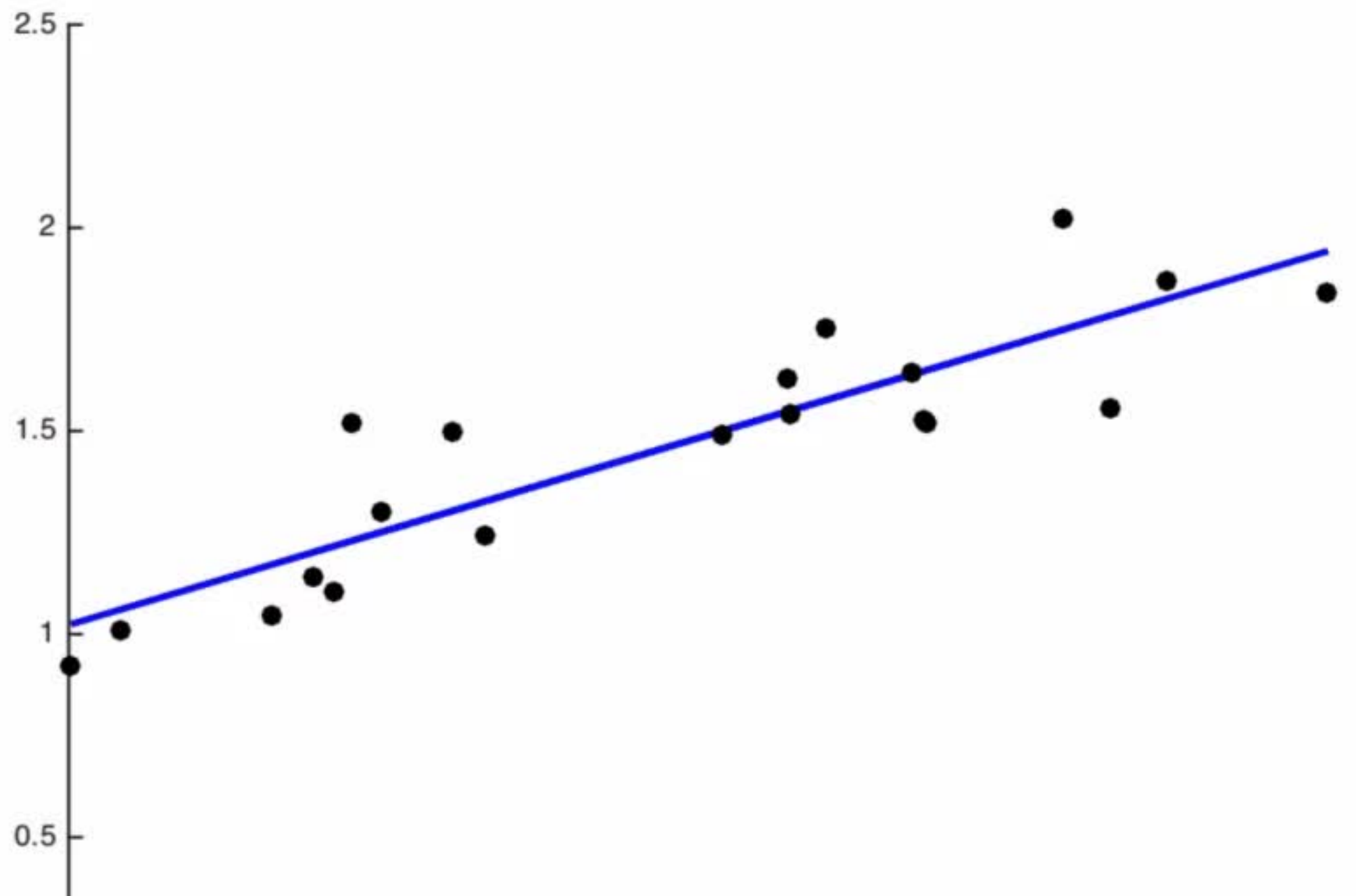


