Lin-bo Qiao, Bo-feng Zhang, Jin-shu Su, Xi-cheng Lu , 2017. A systematic review of structured sparse learning. *Frontiers of Information Technology & Electronic Engineering*, **18**(4): 445-463.
http://dx.doi.org/10.1631/FITEE.1601489

A systematic review of structured sparse learning

Key words: Sparse learning; Structured sparse learning; Structured regularization; Algorithms; Applications

Contact: Linbo Qiao
E-mail: qiao.linbo@nudt.edu.cn
ORCID: http://orcid.org/0000-0002-8285-2738

Introduction

- The increasing data becomes a great challenge for contemporary statistical learning algorithms.
- Sparse learning has become a popular tool with the development of theoretical frameworks and various efficient algorithms.
- Structured sparse learning encodes the structured information, provides a great advantage beyond the traditional sparse learning algorithms to pursue the structured models.

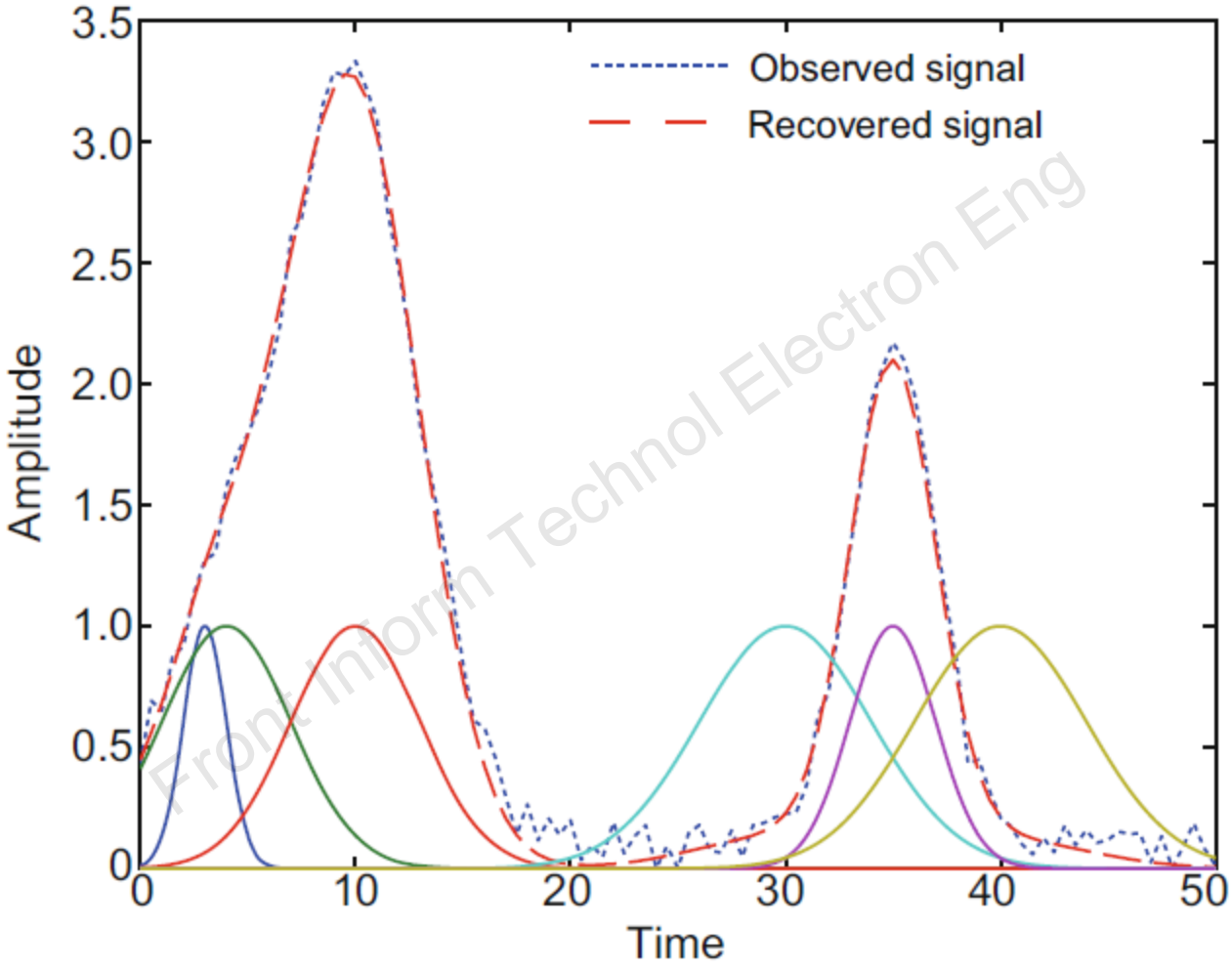


Illustration of Lasso and its extensions

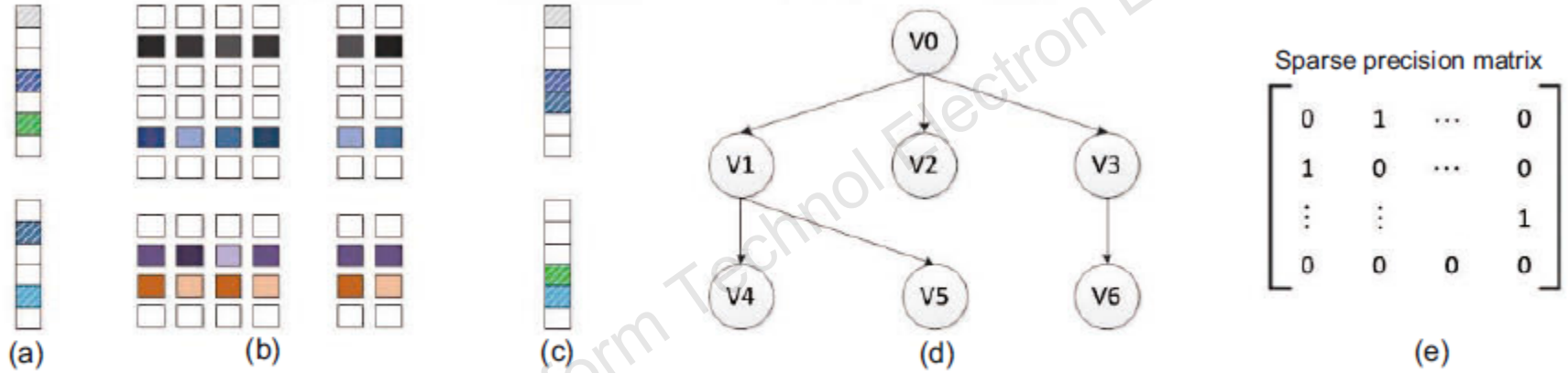


Fig. 2 Illustration of sparsity and its extensions: (a) standard sparsity; (b) grouped sparsity; (c) fused sparsity; (d) hierarchical sparsity; (e) graphical sparsity (References to color refer to the online version of this figure)

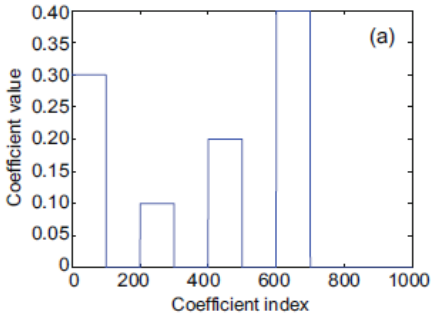
Table 2 An overview of formulations, processing algorithms and software packages for Lasso and its structured extensions

