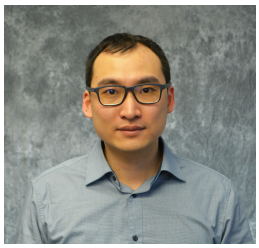
MISSOURI S&T

# Mathematics & Statistics Colloquium

**Friday, February 26, 2021, 4:15pm-5:15pm**
**Zoom Meeting ID: 941 6389 5998**
**Password (if prompted): 371814**

---




## Shih-Kang Chao

Assistant Professor

Department of Statistics

University of Missouri-Columbia


---

# Directional pruning of deep neural networks


**Abstract.** Deep neural networks are heavily overparameterized models that require significant computational power, which limits its implementation on small devices such as smart phones and robots. Pruning deep neural networks is a popular approach for reducing the requirement of computational resource and facilitating on-device deep learning. In the light of the fact that the stochastic gradient descent (SGD) often finds a flat minimum valley in the training loss, we propose a novel directional pruning which searches for a sparse minimizer in the flat region, while the re-training and the expert knowledge to decide the sparsity level are not required. To overcome the computational formidability of estimating the flat directions, we propose to use an l1 proximal gradient algorithm which can provably achieve the directional pruning with small learning rate after sufficient training. The empirical experiments show that our algorithm performs competitively in high sparse regime (92% sparsity) among many existing pruning methods on the ResNet50 with the ImageNet, while using only slightly higher wall time and memory footprint than the SGD. Using the VGG16 and the wide ResNet 28x10 on the CIFAR-10 and CIFAR-100, we show our algorithm reaches the same minima valley as the SGD, while our algorithm has a training trajectory less restricted in the low dimensional subspace associated with the leading eigenvectors of the Hessian than the SGD.


**Biographical Sketch.** Dr. Shih-Kang Chao is an assistant professor in statistics at the University of Missouri-Columbia. He obtained doctoral degree from the Humboldt University of Berlin, Germany. He was a postdoctoral research fellow and a visiting assistant professor in the statistics department at Purdue University. His research area includes computational resource-aware statistical inference, e.g. distributed learning and stochastic optimization, with applications to robust statistics, structure-aware stochastic optimization and deep learning.