Robust Bayesian Regression via Hard Thresholding

Zhaohui Li

Academy of Mathematics and System Sciences, Institute of Systems Science

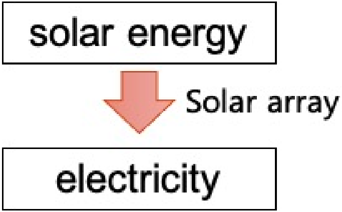
A joint work with Zheyi Fan and Qingpei Hu

Overview

1. Introduction and Motivation
2. Robust Bayesian Regression via Hard Thresholding
3. Satellite Solar Cell Data Reconstruction Algorithm
4. Result
5. Conclusion

1. Introduction and Motivation
2. Robust Bayesian Regression via Hard Thresholding
3. Satellite Solar Cell Data Reconstruction Algorithm
4. Result
5. Conclusion

Introduction



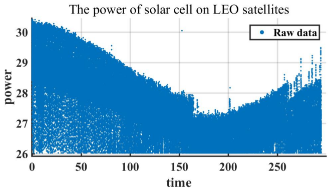
solar energy

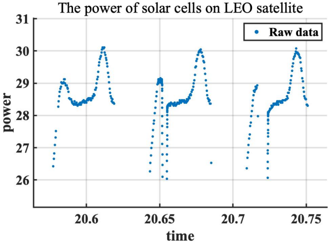
Solar array

electricity

- Solar array is a very important component of satellite, which is the key factor for satellite to work normally.
- The solar array provides power for the entire satellite platform, so the reliability of the solar array is a very important research topic.
- By analyzing the on orbit data of the solar array, we can obtain various reliability indexes and predict the life of solar array.

Problem





Data characteristics

- The data is periodic and has an overall change trend.
- The total amount of data is large, about 20 million in total.
- The outliers account for a high proportion and their distribution is rather complex, so it is difficult to deal with them by common denoising algorithms.

Our Goals: To reconstruct the data and extract the overall change trend.

The weaknesses of previous Methods

Low breakdown point

- (McWilliams et al. 2014): $\alpha = O(1/\sqrt{d})$.
- (Prasad et al. 2018): $\alpha = O(1/\log d)$.

Can only resist OAA (Oblivious adversarial attack)

- (Bhatia et al. 2017): proposed the first efficient consistent estimator.
- (Suggala et al. 2019): α is close to 1 as $n \to \infty$.

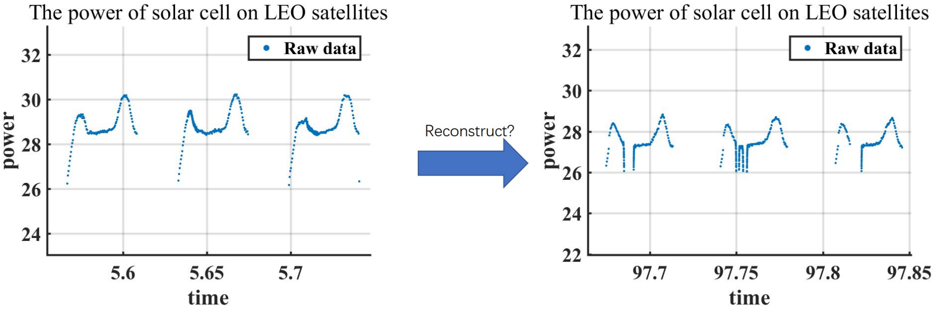
The weaknesses of previous Methods

Can resist AAA (Adaptive adversarial attack)

- Have a low breakdown point
  - (Cherapanamje et al. 2020)(Jambulapat et al.2021)(Pensia et al. 2020)
- (Bhatia et al. 2015): only in noiseless case.
- (Diakonikolas et al. 2019): requires accurate information of data covariance.

How to **increase** the breakdown point of robust regression under AAA?

Motivation



Reconstruct?

- We know the distribution pattern of normal data. Can we use this information to reconstruct data?
- In the reconstructed data, we can select a representative point in each period to get the overall data trend.