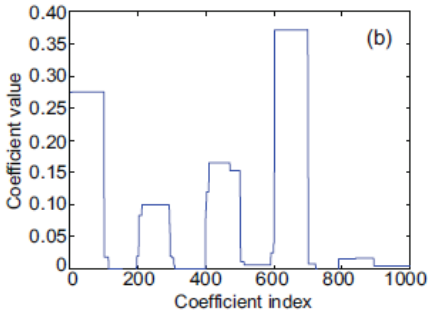
Methods	Losses, $l(x)$	Regularizations, $R(x)$	Optimization algorithms	Software packages
Lasso (Tibshirani, 1996)	$\|\|y - Ax\|\|_F^2$	$\lambda\|x\|_1$	Generic QP methods after reformulation (Bach *et al.*, 2012a), alternating direction methods (Boyd *et al.*, 2011), proximal methods (Parikh and Boyd, 2014), block coordinate descent methods (Tseng and Yun, 2009; Wen *et al.*, 2012; Peng *et al.*, 2016), iteratively reweighted methods (Chartrand and Yin, 2008; Lai *et al.*, 2013), working-set and homotopy methods (Bach *et al.*, 2012a)	CVX (Grant and Boyd, 2013), SDPT3 (Toh *et al.*, 2006), YALL1 (Zhang *et al.*, 2011), SPGL1 (van den Berg and Friedlander, 2007), SLEP (Liu *et al.*, 2009b), SPAMs (Mairal *et al.*, 2011), SparseLab (Donoho *et al.*, 2007)
Grouped Lasso (Ming and Yanping, 2006)	$\|\|y - \beta_0 - \sum_{i=1}^{g} X_i^T \beta_i\|\|_F^2$	$\lambda \sum_{i=1}^{g} \sqrt{p_i}\|\|\beta_i\|\|_2$	Generic SOCP methods after reformulation (Bach, 2008b), alternating direction methods (Boyd *et al.*, 2011), proximal methods (Parikh and Boyd, 2014), block coordinate descent methods (Tseng and Yun, 2009; Wen *et al.*, 2012; Peng *et al.*, 2016), iteratively reweighted methods (Chartrand and Yin, 2008; Lai *et al.*, 2013), working-set and homotopy methods (Bach *et al.*, 2012a)	CVX (Grant and Boyd, 2013), SDPT3 (Toh *et al.*, 2006), YALL1 (Zhang *et al.*, 2011), SLEP (Liu *et al.*, 2009b), SPAMs (Mairal *et al.*, 2011)
Fused Lasso (Tibshirani *et al.*, 2005)	$\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2$	$\lambda_1 \sum_{j=1}^{p}\|\|\beta_j\|\|_1 + \lambda_2 \sum_{j=1}^{p-1}\|\|\beta_{j+1} - \beta_j\|\|_1$	Generic SDP methods after reformulation (Selesnick and Bayram, 2014), alternating direction methods (Boyd *et al.*, 2011), proximal methods (Parikh and Boyd, 2014), block coordinate descent methods (Tseng and Yun, 2009; Wen *et al.*, 2012; Peng *et al.*, 2016), iteratively reweighted methods (Chartrand and Yin, 2008; Lai *et al.*, 2013) working-set and homotopy methods (Bach *et al.*, 2012a)	CVX (Grant and Boyd, 2013), SDPT3 (Toh *et al.*, 2006), SLEP (Liu *et al.*, 2009b), SPAMs (Mairal *et al.*, 2011)
Hierarchical Lasso (Zhao *et al.*, 2009)	$\|\|y - \beta_0 - \sum_{i=1}^{g} X_i^T \beta_i\|\|_F^2$	$\lambda \sum_{g\in\mathcal{G}} w_g\|\|\alpha_{\|g}\|\|$	Generic SDP methods after reformulation (Francis, 2008a), alternating direction methods (Boyd *et al.*, 2011), proximal methods (Parikh and Boyd, 2014), block coordinate descent methods (Tseng and Yun, 2009; Wen *et al.*, 2012; Peng *et al.*, 2016), iteratively reweighted methods (Chartrand and Yin, 2008; Lai *et al.*, 2013), working-set and homotopy methods (Bach *et al.*, 2012a)	CVX (Grant and Boyd, 2013), SDPT3 (Toh *et al.*, 2006), SLEP (Liu *et al.*, 2009b), SPAMs (Mairal *et al.*, 2011)
Graphical Lasso (Meinshausen and Bühlmann, 2006)	$\log\det\Theta - \mathrm{trace}(S\Theta)$	$\lambda\|\|\Theta\|\|_1$	Generic SDP methods after reformulation (Francis, 2008a), alternating direction Methods (Boyd *et al.*, 2011), proximal methods (Parikh and Boyd, 2014), block coordinate descent methods (Tseng and Yun, 2009; Wen *et al.*, 2012; Peng *et al.*, 2016), iteratively reweighted methods (Chartrand and Yin, 2008; Lai *et al.*, 2013), working-set and homotopy methods (Bach *et al.*, 2012a)	CVX (Grant and Boyd, 2013), SDPT3 (Toh *et al.*, 2006), SLEP (Liu *et al.*, 2009b), SPAMs (Mairal *et al.*, 2011)

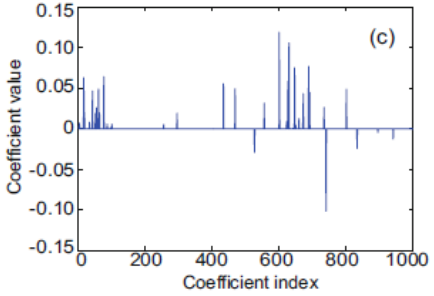
An overview of optimization methods' computational complexity

