Support recovery guarantees for group Lasso estimator

Mojtaba Kadkhodaie Elyaderani, Swayambhoo Jain, and Jarvis Haupt
Dept. of Electrical and Computer Engineering
University of Minnesota - Twin Cities

Jeff Druce and Stefano Gonella
Dept. of Civil, Environmental, and Geo- Engineering
University of Minnesota - Twin Cities

***Abstract*—We consider the problem of estimating a high dimensional signal from noisy low-dimensional linear measurements, where the desired unknown signal exhibits a *group-sparse* structure. Assuming the non-zero groups of the group-sparse signal possess enough strength and are generated according to certain statistical assumptions, we provide conditions to guarantee that the signal support can be exactly recovered via solving the group Lasso problem.**

I. Introduction

Let the measurement vector $\boldsymbol{y} \in \mathbb{R}^n$ be generated according to

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{w}, \tag{1}$$

where $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ is a (given) dictionary, $\boldsymbol{\beta}^* \in \mathbb{R}^p$ is an unknown signal, and $\boldsymbol{w} \in \mathbb{R}^n$ represents noise and/or model inaccuracies. This work studies the high dimensional scenario ($p \gg n$) in which the signal $\boldsymbol{\beta}^*$ is group-sparse. Specifically, assuming there exists a predefined partition of $\boldsymbol{\beta}^*$ as $(\boldsymbol{\beta}^*)^T = \left[(\boldsymbol{\beta}^*_{\mathcal{I}_1})^T\,(\boldsymbol{\beta}^*_{\mathcal{I}_2})^T \cdots (\boldsymbol{\beta}^*_{\mathcal{I}_G})^T\right]$, where $\boldsymbol{\beta}^*_{\mathcal{I}_g} \in \mathbb{R}^{d_g}$ for $g \in [G] := \{1, \cdots, G\}$ denotes the d_g entries of $\boldsymbol{\beta}^*$ corresponding to the index set $\mathcal{I}_g \subset [p]$, then a small fraction of the groups, say s out of the entire G ones, are non-zero. Given this assumption, a well-studied estimator is the group Lasso estimator

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{2}||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||_2^2 + \sum_{g=1}^{G} \lambda_g ||\boldsymbol{\beta}_{\mathcal{I}_g}||_2, \tag{2}$$

where $\lambda_g > 0$ is a regularization constant. Numerous studies provide statistical guarantees for problem (2) when the measurements are generated as in (1). Some works assume $\boldsymbol{X}$ is generated according to a random (e.g. Gaussian) distribution [1]–[3], which narrows down the applicability of the recovery result. In terms of the requirements for successful recovery, various conditions are proposed so far: the group RIP condition of [4] and the restricted group eigenvalue condition of [5], [6] are among the most popular ones. Since verifying such conditions for structured measurement matrices can be computationally prohibitive, we do not base our analysis on them and instead use the concept of block coherence (see Definition I.1) which is computable in polynomial time [7]. The recent study [8] analyzes group Lasso estimation method using similar conditions as we consider here. However, it focuses on regression error instead of support recovery, which constitutes the primary focus of this study.

Definition I.1. *For the dictionary $\boldsymbol{X} = [\boldsymbol{X}_{\mathcal{I}_1}\boldsymbol{X}_{\mathcal{I}_2}\cdots\boldsymbol{X}_{\mathcal{I}_G}]$, with $\boldsymbol{X}_{\mathcal{I}_g} \in \mathbb{R}^{n\times d_g}$, the block coherence constant $\mu_B(\boldsymbol{X})$ is defined as*

$$\mu_B(\boldsymbol{X}) := \max_{1\le g\ne g'\le G} \|\boldsymbol{X}^T_{\mathcal{I}_g}\boldsymbol{X}_{\mathcal{I}_{g'}}\|_{2\to2}, \tag{3}$$

where $\|\cdot\|_{2\to2}$ denotes the matrix spectral norm and moreover the intra-block coherence parameter $\mu_I(\boldsymbol{X})$ is

$$\mu_I(\boldsymbol{X}) := \max_{g\in[G]} \|\boldsymbol{X}^T_{\mathcal{I}_g}\boldsymbol{X}_{\mathcal{I}_g} - \boldsymbol{I}_{d_g\times d_g}\|_{2\to2}. \tag{4}$$

This work was supported by the DTI grant, NSF Award CCF-1217751, and the DARPA Young Faculty Award N66001-14-1-4047.
Author emails: {kadkh004, jainx174, jdhaupt, druce001, sgonella}@umn.edu.

II. Main Result

Similar to [8], we begin by making some statistical assumptions.