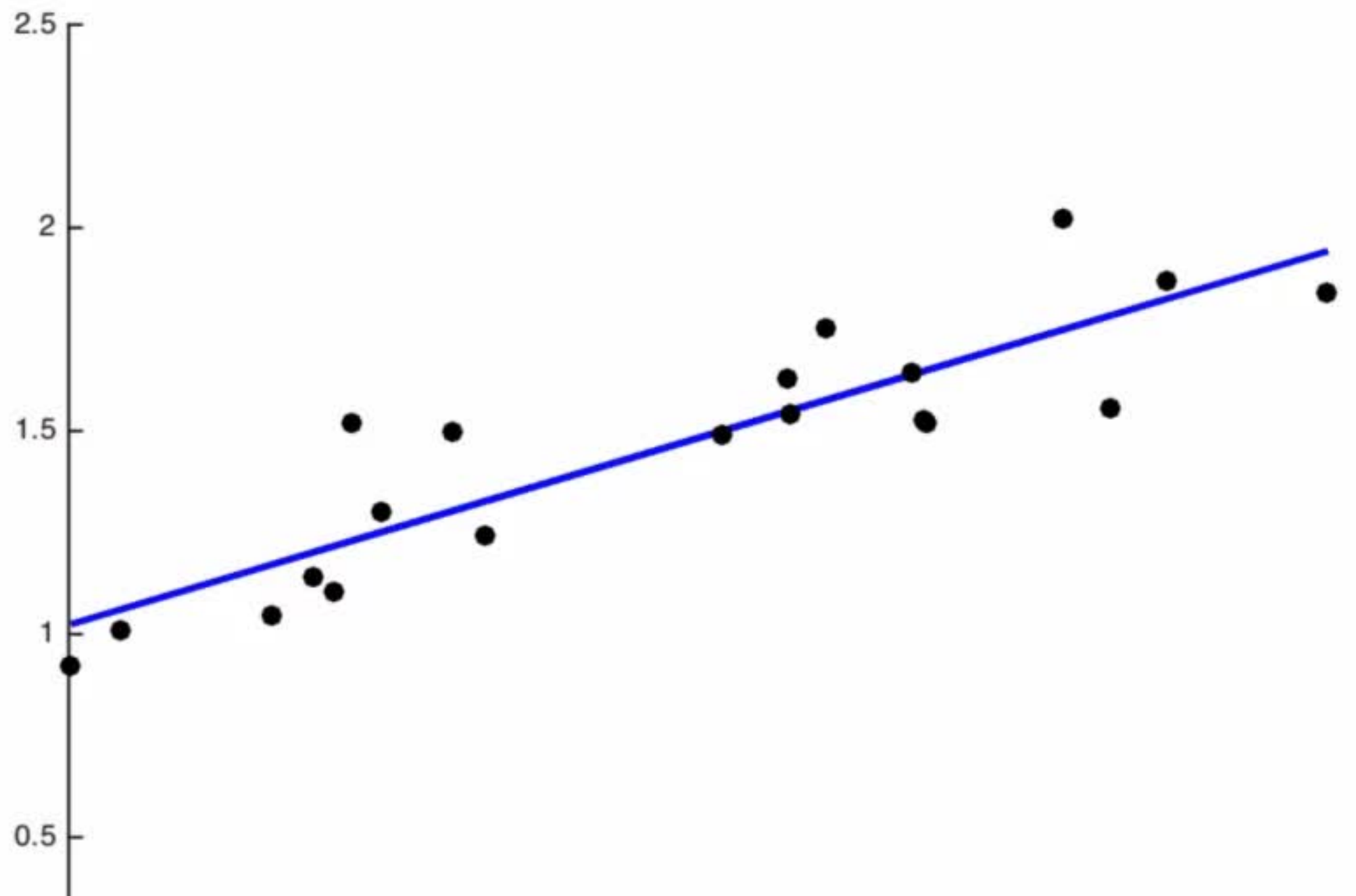
# Trimming Linear and Nonlinear Models.

