Handling Imperfections in Energy Disaggregation

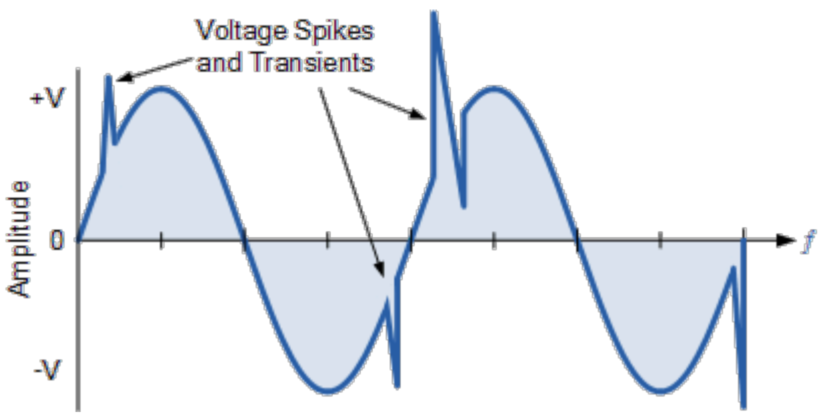
Authors: Angshul Majumdar, Megha Gupta

Presenter: Manoj Gulati

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Spikes or Surges

Short-term increases in the electrical supply voltage, or current or both, are called ‘spikes’ (also called ‘surges’)

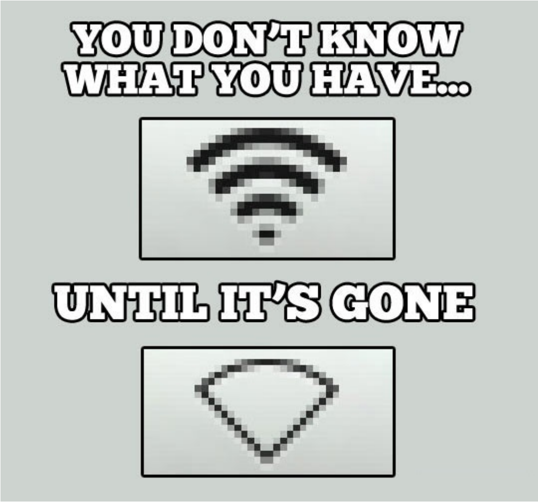


Spikes or Surges

- They are inevitable.
- Why?
  - Lightning strikes & Power outages
  - Tripped circuit breakers & Short circuits
  - Power transitions in other large equipment on the same power line
  - Malfunctions caused by the power company
  - Electromagnetic pulses (EMP)
    with electromagnetic energy distributed typically up to the 100 kHz and 1 MHz frequency range.

Missing Data

- Power-meter acquires data, but cannot be transmitted.
- (If there is any other reason – add)



Sparse coding - Training

 — 

$$X_{dishwasher} = D_1 Z_1 \equiv \min_{D_1 Z_1} \left\| X_{dishwasher} - D_1 Z_1 \right\|_F^2 + \lambda \left\| Z_1 \right\|_1$$

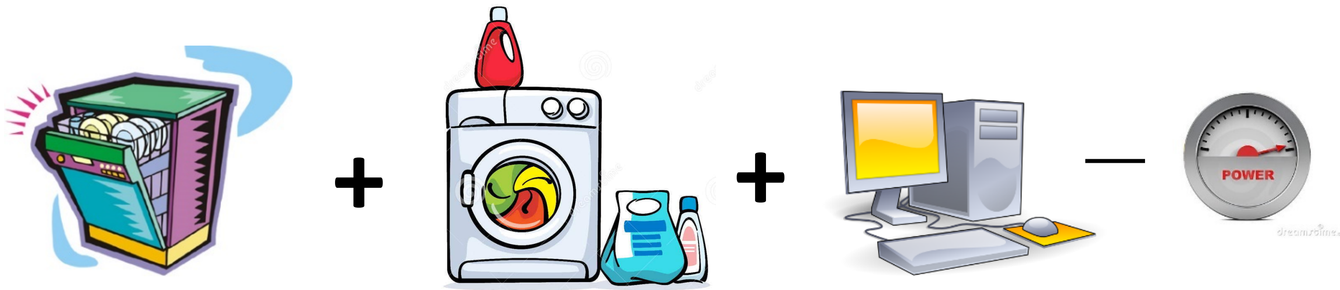
 — 

$$X_{washer} = D_2 Z_2 \equiv \min_{D_2 Z_2} \left\| X_{washer} - D_2 Z_2 \right\|_F^2 + \lambda \left\| Z_2 \right\|_1$$

 — 

$$X_{desktop} = D_3 Z_3 \equiv \min_{D_3 Z_3} \left\| X_{desktop} - D_3 Z_3 \right\|_F^2 + \lambda \left\| Z_3 \right\|_1$$