M_1) The group-level support of $\boldsymbol{\beta}^*$, denoted by $\mathcal{G}^* \subseteq [G]$, comprises $s := |\mathcal{G}^*|$ non-zero blocks, whose indices are selected uniformly at random from all subsets of $[G]$ that are of size s.

M_2) The non-zero entries of $\boldsymbol{\beta}^*$ are equally likely to be positive or negative: $\mathbb{E}\,\mathrm{sign}(\beta^*_j) = 0$ for $j \in [p]$.

M_3) The non-zero blocks of $\boldsymbol{\beta}^*$ have statistically independent "directions."

Given these assumptions and $d_{\min} := \min_{g\in[G]} d_g$, $d_{\max} := \max_{g\in[G]} d_g$, $d^*_{\mathcal{G}} = \sum_{g\in\mathcal{G}^*} d_g$, our main contribution is [9], [10]:

Theorem II.1. *For the model in* (1) *with $\boldsymbol{w} \sim \mathcal{N}(0, \sigma^2\boldsymbol{I}_{n\times n})$ if*

1) $\mu_I(\boldsymbol{X}) \le c_0$ *and* $\mu_B(\boldsymbol{X}) \le \sqrt{\frac{d_{\min}}{d^2_{\max}}}\,\frac{c_1}{\log p}$,
2) $|\mathcal{G}^*| \le \min\{\frac{c_2\,G}{\|\boldsymbol{X}\|^2_{2\to2}\log p}, \frac{d_{\min}}{d^2_{\max}}\,\frac{c_3\,\mu_B^{-2}(\boldsymbol{X})}{\log p}\}$,
3) $\|\boldsymbol{\beta}^*_{\mathcal{I}_g}\|_2 \ge 10\sigma(1+\epsilon)(\sqrt{d^*_{\mathcal{G}}} + \sqrt{d_g})\,\max\{1, \sqrt{\frac{s}{d_{\max}\,\log p}}\}$,

all hold for some non-negative constants c_0, $c_1 \le 0.004$, $c_2 \le \frac{1}{14}(\frac{1}{4} - 3\,c_0 - 48\,c_1)$, $c_3 = \min\{c_2, 0.0004\}$, and some

$$\epsilon \ge \sqrt{\frac{(1+\mu_I(\boldsymbol{X}))\,\log(p\,G)}{d_{\min}}},$$

then the solution $\widehat{\boldsymbol{\beta}}$ of (2)*, with $\lambda_g = 4\sigma(1+\epsilon)\sqrt{d_g}$ for every $g \in [G]$, is unique and has the same group-level support as $\boldsymbol{\beta}^*$ and*

$$\|\widehat{\boldsymbol{\beta}}_{\mathcal{I}_g} - \boldsymbol{\beta}^*_{\mathcal{I}_g}\|_2 \le 5\sigma(1+\epsilon)(\sqrt{d_g} + \sqrt{d^*_{\mathcal{G}}}),\ \forall g \in \mathcal{G}(\boldsymbol{\beta}^*),$$

with probability at least $1 - 12\,p^{-2\log 2}$.

III. Numerical Experiments

Assume the dictionary $\boldsymbol{X}$ is the concatenation of two orthonormal bases, i.e. $\boldsymbol{X} := \left[\boldsymbol{X}_{(1)}\middle|\boldsymbol{X}_{(2)}\right] \in \mathbb{R}^{n\times 2n}$, where $\boldsymbol{X}_{(1)} \in \mathbb{R}^{n\times n}$ is the discrete cosine transform (DCT) matrix and $\boldsymbol{X}_{(2)} \in \mathbb{R}^{n\times n}$ is the identity matrix. The authors leveraged this widely-studied dictionary in the context of structural anomaly detection using propagating wave-field measurements [11], where $\boldsymbol{X}_{(2)}$ was column-wise partitioned into groups of size $d_g = d$ and $\boldsymbol{X}_{(1)}$ was divided into singleton groups of size $d_g = 1$. For such $\boldsymbol{X}$ with the specified partition, it can be shown that $\mu_B(\boldsymbol{X}) \le \sqrt{4d/n}$, $\mu_I(\boldsymbol{X}) = 0$, $\|\boldsymbol{X}\|^2_{2\to2} = 2$, and $G = n\,(1 + 1/d)$. Substituting these in the expressions of the above theorem implies that if $\frac{\sqrt{n}}{\log(2n)} \ge \frac{2\sqrt{d^3}}{c_1}$, $s \le \frac{c_2 n}{d^3\log(2n)}$, and