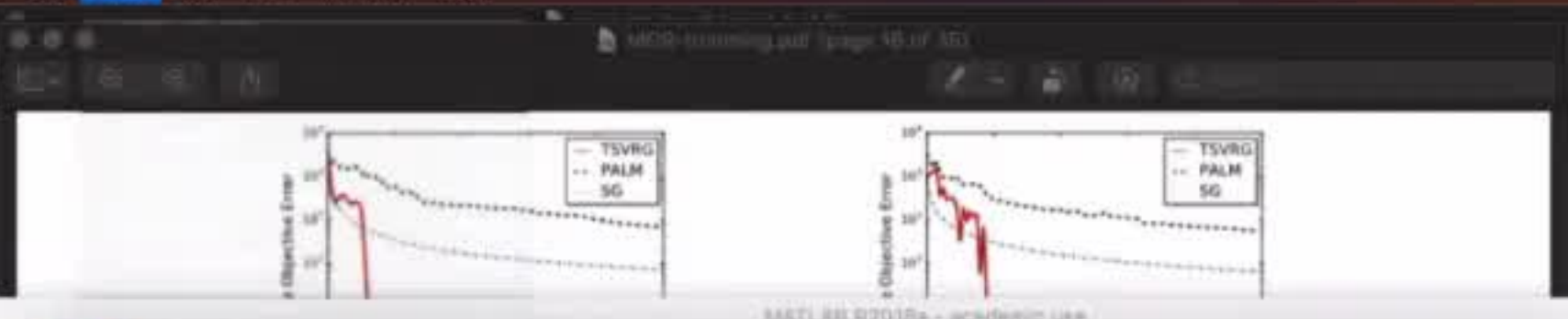
Aleksandr Aravkin<sup>1</sup>, Damek Davis<sup>2</sup>, and Peng Zheng<sup>3</sup>

$$\underset{x \in \mathbb{R}^d}{\text{minimize}} \quad \frac{1}{n} \sum_{i=1}^n \|a_i^T x - b_i\|^2$$



$$\underset{x \in \mathbb{R}^d}{\text{minimize}} \quad \frac{1}{n} \sum_{i=1}^n \|a_i^T x - b_i\|^2$$





HOME PLOTS

New Script New Live Script

Current Folder

- Name a
- G\_mat.jl
- LmNewton.jl
- LmSolversIP.jl
- nonworking\_case.jl
- README.md
- simulate\_ARp.jl
- single\_run\_functions.jl
- utils.jl

Select a file to view

SiamCSETrimming.pdf (page 2 of 32)

Cornell University, Operations Research and Information Engineering  
 University of Washington, Applied Math

6 / 18

$$\text{minimize}_{x \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^n \|a_i^T x - b_i\|^2$$

A scatter plot with x-axis from 0 to 1 and y-axis from 0 to 2.5. Black dots represent data points. A solid blue line represents a linear fit. At the bottom of the plot, there are several blue circles, likely representing the residuals or weights of the data points.

1 / 18

Search Documentation Log In

Workspace

Name	Value
ans	[1.1102e-16; 2; 4; 4]
X	5x4 double

2019

ArxivTODO.rtf 1-509Docs

AvnerNote 351Final

ben 352Ex

boosting 352Stuff

Debug1.24.rtf

micrum\_signaling acms

DC2018Reviews AmazonVisit

Clustering AnnaSasha

## Application: Trimmed Digit Classification

$$\underset{X \in \mathbb{R}^{n \times K}}{\text{minimize}} \frac{1}{n} \sum_{i=1}^n w_i (\text{LSE}(X a_i) - a_i^T X y_i) + \frac{\lambda}{2n} \|X\|_F^2 + r_1(w).$$

- $K$ -class classification with multinomial logistic regression
- Each column of  $X$  is a set of weights for recognizing the  $K$ th class.
- If example  $i$  is in class 1, then  $a_i^T X y_i = x_1^T a_i$ , so want

$$\text{LSE}(X a_i) = \log \left( \sum_{s=1}^K \exp(x_s^T a_i) \right) \approx x_1^T a_i.$$

- Formulation tries to make  $x_1^T a_i \gg x_k^T a_i$  for  $k \neq 1$ .
- **Dataset:** MNIST's  $n = 60000$  handwritten digits (0-9)
- Trim with  $h = 0.99n$ , find 1% worst outliers. **Labels?**



?



?



?



?



?



?