1. Introduction and Motivation
2. Robust Bayesian Regression via Hard Thresholding
3. Satellite Solar Cell Data Reconstruction Algorithm
4. Result
5. Conclusion

Basic aspects in robust regression

Given a data matrix $X = [\mathbf{x}_1, ..., \mathbf{x}_n] \in \mathbb{R}^{d\times n}$, its corresponding response vector $\mathbf{y} \in \mathbb{R}^n$, and a certain number k which represents the number of corruptions in the data, the problem of RLSR can be described as:

$$(\hat{\mathbf{w}}, \hat{S}) = \arg \min_{\substack{\mathbf{w}\in\mathbb{R}^p, S\subset[n] \\ |S|=n-k}} \sum_{i\in S} (y_i - \mathbf{x}_i^T \mathbf{w})^2$$

Basic aspects in robust regression

A commonly used data generation model is：

$$\mathbf{y} = X^T\mathbf{w}^* + \mathbf{b}^* + \boldsymbol{\epsilon}$$

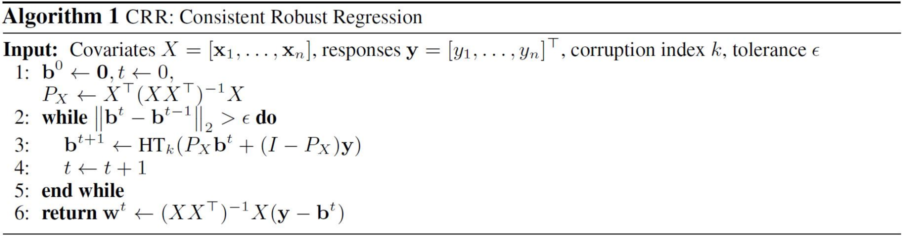
$\mathbf{w}^*$: the true regression coefficient we want to recover.

$\boldsymbol{\epsilon}$: The dense white noise vector subject to a specific distribution.

$\mathbf{b}^*$: a k-sparse vector with only k non 0-values, representing k unbounded noise in the response vector.

Based Work

Consistent Robust Regression (CRR) (Bhatia et al. 2017)



Algorithm 1 CRR: Consistent Robust Regression

```
Input: Covariates $X = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, responses $\mathbf{y} = [y_1, \ldots, y_n]^\top$, corruption index $k$, tolerance $\epsilon$
$\mathbf{b}^0 \leftarrow \mathbf{0}, t \leftarrow 0$,
$P_X \leftarrow X^\top (XX^\top)^{-1} X$
while $\left\|\mathbf{b}^t - \mathbf{b}^{t-1}\right\|_2 > \epsilon$ do
    $\mathbf{b}^{t+1} \leftarrow \mathrm{HT}_k(P_X \mathbf{b}^t + (I - P_X)\mathbf{y})$
    $t \leftarrow t + 1$
end while
return $\mathbf{w}^t \leftarrow (XX^\top)^{-1} X(\mathbf{y} - \mathbf{b}^t)$
```

The key step in this CRR algorithm is $\mathbf{b}^{t+1} \leftarrow HT_k(P_X\mathbf{b}^t + (I - P_X)\mathbf{y})$. This step can be divided into two sub steps:

$$\mathbf{w}^{t+1} \leftarrow (XX^T)^{-1} X(\mathbf{y} - \mathbf{b}^t)$$

$$\mathbf{b}^{t+1} \leftarrow HT_k(\mathbf{y} - X^T\mathbf{w}^{t+1})$$

This means $\mathbf{w}^{t+1} = \arg\min_{\mathbf{w}} \|\mathbf{y} - \mathbf{b}^t - X^T\mathbf{w}\|^2$

