

Figure 2: (a) Intensity image from perceptor Ladar USF Range Database (b) Cartesian coordinates extracted from the range data corresponding to (a).
Figure 3: Three estimated planes 1, 2, and 3 overlayed on the data samples. The planes are shown in different colors. (rotated version of 2(b) for ease of illustration)

Figure 4: Estimated planes 4, 5, and 6 overlayed on the data samples. The planes are shown in different colors. (rotated version of 2(b) for ease of illustration)

This data was extracted from the range image shown in Figure 2(a). There are six planes to be detected in this image. Figures 3 and 4 show the estimated planes overlayed onto the data points. The algorithm was, thus, able to detect all the instances, along with their parameters, in this multiple structure detection problem.

5. Conclusions

In this paper, we presented a novel formalism to solve the problem of robust model fitting using the linear EIV framework. We used nonparametric density estimation to estimate the unknown noise density and use a maximum likelihood approach for robust estimation of model parameters. The implication of such an approach was that the data could consist of multiple model instances and unknown noise. We also proposed a provably convergent iterative algorithm to solve the resultant optimization problem. The algorithm uses iterative quadratic approximation to the likelihood function based on a variation formulation using the convexity of density kernel functions. The performance of the proposed algorithm favorably compares to two popular algorithms — the Least Trimmed Squares (LTS), and the Total Least Squares (TLS). Results for model fitting on real range data are also provided.

References