## Application: Trimmed Digit Classification

$$\underset{X \in \mathbb{R}^{n \times K}}{\text{minimize}} \frac{1}{n} \sum_{i=1}^n w_i (\text{LSE}(X v_i) - v_i^T X y_i) + \frac{\lambda}{2n} \|X\|_F^2 + r_1(w).$$

- $K$ -class classification with multinomial logistic regression
- **Dataset:** MNIST's  $n = 60000$  handwritten digits (0-9)
- Trim with  $h = 0.99n$ , find 1% worst outliers. **Labels?**



9



3



5



9

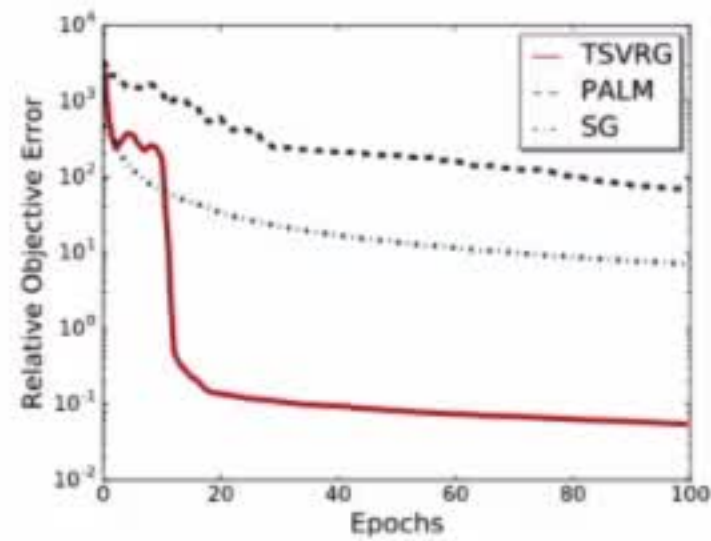


4

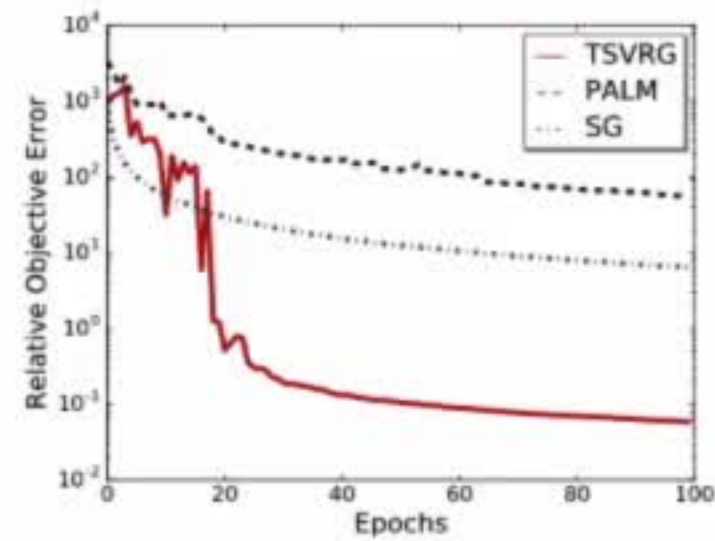


8

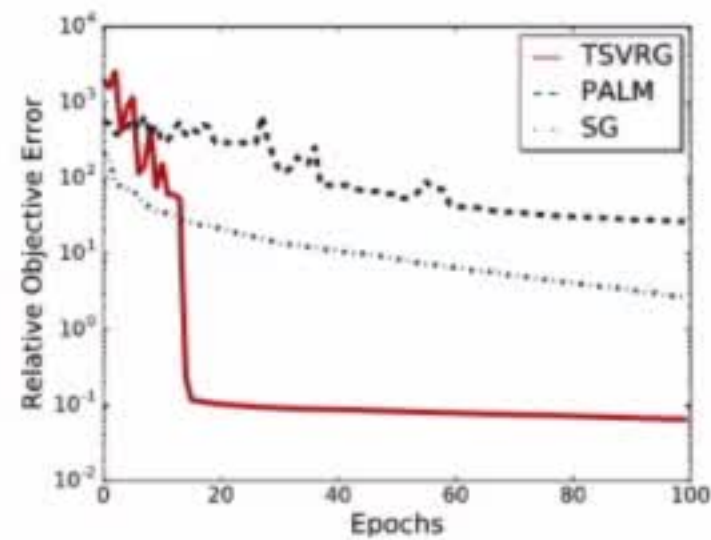
# Trimmed MNIST with PALM, SGD, and TSVRG