TRIP:Hard Thresholding Approach to Robust Regression with Simple Prior

CRR

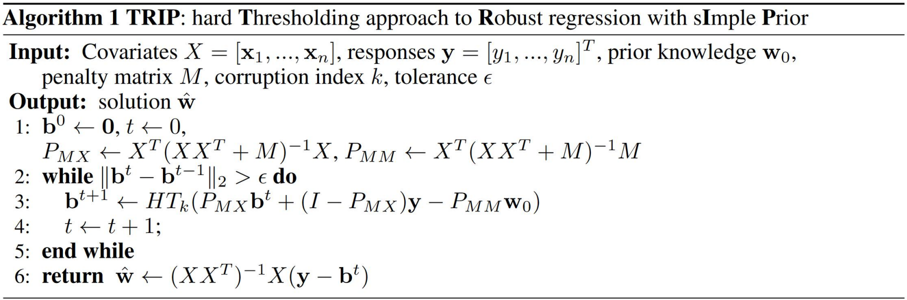
$$\mathbf{w}^t = \arg\min_{\mathbf{w}} \|\mathbf{y} - \mathbf{b}^t - X^T\mathbf{w}\|^2$$
$$\mathbf{w}^t \leftarrow (XX^T)^{-1}X(\mathbf{y} - \mathbf{b}^t)$$
$$\mathbf{b}^{t+1} \leftarrow HT_k(\mathbf{y} - X^T\mathbf{w}^t)$$

TRIP:Hard Thresholding Approach to Robust Regression with Simple Prior (Ours)

$$\mathbf{w}^t = \arg\min_{\mathbf{w}} \|\mathbf{y} - \mathbf{b}^t - X^T\mathbf{w}\|^2 + (\mathbf{w} - \mathbf{w}_0)^T M(\mathbf{w} - \mathbf{w}_0)$$
$$\mathbf{w}^t \leftarrow (XX^T + M)^{-1}X(\mathbf{y} - \mathbf{b}^t + M\mathbf{w}_0)$$
$$\mathbf{b}^{t+1} \leftarrow HT_k(\mathbf{y} - X^T\mathbf{w}^t)$$

TRIP can be viewed as using MAP to estimate $\mathbf{w}$ by giving $\mathbf{w}$ a prior $\mathcal{N}(\mathbf{w}_0, \Sigma_0)$ and $M = (\Sigma_0/\sigma^2)^{-1}$ in Bayesian view.

TRIP



Algorithm 1 TRIP: hard **T**hresholding approach to **R**obust regression with s**I**mple **P**rior

```
Input: Covariates $X = [\mathbf{x}_1, ..., \mathbf{x}_n]$, responses $\mathbf{y} = [y_1, ..., y_n]^T$, prior knowledge $\mathbf{w}_0$,
    penalty matrix $M$, corruption index $k$, tolerance $\epsilon$
Output: solution $\hat{\mathbf{w}}$
$\mathbf{b}^0 \leftarrow \mathbf{0}, t \leftarrow 0,$
$P_{MX} \leftarrow X^T(XX^T + M)^{-1}X, P_{MM} \leftarrow X^T(XX^T + M)^{-1}M$
while $\|\mathbf{b}^t - \mathbf{b}^{t-1}\|_2 > \epsilon$ do
    $\mathbf{b}^{t+1} \leftarrow HT_k(P_{MX}\mathbf{b}^t + (I - P_{MX})\mathbf{y} - P_{MM}\mathbf{w}_0)$
    $t \leftarrow t + 1;$
end while
return $\hat{\mathbf{w}} \leftarrow (XX^T)^{-1}X(\mathbf{y} - \mathbf{b}^t)$
```

Theorem (The convergence of TRIP)

Under the conditions of Theorem 1, and assuming that $\mathbf{x}_i \in \mathbb{R}^d$ are generated from the standard normal distribution, if $k > k^*$, it is guaranteed with a probability of at least $1 - \delta$ that, for any $\varepsilon, \delta > 0$, the current estimation coefficient $\mathbf{w}_{T_0}$ satisfies $\|\mathbf{w}_{T_0} - \mathbf{w}^*\|_2 \leq O(\frac{1}{\sqrt{n}})(\varepsilon + e_0) + O(\frac{\sqrt{k+k^*}\lambda_{\max}(M)}{n^{3/2}})\|\mathbf{w}^* - \mathbf{w}_0\|_2$ after $T_0 = O(\log(\frac{\|\mathbf{b}^*\|_2}{\varepsilon}))$ steps.

The Convergence Condition

The Convergence Condition of CRR

$$2\frac{\Lambda_{k+k^*}}{\lambda_{min}(XX^T)} < 1$$

The Convergence Condition of TRIP

$$2\frac{\Lambda_{k+k^*}}{\lambda_{min}(XX^T+M)} < 1$$

Notice that $\lambda_{min}(XX^T + M) \geq \lambda_{min}(XX^T) + \lambda_{min}(M)$, so TRIP need a weaker condition for convergence than CRR.