Sparse Coding - Disaggregation



$$X_{dishwasher} + X_{washer} + X_{desktop} = X$$

$$D_1 Z_1 + D_2 Z_2 + D_3 Z_3 = X$$

Dictionaries are already learnt in the training phase

$$\min_{Z_1, Z_2, Z_3} \left\| X - \left[D_1 \mid D_2 \mid D_3 \right] \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} \right\|_F^2 + \lambda \left\| \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} \right\|_1$$

$$\hat{X}_{dishwasher} = D_1 Z_1 ; \hat{X}_{washer} = D_2 Z_2 ; \hat{X}_{desktop} = D_2 Z_2$$

HANDLING SURGES

Gaussianity & Mean Squared Error

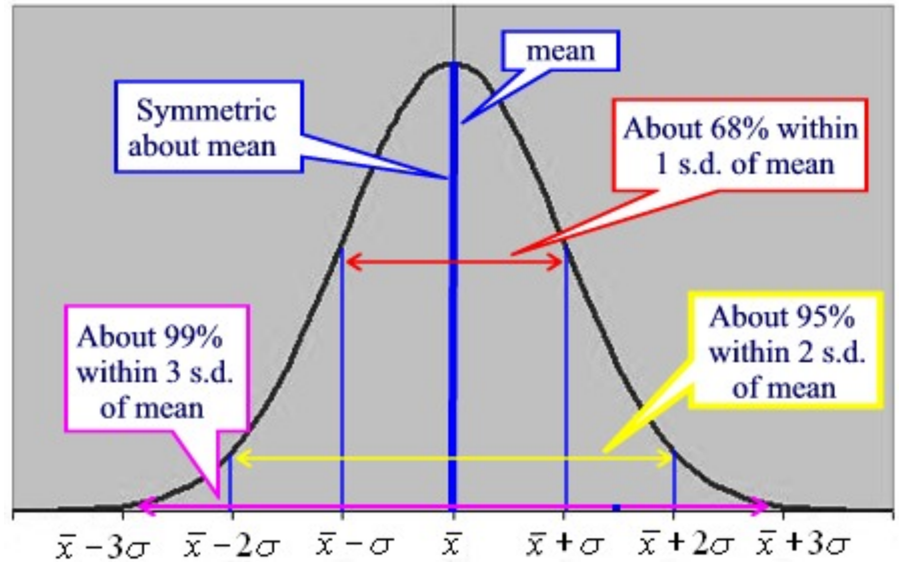
Zhou Wang and Alan C. Bovik

For more than 50 years, the mean-squared error (MSE) has been the dominant quantitative performance metric in the field of signal processing. It remains the standard criterion for the assessment of signal quality and fidelity; it is the method of choice for comparing competing signal processing methods and systems, and, perhaps most importantly, it is the nearly ubiquitous preference of design engineers seeking to optimize signal processing algorithms. This is true despite the fact that in many of these applications, the MSE exhibits weak performance and has been widely criticized for serious shortcomings, especially when dealing with perceptually important signals such as speech and images. Yet the MSE has exhibited remarkable staying power, and prevailing attitudes towards the MSE seem to range from "it's easy to use and not so bad" to "everyone else uses it."

So what is the secret of the MSE—why is it still so popular? And is this popularity misplaced? What is wrong with the MSE when it does not work well? Just how wrong is the MSE in these cases? If not the MSE, what else can be used? These are the questions we'll be concerned with in this article. Our backgrounds are primarily in the field of image processing, where the MSE has a particularly bad reputation, but where, ironically, it is used nearly as much as in other areas of signal processing. Our discussion will often deal with the role of the MSE (and alternative methods) for processing visual signals. Owing to the poor performance of the MSE as a visual metric, interesting alternatives are arising in the image processing field. Our goal is to stimulate fruitful thought and discussion regarding the role of the MSE in processing other types of signals. More specifically, we hope to inspire signal processing engineers to rethink whether the MSE is truly the criterion of choice in their own theories and applications, and whether it is time to look for alternatives.

Mean Squared Error: Love It or Leave It?