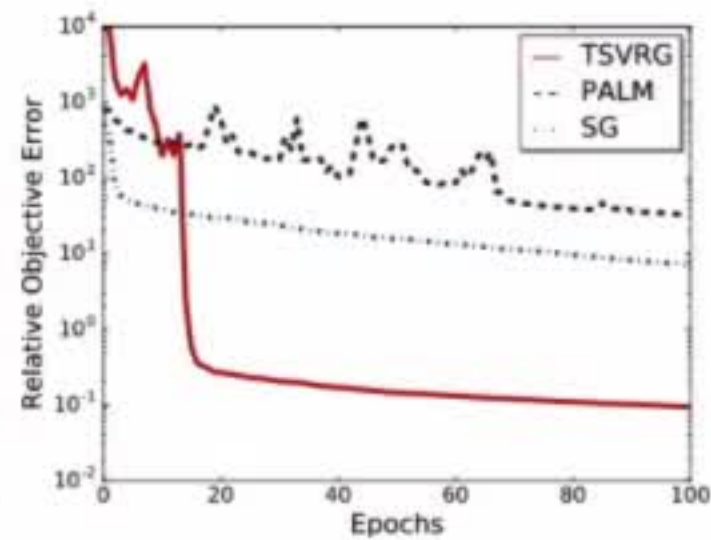
(a) 10% contamination



(b) 20% contamination



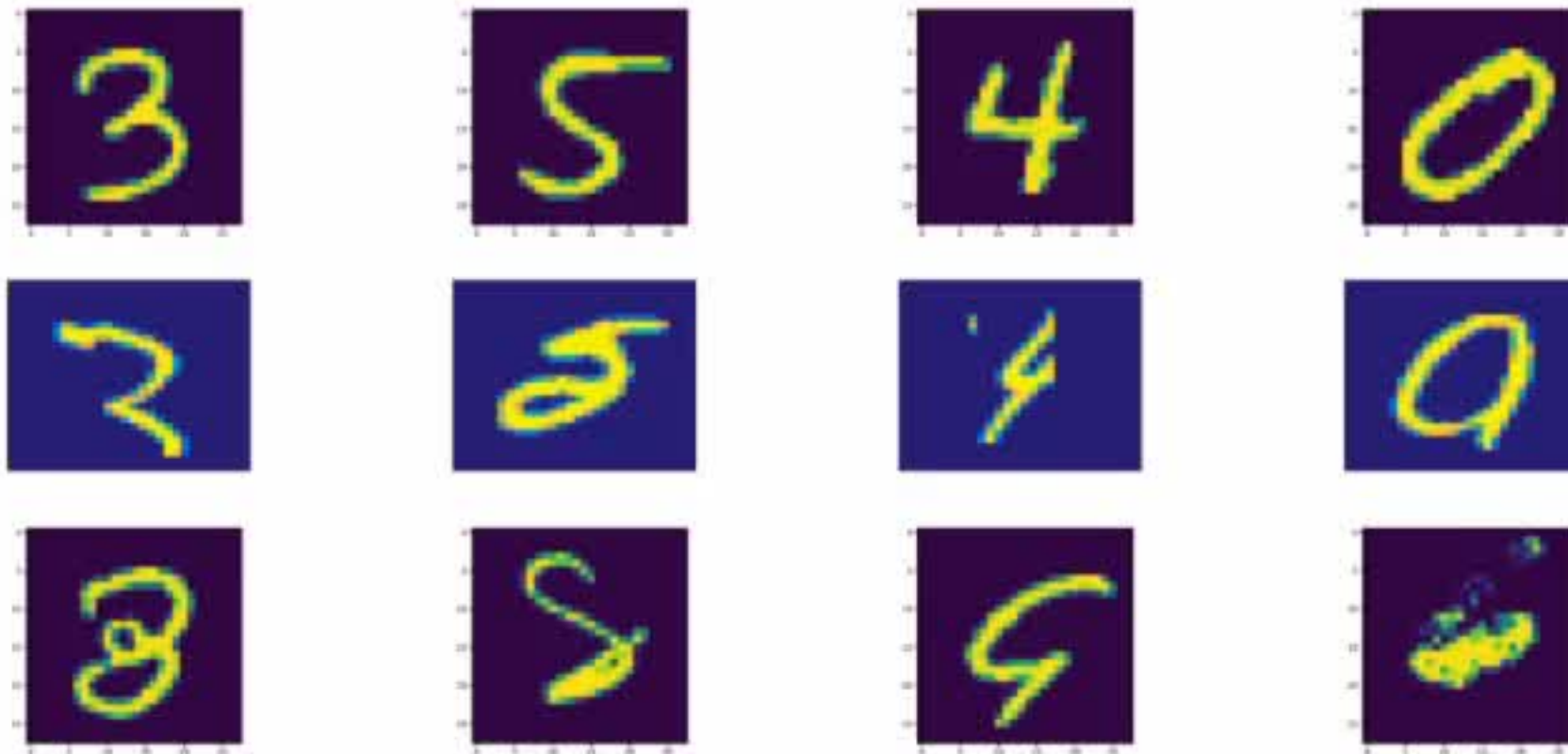
(c) 30% contamination



(d) 40% contamination

## Trimming Neural Nets

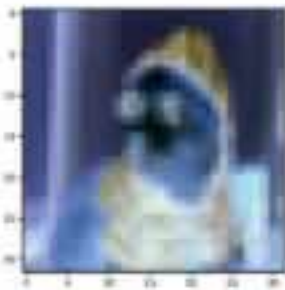
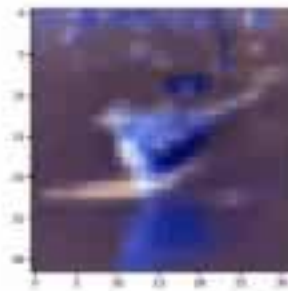
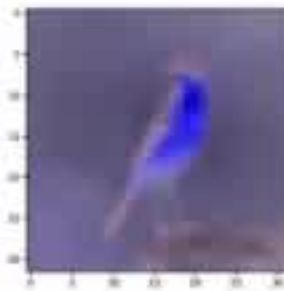
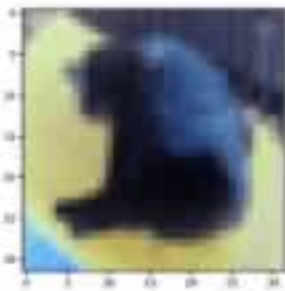
- Peng Zheng implemented TSVRG in TensorFlow for several types of nets.
- Neural nets are more expressive, so the outliers are more extreme.





## Trimming Convolutional Neural Nets

- Convolutional neural nets (CNNs) are used to classify images.
- CNNs control model complexity by connecting output neurons to small regions of input space. They add layers that 'downsample' inputs.
- Separating cats and birds: inliers vs. outliers



## Application: Trimmed Homography Estimation



(a) Image 1



(b) Image 2

- **Task:** Estimate *homography*  $H \in \mathbb{R}^{3 \times 3}$  between two images of the same scene
- Given points  $(u_1, v_1)$  in image 1 and  $(u_2, v_2)$  in image 2, we have

$$H \begin{bmatrix} u_1 & v_1 & 1 \end{bmatrix}^T = \begin{bmatrix} u_2 & v_2 & 1 \end{bmatrix}^T$$