Under the condition $\lim_{n\to\infty} \frac{\lambda_{min}(M)}{n} = \xi$, we can give an approximate expression of the breakdown point for TRIP when ξ is not too large

$$k^* \leq k \leq (0.3023 - \sqrt{0.0887 - 0.0040\xi})n$$

How To Decrease The Bias?

In the convergence theory of TRIP, there is a unavoidable bias in the estimation:

$$O(\frac{\sqrt{k+k^*}\lambda_{\max}(M)}{n^{3/2}})\|\mathbf{w}^* - \mathbf{w}_0\|_2$$

In TRIP, the estimation of $\mathbf{w}^*$ in every iteration still uses least-squares with a penalty term, which is still easily affected by the corrupted points. This disadvantage forces us to assign a higher weight to the prior to resist severe data corruption in the case of AAAs. However, a higher weight on the prior means a larger estimation bias.

If every iteration step is **more robust**, will the estimation bias decrease?

BRHT: Robust Bayesian Reweighting Regression via Hard Thresholding

TRIP:Hard Thresholding Approach to Robust Regression with Simple Prior

$$
\begin{aligned}
&\mathbf{w}^t = \arg\min_{\mathbf{w}} \|\mathbf{y} - \mathbf{b}^t - X^T\mathbf{w}\|^2 + (\mathbf{w} - \mathbf{w}_0)^T M(\mathbf{w} - \mathbf{w}_0) \\
&\mathbf{w}^t \leftarrow (XX^T + M)^{-1} X(\mathbf{y} - \mathbf{b}^t + M\mathbf{w}_0) \\
&\mathbf{b}^{t+1} \leftarrow HT_k(\mathbf{y} - X^T\mathbf{w}^t)
\end{aligned}
$$

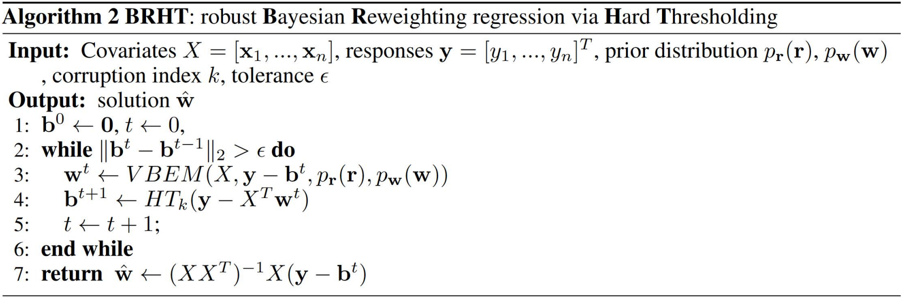
BRHT: Robust Bayesian Reweighting Regression via Hard Thresholding

$$
(\mathbf{w}^t, \mathbf{r}^t) = \arg \max_{\mathbf{w}\in\mathbb{R}^d, \mathbf{r}\in\mathbb{R}_+^n} \log p_{\mathbf{w}}(\mathbf{w}) + \log p_{\mathbf{r}}(\mathbf{r}) + \sum_{i=1}^{n} r_i \log \ell(y_i - b_i^t \mid \mathbf{w}, \mathbf{x}_i, \sigma^2)
$$

$$
\mathbf{b}^{t+1} \leftarrow HT_k(\mathbf{y} - X^T\mathbf{w}^t)
$$

The estimation of $\mathbf{w}$ is based on Bayesian Reweighting from (Wang et al. 2017), which is a robust Bayesian model for estimating parameters. $\mathbf{r}$ is the weight on each point. The prior $\mathbf{w}$ is also in the form $\mathcal{N}(\mathbf{w}_0, \Sigma_0)$.