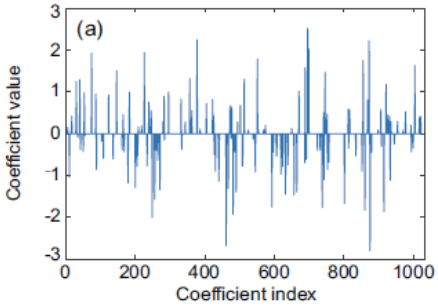
Algorithm	Formulation	Convex	Strongly Convex
Subgradient descent	$\min_{x \in \mathbb{R}^n} f(x)$	$O(1/\varepsilon^2)$ (Nesterov, 2004)	$O(1/\varepsilon)$ (Lacoste-Julien *et al.*, 2012)
Mirror descent	$\min_{x \in C} f(x)$	$O(1/\varepsilon^2)$ (Beck and Teboulle, 2003)	$O(1/\varepsilon)$ (Nemirovski, 2004)
Dual averaging	$\min_{x \in C} f(x)$	$O(1/\varepsilon^2)$ (Nesterov, 2009)	$O(\log(1/\varepsilon))$ (Suzuki, 2013)
Gradient descent	$\min_{x \in \mathbb{R}^n} f(x)$	$O(1/\varepsilon)$ (Nesterov, 2004)	$O(\log(1/\varepsilon))$ (Hazan *et al.*, 2007)
Accelerated gradient descent	$\min_{x \in \mathbb{R}^n} f(x)$	$O(1/\sqrt{\varepsilon})$ (Su *et al.*, 2014)	$O(\log(1/\varepsilon))$ (Tseng, 2008)
Proximal gradient descent	$\min_{x \in C} f(x) + g(x)$	$O(1/\varepsilon)$ (Combettes and Pesquet, 2011)	$O(\log(1/\varepsilon))$ (Suzuki, 2013)
Accelerated proximal gradient descent	$\min_{x \in C} f(x) + g(x)$	$O(1/\sqrt{\varepsilon})$ (Mairal, 2013)	$O(\log(1/\varepsilon))$ (Lin *et al.*, 2015)
Frank-Wolfe algorithm/conditional gradient algorithm	$\min_{x \in C} f(x)$	$O(1/\varepsilon)$ (Jaggi, 2013)	$O(1/\sqrt{\varepsilon})$ (Garber and Hazan, 2015)

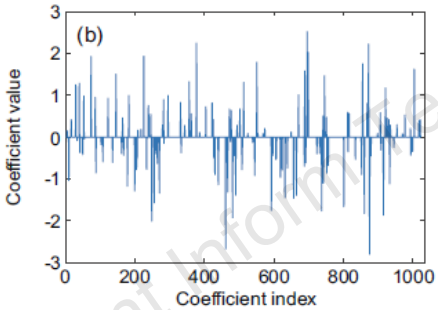
Experiments

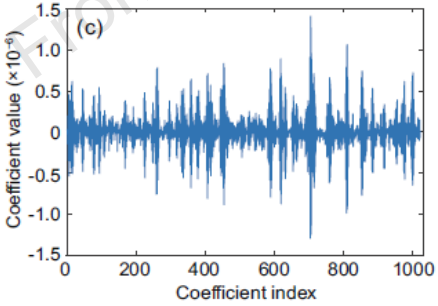


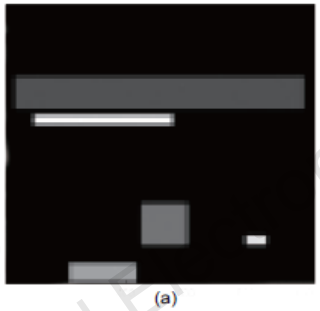






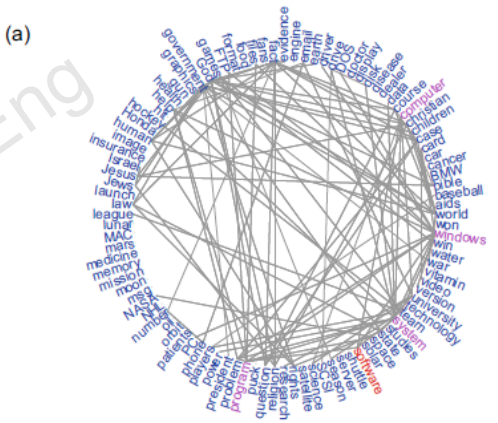


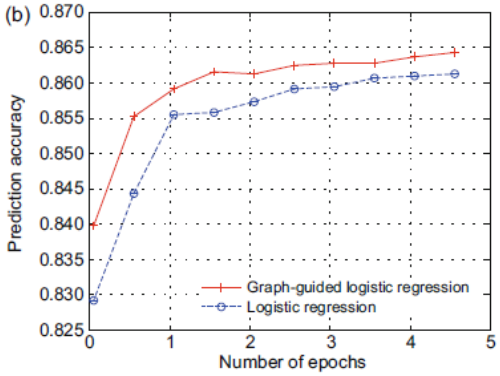



(a)


(b)


(c)





Conclusions

- We reviewed the development of the formulations, algorithms, and applications of the latest structured sparse learning methods, including grouped structured sparsity, fused structured sparsity, hierarchical structured sparsity, and graphical structured sparsity.
- Experiments are conducted to demonstrate the advantage of structured sparse learning algorithms beyond standard sparse learning methods.