$$\|\boldsymbol{\beta}^*_{\mathcal{I}_g}\|_F \ge 10\sigma(1+\epsilon)\left(\sqrt{d} + \sqrt{s_1 + s_2\,d}\right)\max\left\{1, \sqrt{\frac{s}{d\,\log(2n)}}\right\}$$

for all $g \in \mathcal{G}^*$, hold simultaneously for $\epsilon \ge \sqrt{2\log n}$, then exact recovery is possible. Fig. 1 shows the result of simulations designed to investigate the above relationship between the number of non-zero groups s in $\boldsymbol{\beta}^*$ and their magnitudes $\{\|\boldsymbol{\beta}^*_{\mathcal{I}_g}\|_F\}_{g\in\mathcal{G}^*}$.

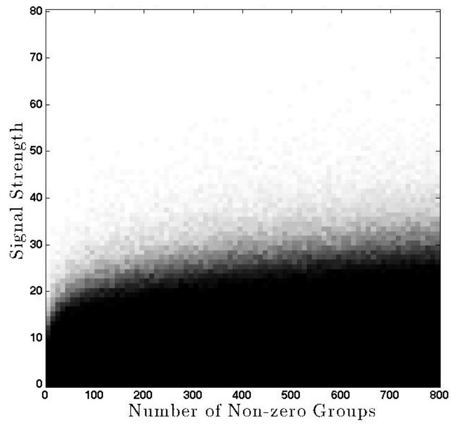

Fig. 1. The phase transition depicts the signal strength, controlled by the positive scalar α in $\boldsymbol{y} = \alpha \boldsymbol{X} \boldsymbol{\beta}^* + \boldsymbol{w}$, where $\boldsymbol{\beta}^*$ is generated according to M1 to M3 (with its non-zero entries drawn from standard Gaussian distribution) and $\boldsymbol{w} \sim \mathcal{N}(0, \boldsymbol{I}_n)$, versus the number of non-zero groups s. Here we have set $n = 10^4$, $d = 4$, $\lambda_1 = 10$ and $\lambda_2 = 10\sqrt{d}$, where the regularization constants correspond to the groups over the DCT and identity components, respectively. The plot shown is obtained after averaging over 100 trials.

REFERENCES

[1] Francis R Bach, "Consistency of the group lasso and multiple kernel learning," *The Journal of Machine Learning Research*, vol. 9, pp. 1179–1225, 2008.

[2] Guillaume Obozinski, Martin J Wainwright, and Michael I Jordan, "Support union recovery in high-dimensional multivariate regression," *The Annals of Statistics*, pp. 1–47, 2011.

[3] Nikhil S Rao, Ben Recht, and Robert D Nowak, "Universal measurement bounds for structured sparse signal recovery," in *International Conference on Artificial Intelligence and Statistics*, 2012, pp. 942–950.

[4] Junzhou Huang and Tong Zhang, "The benefit of group sparsity," *The Annals of Statistics*, vol. 38, no. 4, pp. 1978–2004, 2010.

[5] Karim Lounici, Massimiliano Pontil, Sara Van De Geer, and Alexandre B Tsybakov, "Oracle inequalities and optimal inference under group sparsity," *The Annals of Statistics*, pp. 2164–2204, 2011.

[6] Sahand Negahban, Pradeep Ravikumar, Martin J Wainwright, and Bin Yu, "A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers," *Manuscript, University of California, Berkeley, Dept. of Statistics and EECS*, 2011.

[7] Yonina C Eldar, Patrick Kuppinger, and Helmut Bölcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, 2010.

[8] Waheed U Bajwa, Marco F Duarte, and Robert Calderbank, "Conditioning of random block subdictionaries with applications to block-sparse recovery and regression," *IEEE Transactions on Information Theory*, vol. 61, no. 7, pp. 4060–4079, 2015.

[9] Mojtaba Kadkhodaie Elyaderani, Swayambhoo Jain, Jeff Druce, Stefano Gonella, and Jarvis D. Haupt, "Group-level support recovery guarantees for group lasso estimator, with applications to structural health monitoring," *In Preparation*.

[10] Mojtaba Kadkhodaie Elyaderani, Swayambhoo Jain, Jeff Druce, Stefano Gonella, and Jarvis D. Haupt, "Group-level support recovery guarantees for group lasso estimator," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2017.

[11] Mojtaba Kadkhodaie, Swayambhoo Jain, Jarvis Haupt, Jeff Druce, and Stefano Gonella, "Locating rare and weak material anomalies by convex demixing of propagating wavefields," in *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015 IEEE 6th International Workshop on*. IEEE, 2015, pp. 373–376.