BRHT



Algorithm 2 BRHT: robust **B**ayesian **R**eweighting regression via **H**ard **T**hresholding

```
Input: Covariates $X = [\mathbf{x}_1, ..., \mathbf{x}_n]$, responses $\mathbf{y} = [y_1, ..., y_n]^T$, prior distribution $p_\mathbf{r}(\mathbf{r})$, $p_\mathbf{w}(\mathbf{w})$
    , corruption index $k$, tolerance $\epsilon$
Output: solution $\hat{\mathrm{w}}$
$\mathbf{b}^0 \leftarrow \mathbf{0}$, $t \leftarrow 0$,
while $\|\mathbf{b}^t - \mathbf{b}^{t-1}\|_2 > \epsilon$ do
    $\mathbf{w}^t \leftarrow VBEM(X, \mathbf{y} - \mathbf{b}^t, p_\mathbf{r}(\mathbf{r}), p_\mathbf{w}(\mathbf{w}))$
    $\mathbf{b}^{t+1} \leftarrow HT_k(\mathbf{y} - X^T\mathbf{w}^t)$
    $t \leftarrow t + 1$;
end while
return $\hat{\mathrm{w}} \leftarrow (XX^T)^{-1}X(\mathbf{y} - \mathbf{b}^t)$
```

Theorem (The convergence of BRHT)

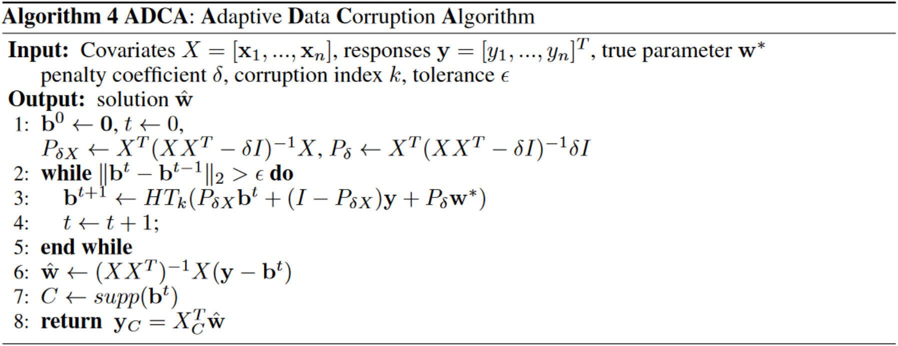
Under the conditions of Theorem 4 and with $\mathbf{x}_i \in \mathbb{R}^d$ generated from the standard normal distribution, there exist $\alpha > 0$ and $0 < \gamma \leq 1 + \epsilon$, where ϵ is a small number, such that if $k > k^*$ and Σ in the prior $p_\mathbf{w}(\mathbf{w})$ is $\alpha\sigma^2 M^{-1}$, it can be guaranteed with a probability of at least $1 - \delta$ that, for any $\varepsilon, \delta > 0$, the current estimation coefficient $\mathbf{w}_{T_0}$ satisfies $\|\mathbf{w}_{T_0} - \mathbf{w}^*\|_2 \leq O(\frac{1}{\sqrt{n}})(\varepsilon + e_0) + O(\frac{\sqrt{k+k^*}\lambda_{\max}(M)}{n^{3/2}})\gamma\|\mathbf{w}^* - \mathbf{w}_0\|_2$ after $T_0 = O(\log(\frac{\gamma\|\mathbf{b}^*\|_2}{\varepsilon}))$ steps.

The design of OAA and AAA

OAA: The set of corrupted points S is selected as a uniformly random k-sized subset of $[n]$, and the corresponding response variables are set as $y_i = \mathbf{x}_i^T \mathbf{w}^* + b_i + \epsilon_i$, where b_i are sampled from the uniform distribution $U[0, 10]$ and the white noise $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

AAA: We use all information from the true data distribution to corrupt the data, and propose an adaptive data corruption algorithm (ADCA). δ is set to $0.1n$ for $n = 1000$, $p = 200$, and to $0.2n$ for $n = 2000$, $p = 100$.