- Given a set of correspondences  $B_1$  and  $B_2$ , get  $H$  via

$$\min_{\|H\|_F=1} \|HB_1 - B_2\|_F^2.$$

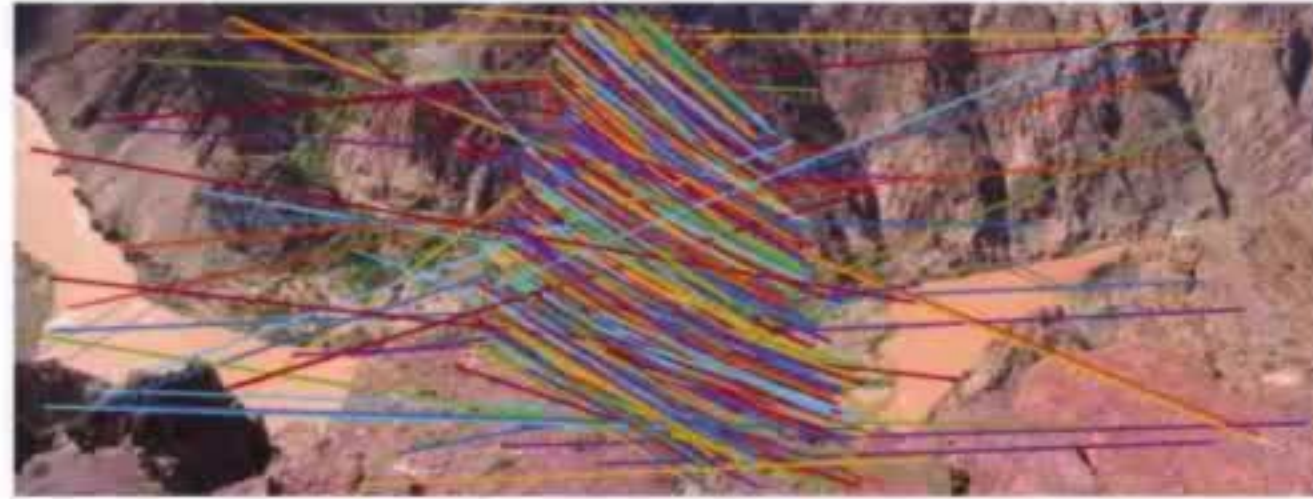
- Nonconvex problem!

## Application: Trimmed Homography Estimation

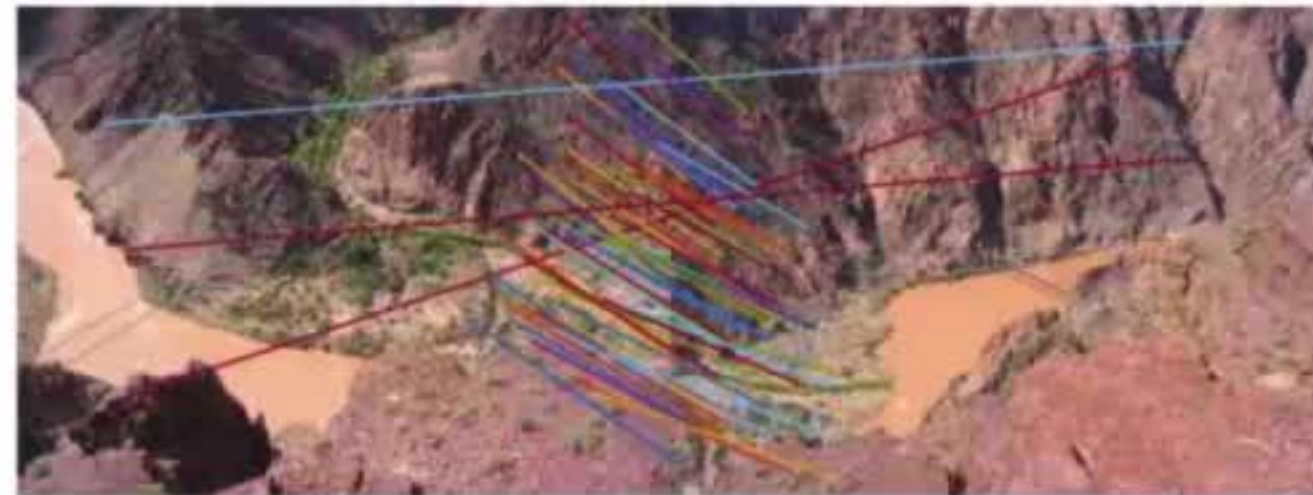
- Only need four correspondences to get a homography.
- **Problem:** one bad correspondence can completely ruin homography
- Trim instead:

$$\underset{w \in \Delta^k, \|H\|_F = 1}{\text{minimize}} \sum_{i=1}^n w_i \|Hb_{1,i} - b_{2,i}\|^2$$

## Application: Trimmed Homography Estimation



(a) Tentative SIFT matches



(b) Best 10% matches found by TSVRG (We choose  $h = .1n$ )

- Matches extremely noisy
- In practice, RANSAC (brute force) used to find four inlying points

## Application: Trimmed Homography Estimation

- Only need four correspondences to get a homography.
- **Problem:** one bad correspondence can completely ruin homography
- Trim instead:

$$\underset{w \in \Delta^h, \|H\|_F=1}{\text{minimize}} \sum_{i=1}^n w_i \|Hb_{1,i} - b_{2,i}\|^2$$