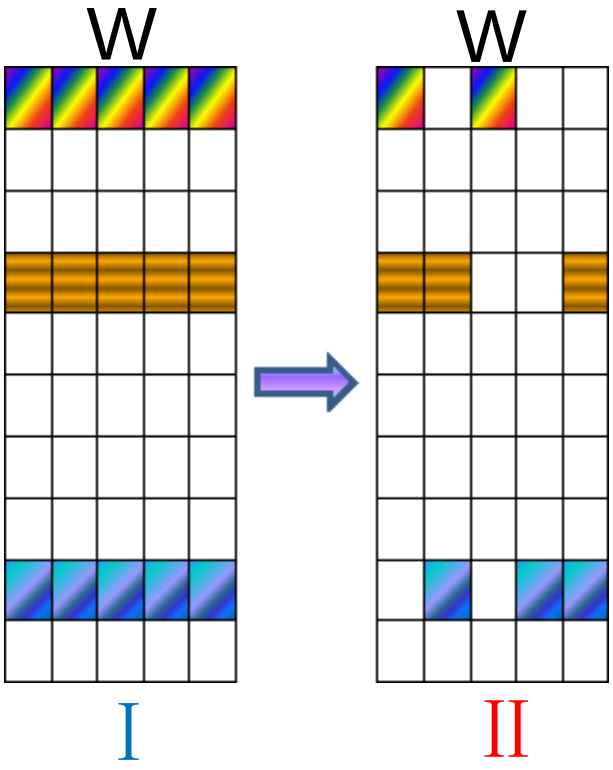
Multi-Stage Multi-Task Feature Learning

Pinghua Gong[1], ***Jieping Ye***[2], ***Changshui Zhang***[1]

[1]*Tsinghua University, Beijing, China* [2]*Arizona State University, Tempe, AZ*

gph08@mails.tsinghua.edu.cn, *jieping.ye@asu.edu*, *zcs@mail.tsinghua.edu.cn*



$$\min_{W \in \mathbb{R}^{d \times m}} \{ l(W) + r(W) \}$$

I Joint feature selection

II Shared features + Task specific Features

Proposed Non-convex Regularizer

$$r(W) = \lambda \|W\|_{1,q} \ (q = 2, \infty)$$

$$r(W) = \lambda \|W\|_{1,1}$$

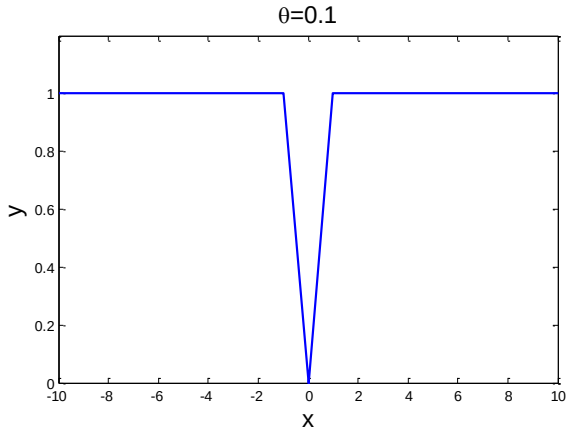
$$r(W) = \lambda_1 \|W\|_{1,1} + \lambda_2 \|W\|_{1,q} \ (q = 2, \infty)$$