Algorithm of AAA attack

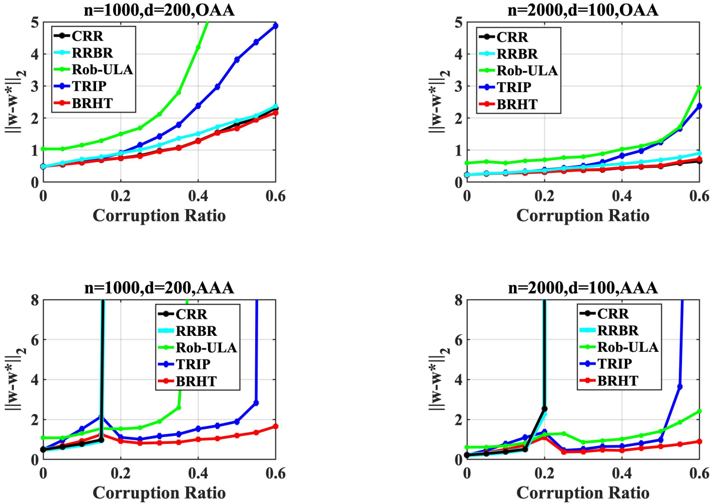


Algorithm 4 ADCA: Adaptive Data Corruption Algorithm

```
Input: Covariates $X = [\mathbf{x}_1, ..., \mathbf{x}_n]$, responses $\mathbf{y} = [y_1, ..., y_n]^T$, true parameter $\mathbf{w}^*$
    penalty coefficient $\delta$, corruption index $k$, tolerance $\epsilon$
Output: solution $\hat{\mathbf{w}}$
$\mathbf{b}^0 \leftarrow \mathbf{0}, t \leftarrow 0,$
$P_{\delta X} \leftarrow X^T(XX^T - \delta I)^{-1}X,\ P_\delta \leftarrow X^T(XX^T - \delta I)^{-1}\delta I$
while $\|\mathbf{b}^t - \mathbf{b}^{t-1}\|_2 > \epsilon$ do
    $\mathbf{b}^{t+1} \leftarrow HT_k(P_{\delta X}\mathbf{b}^t + (I - P_{\delta X})\mathbf{y} + P_\delta \mathbf{w}^*)$
    $t \leftarrow t + 1;$
end while
$\hat{\mathbf{w}} \leftarrow (XX^T)^{-1}X(\mathbf{y} - \mathbf{b}^t)$
$C \leftarrow supp(\mathbf{b}^t)$
return $\mathbf{y}_C = X_C^T\hat{\mathbf{w}}$
```

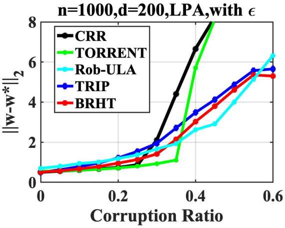
ADCA is similar with TRIP algorithm, but will seriously destroys the data, which conventional methods can't resist it.

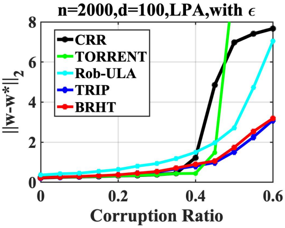
The recovery effects of TRIP and BRHT

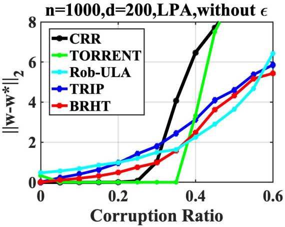


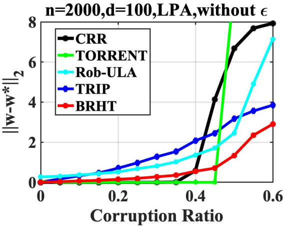
Our method is more robust than other methods and shows excellent performance under complex attacks.

The reliability of TRIP and BRHT under other attack



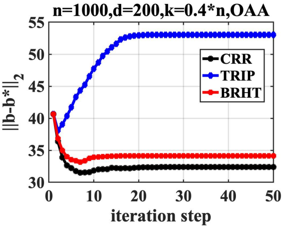


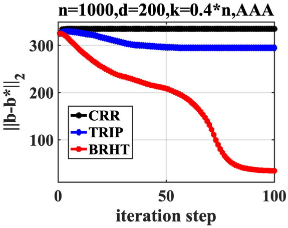


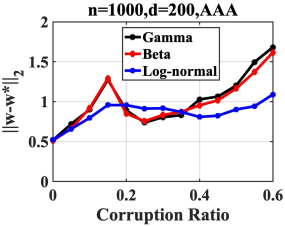


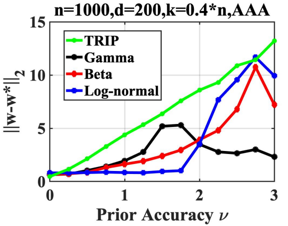
Under this new severe attack, our methods still maintain good performance.