[A new look at signal fidelity measures]

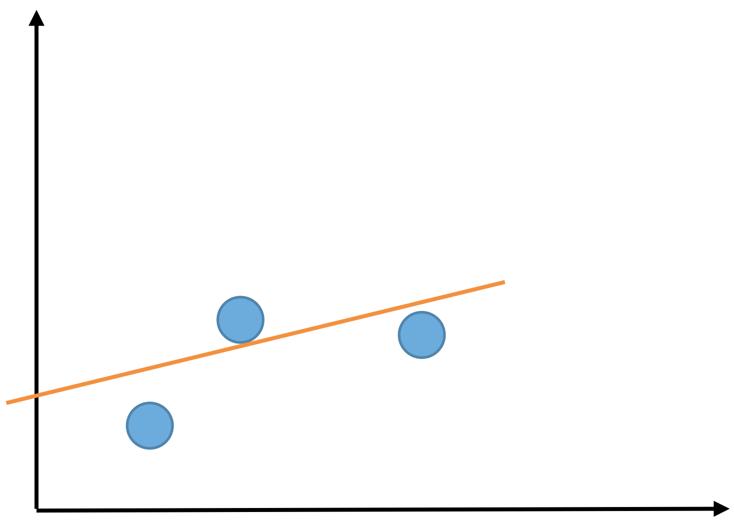


Smooth cost function – easy to optimize

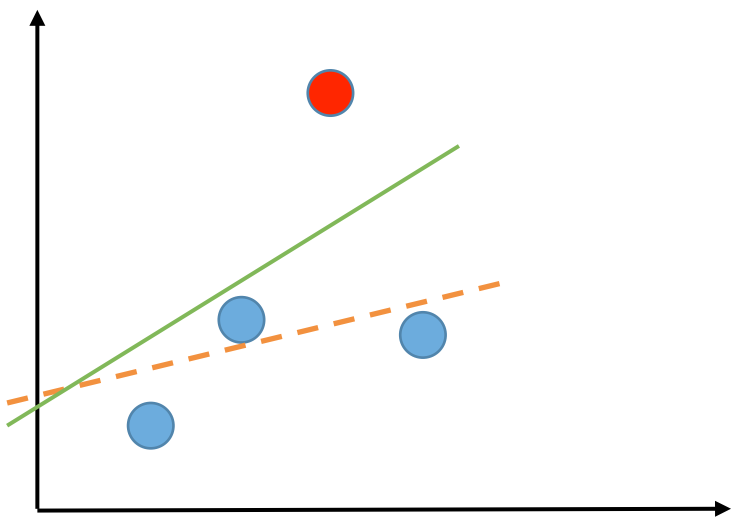
Perfect to use when error is small / probability of large error is negligible

But not for large outliers!

Line-fitting – with outliers



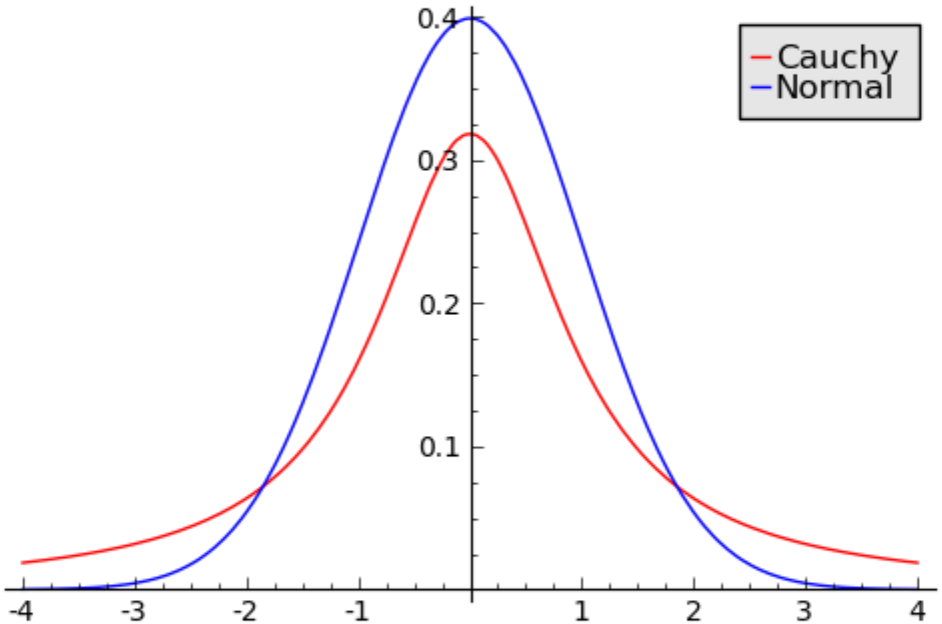
Fitting a line in presence of noise (small errors) with least squares



Fitting a line in presence of outliers (large errors) with least squares

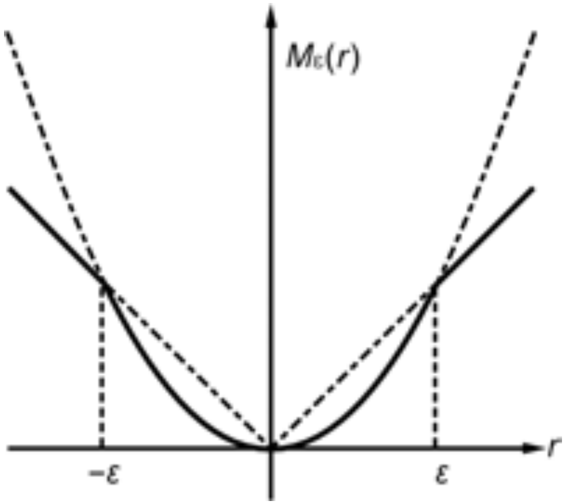
Outliers – Heavy Tailed Distribution

- Heavy tails – probability of large values are not small, e.g. Cauchy



- Used for modelling outliers (Cauchy, Exponential, etc.)