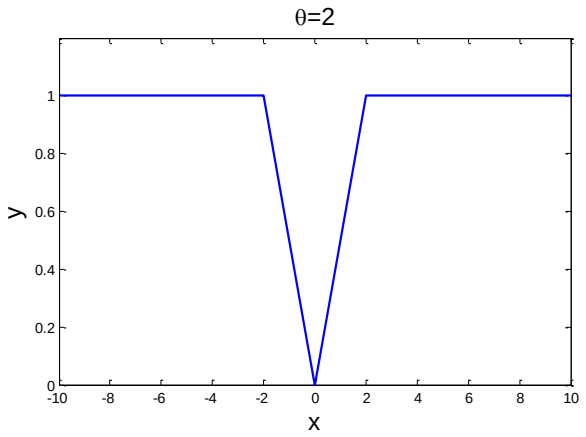
Convex: Over-penalization

$$r(W) = \lambda \sum_{j=1}^{d} \min\left(\|\mathbf{w}^j\|_1, \theta\right)$$

Proposed: Non-convex

$$y = \frac{1}{\theta} \min(|x|, \theta)$$





Optimization Algorithm

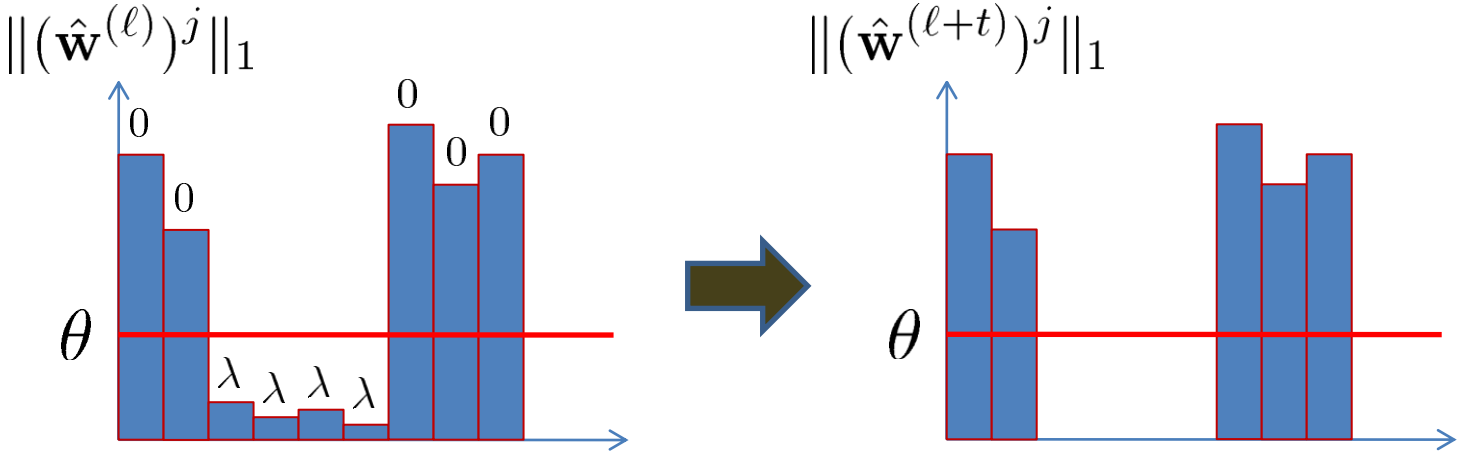
MSMTFL: Multi-Stage Multi-Task Feature Learning

1. Initialize $\lambda_j^{(0)} = \lambda$

repeat

2. $\hat{W}^{(\ell)} = \arg\min_{W \in \mathbb{R}^{d \times m}} \left\{ l(W) + \sum_{j=1}^{d} \lambda_j^{(\ell-1)} \|\mathbf{w}^j\|_1 \right\}$ reweighted Lasso

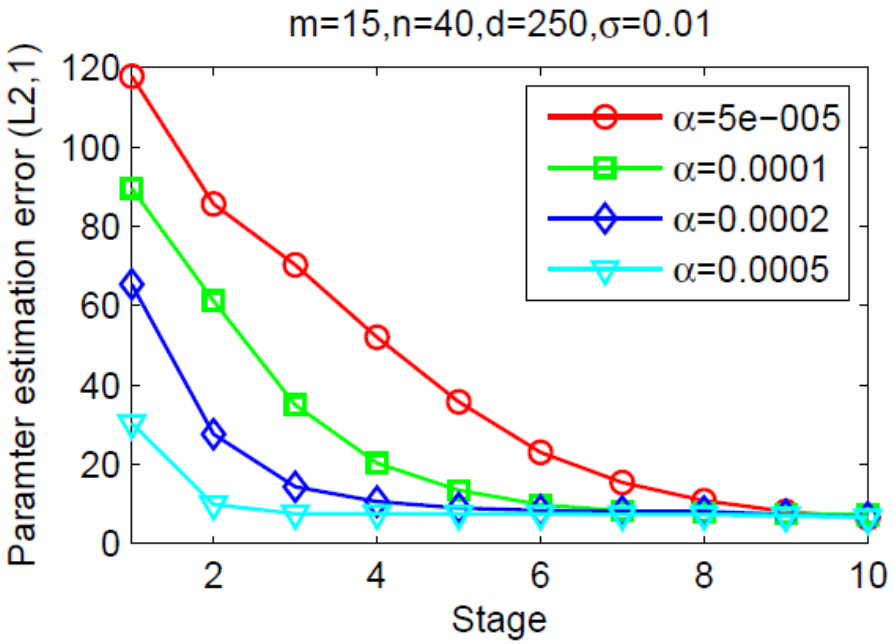
3. $\lambda_j^{(\ell)} = \lambda I(\|(\hat{\mathbf{w}}^{(\ell)})^j\|_1 < \theta) \ (j = 1, \cdots, d)$ penalize small rows



Parameter Estimation Error Bound

$$\|\hat{W}^{(\ell)} - \bar{W}\|_{2,1} = 0.8^{\ell/2} O\left(m\sqrt{\bar{r}\ln(dm/\eta)/n}\right) + O\left(m\sqrt{\bar{r}/n + \ln(1/\eta)/n}\right)$$

Exponential shrinkage & stage-wise Improvement



Lasso: $\|\hat{W}^{Lasso} - \bar{W}\|_{2,1} = O\left(m\sqrt{\bar{r}\ln(dm/\eta)/n}\right)$

MSMTFL: $\|\hat{W}^{(\ell)} - \bar{W}\|_{2,1} = O\left(m\sqrt{\bar{r}/n + \ln(1/\eta)/n}\right)$