The stability of TRIP and BRHT





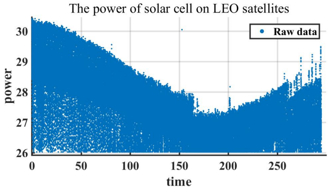


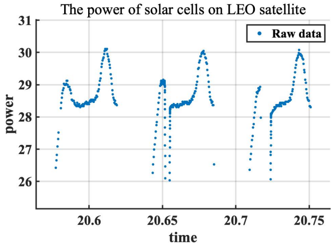


BRHT is more robust than TRIP under AAAS, and both methods are not too sensitive to the prior.

1. Introduction and Motivation
2. Robust Bayesian Regression via Hard Thresholding
3. Satellite Solar Cell Data Reconstruction Algorithm
4. Result
5. Conclusion

The introduction of Satellite Solar Cell Data





Methods to solve data reconstruction problem

- Select a period with few corruptions as the prior data.
- Remove the time deviation of data in each period.
- Using robust Bayesian regression method to reconstruct the data of each period.
- Select representative points in each period to find the overall trend.

Prior Information

We used the parameters of one standard period as the prior data. The standard period is obtained by manually removing the outliers.

Fitting method

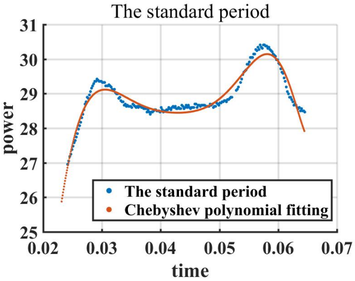
- Chebyshev polynomials are used as the basis of linear regression because the results obtained are relatively stable

$$p_{st}(x) = \sum_{i=0}^{m} \gamma_i^{st} T_i(t) \qquad (1)$$

$$T_0(t) = 1 \qquad (2)$$

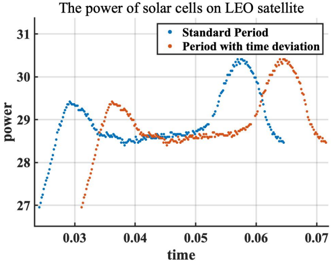
$$T_1(t) = t \qquad (3)$$

$$T_{n+1}(t) = 2tT_n(t) - T_{n-1}(t) \qquad (4)$$



Remove Time Deviation

The time deviation in each period will cause large deviation in regression coefficient.



Remove main time deviation

$$\tilde{t}_{ij} = t_{ij} - \bar{t}_i \tag{5}$$

Further remove minor time deviation

$$p_i(t) = p_0^i(t + h_i) \approx \sum_{i=0}^{m} \gamma_{ij} T_i(t) + h_i p_0^{i\,'}(t) \tag{6}$$

As for each period, the shape of the data is roughly similar, so we can use $p_{st}^{'}(t)$ to replace $p_0^{i\,'}(t)$.

$$p_i(t) = \sum_{i=0}^{m} \gamma_{ij} T_i(t) + h_i p_{st}^{'}(t) \tag{7}$$

Where $p_{st}^{'}(t)$ can be calculated by $p_{st}^{'}(t) = \sum_{j=0}^{m} \gamma_j^{st} T_i^{'}(t)$.

Data reconstruction of each period

So the data reconstruction problem for each period becomes:

$$p_i(t) = \sum_{i=0}^{m} \gamma_{ij} T_i(t) + h_i p'_{st}(t) \tag{8}$$

The reconstruction method adopts **TRIP** algorithm as its high-speed calculation. The penalty matrix adopts the following form

$$M = s \begin{bmatrix} 0 & 0 & 0 \\ 0 & I_m & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{9}$$

Subsequently, the parameters $[\gamma_0^{st}, ..., \gamma_m^{st}, h_{st}]$ of the standard period are used as the prior information, where $h_{st} = 0$.