Robust Estimation



- Huber function – have been widely used for robust estimation.
- But more recently absolute deviations are being minimized.

- P. J. Huber, “Robust Estimation of a Location Parameter”, The Annals of Mathematical Statistics, Vol. 35 (1), pp. 73-101, 1964.
- R. L. Branham Jr., "Alternatives to least squares", Astronomical Journal 87, pp. 928–937, 1982.
- M. Shi and M. A. Lukas, "An L1 estimation algorithm with degeneracy and linear constraints". Computational Statistics & Data Analysis, Vol. 39 (1), pp. 35–55, 2002.
- L. Wang, M. D. Gordon and J. Zhu, "Regularized Least Absolute Deviations Regression and an Efficient Algorithm for Parameter Tuning". IEEE ICDM. pp. 690–700, 2006.

Robust Dictionary Learning - NILM

- Change all the cost function from l2-norm to l1-norm.
- During training –

$$\min_{D_i Z_i} \left\| X_i - D_i Z_i \right\|_1 + \lambda \left\| Z_i \right\|_1$$

- During Disaggregation –

$$\min_{Z_1,\ldots,Z_3} \left\| X - \left[D_1 \mid \ldots \mid D_N \right] \begin{bmatrix} Z_1 \\ \ldots \\ Z_N \end{bmatrix} \right\|_1 + \lambda \left\| \begin{bmatrix} Z_1 \\ \ldots \\ Z_N \end{bmatrix} \right\|_1$$

A. Majumdar and R. K. Ward "Robust Dictionary Learning: Application to Signal Disaggregation", ICASSP 2016

HANDLING MISSING DATA

Modelling Missing Data

- Prior techniques were based on interpolation
  - Nearest neighbor
  - Previous value
  - Bicubic
- Interpolating leads to ‘errors’.
- Mathematically, a more optimal way would be to simply model the missing values, i.e.

$$Y = R \odot X$$

$X -$ actual data collected by meter

$Y -$ data received

$R -$ binary sampling mask

Training and Disaggregation

- Blind Compressed Sensing (BCS) for training –

$$Y_i = R_i \odot X_i = R_i \odot D_i Z_i$$

$$BCS: \min_{D_i, Z_i} \left\| Y_i - R_i \odot D_i Z_i \right\|_F^2 + \lambda \left\| Z_i \right\|_1$$

- Simple Compressed Sensing type for Disaggregation

$$Y = R \odot X = R \odot \left[D_1 | \ldots | D_N \right] \begin{bmatrix} Z_1 \\ \ldots \\ Z_N \end{bmatrix}$$

$$BCS: \min_{Z_1, \ldots, Z_N} \left\| Y - R \odot \left[D_1 | \ldots | D_N \right] \begin{bmatrix} Z_1 \\ \ldots \\ Z_N \end{bmatrix} \right\|_F^2 + \lambda \left\| \begin{bmatrix} Z_1 \\ \ldots \\ Z_N \end{bmatrix} \right\|_1$$

FULL FORMULATION

Handling Both

- Training

$$\min_{D_i, Z_i} \left\| Y_i - R_i \odot D_i Z_i \right\|_1 + \lambda \left\| Z_i \right\|_1$$

Missing Data Surges / Spikes

- Disaggregation

$$\min_{Z_1, \ldots, Z_N} \left\| Y - R \odot \left[D_1 \mid \ldots \mid D_N \right] \begin{bmatrix} Z_1 \\ \ldots \\ Z_N \end{bmatrix} \right\|_1 + \lambda \left\| \begin{bmatrix} Z_1 \\ \ldots \\ Z_N \end{bmatrix} \right\|_1$$

Results

REDD Dataset: Standard Protocol

House	Training Accuracy			Testing Accuracy		
	Robust DL	Powerlet	Proposed	Robust	Powerlet	Proposed
1	75.5	81.6	77.0	53.0	46.0	54.5
2	66.7	79.0	69.1	56.3	49.2	58.0
3	65.2	61.8	67.0	43.9	31.7	45.7
4	63.7	58.5	65.9	60.1	50.9	61.6
6	68.5	79.1	70.2	60.2	54.5	62.0

Future Work

- We proposed ‘deep sparse coding’
  - Instead of learning a single level of dictionary, we learn deeper representations.
- In future, we will combine the proposed work with deep sparse coding.

THANK YOU