Final Denoising Algorithm

Parameter setting

The degree of Chebyshev polynomial is 9. We define $X_i = [X_i^1, ..., X_i^{n_i}]$, where $X_i^j = [T_0(\tilde{t}_{ij}), ..., T_9(\tilde{t}_{ij}), p_{st}^{'}(\tilde{t}_{ij})]^T$ is the covariant of the time point t_{ij}, and the penalty matrix is set as

$$M = \begin{pmatrix} 0 & 0 & 0 \\ 0 & I_9 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{10}$$

Algorithm 5 Solar cell denoising algorithm

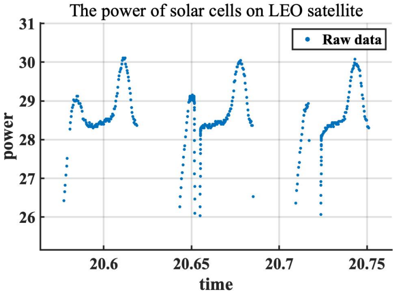
Input: Solar cell data of LEO satellite (t_{ij}, p_{ij}), standard period parameter $\mathbf{w}_0$ penalty matrix M, corruption ratio α, tolerance ϵ

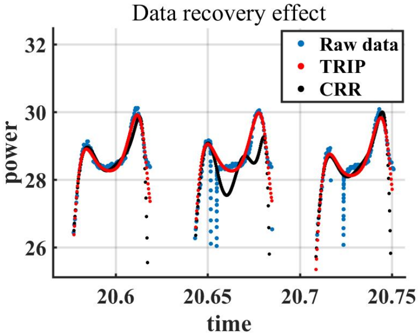
Output: cleaned data $(t_{ij}, p_i(\tilde{t}_{ij}))$

```
while $i < n_{period}$ do
    $\bar{t}_i = (\sum_j t_{ij})/n_i$;
    for $j = 1 : n_i$ do
        $\tilde{t}_{ij} = t_{ij} - \bar{t}_i$
    end for
    $k = \alpha n_i$;
    $\mathbf{w}_i \leftarrow TRIP(X_i, p_i, \mathbf{w}_0, M, k, \epsilon)$
    $(t_{ij}, p_{ij}) \leftarrow (t_{ij}, X_i^j \mathbf{w}_i)$
end while
return $\{(t_{ij}, p_{ij})\}$
```

1. Introduction and Motivation
2. Robust Bayesian Regression via Hard Thresholding
3. Satellite Solar Cell Data Reconstruction Algorithm
4. **Result**
5. Conclusion

Result



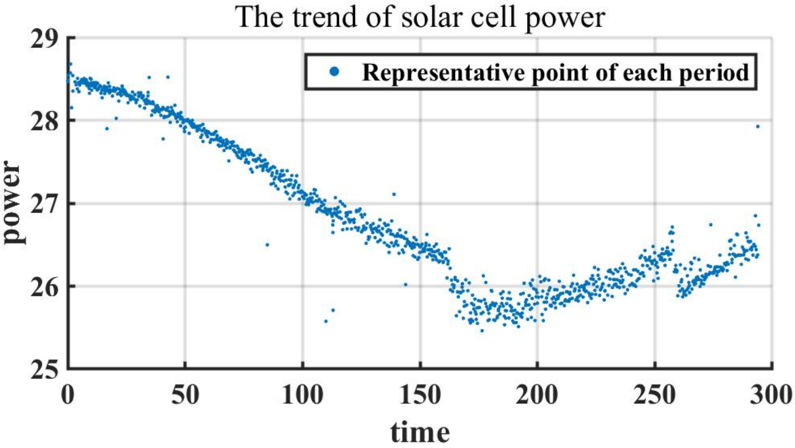


Our method successfully recovered the original data, which means that the recovered data can be further used to extract data trend information

Result

Representative Point Extraction

- In each period, we use a special point $(\bar{t}_i + t_c, p_i(t_c))$ as the representative of this period, where t_c is a determined value in advance.



1. Introduction and Motivation

2. Robust Bayesian Regression via Hard Thresholding

3. Satellite Solar Cell Data Reconstruction Algorithm

4. Result

5. Conclusion

Conclusion

- We propose two novel robust regression algorithms TRIP and BRHT, which can tolerate a larger proportion of outliers by incorporating prior information.
- Through TRIP algorithm, we propose a satellite solar array denoising algorithm, and perfectly recover the corrupted data.
- By extracting representative points from the recovered data, we reveal the power variation trend of the satellite solar array.

Thanks for Listening!