# AMP for SLOPE

*Bartłomiej Polaczyk*

*02 July 2020*

**Abstract**

The aim of this project is to investigate the the approximate message passing algorithm for SLOPE regularization problem based on (Bu et al. 2019) and compare it with classical convex optimization methods. Some numerical experiments regarding the cases that do not fit into the theoretical framework of (Bu et al. 2019) are also performed and analyzed.

## Theoretical bacground

### Introduction

We are interested in solving the standard linear inverse problem

$$y = Ax + w, \tag{1}$$

where $y \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times p}$ are known parameters of the model, $w \in \mathbb{R}^n$ is a random noise vector and $x \in \mathbb{R}^p$ is an unknown vector of paramteres we wish to estimate. We assume $p \gg n$, i.e. the number of features is much greater than the size of the sample data and whence there might be many potential solutions to the problem (1).

To resolve this issue and prevent overfitting, we introduce a penalty function $\phi$ which faforizes sparse solutions of (1), i.e. now we are looking among the minimizers of the following form

$$\widehat{x} = \operatorname*{argmin}_x \{ \frac{1}{2}\|Ax - y\|_2^2 + \phi(x) \}. \tag{2}$$

The usual choices of $\phi$ are scaled $l^2$ penalty (Tikhonov regularization) and $l^1$ penalty (LASSO). This note concerns SLOPE regularization, introduced for the first time in (Bogdan et al. 2015), which assumes $\phi$ to be the sorted $l^1$ penalty, i.e.

$$\phi(x) = \phi_\theta(x) = \sum_{i=1}^{n} \theta_i x_i^{\downarrow},$$

where $x_1^{\downarrow} \geq x_2^{\downarrow} \geq \ldots \geq x_n^{\downarrow}$ is the ordered permutation of the vector $(|x_1|, |x_2|, \ldots, |x_n|)$ and $\theta \in \mathbb{R}_+^n$ is a hyperparameter of the model. Such a choice is a generalization of the $l^1$ regularization, as can be seen by taking $\theta_i = \text{const}$.

The fact that $\phi_\theta$ is non-separable makes the analysis of its teoretical properties much more onerous than in case of classical (separable) models, cf. (Bogdan et al. 2015, Bu et al. (2019)). Nonetheless, it turns out that SLOPE has two advantages over other regularization methods such as LASSO and knocoffs, namely:

1. it achieves certain minimax estimation properties under particular random designs *without* requiring any knowledge of the sparsity degree of $\widehat{x}$, cf. (Bellec, Lecué, and Tsybakov 2018);
2. it controls the false discovery rate in the case of independent predictors, cf. (Bogdan et al. 2015).

We are interested in the algorithmic solutions to the problem (2). Since the objective function in (2) is convex but not smooth, one can not apply directly the classical gradient descent and has to turn to other methods.

A natural alternative solution is the plethora of proximal algorithms, e.g. ISTA (and its improvement – FISTA, cf. (A. Beck and Teboulle 2009)) or more classical alternating direction methods of multipliers (ADMM).

The methods have been throughly studied in the literature, cf. (Beck A 2017) for a detailed treatment of the subject.

In this note we will focus on another approach, via the approximate message passing, considered for the first time in context of the LASSO problem in (Donoho, Maleki, and Montanari 2009) and subsequentially developed in e.g. (Bayati and Montanari 2011), and for the SLOPE regularization in (Bu et al. 2019) – see e.g. (Zdeborová and Krzakala 2016) for an accessible derivation of the method.

In the subsequent sections we will describre briefly these approaches.

### Proximal methods

Denoting $g(x) = \frac{1}{2}\|Ax - y\|_2^2$, the ISTA iteration can be written as:

- gradient step: $r_k = x_k - \alpha \nabla g(x_k)$
- proximal step: $x_{k+1} = \mathrm{prox}_{\gamma\phi}(r_k)$,

where $\gamma$ is the learning rate and prox denotes the proximal operator given by

$$\mathrm{prox}_{\gamma\phi}(y) = \underset{x}{\mathrm{argmin}}\{\, \gamma\phi(x) + \frac{1}{2}|x - y|^2 \,\}.$$

It is known that the rate of convergence of this method is $O(1/k)$, where $k$ is the number of steps.

### Approximate message passing

## Numerical experiments

### Comparison with convex optimization methods

## References

Bayati, M., and A. Montanari. 2011. "The Dynamics of Message Passing on Dense Graphs, with Applications to Compressed Sensing." *IEEE Trans. Inform. Theory* 57 (2): 764–85. doi:10.1109/TIT.2010.2094817.

Beck, A., and M. Teboulle. 2009. "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems." *SIAM J. Imaging Sci.* 2 (1): 183–202. doi:10.1137/080716542.

Beck, A. 2017. *First-Order Methods in Optimization.* Vol. 25. MOS-Siam Series on Optimization. Society for Industrial; Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA. doi:10.1137/1.9781611974997.ch1.

Bellec, P. C., G. Lecué, and A. B. Tsybakov. 2018. "Slope Meets Lasso: Improved Oracle Bounds and Optimality." *Ann. Statist.* 46 (6B): 3603–42. doi:10.1214/17-AOS1670.

Bogdan, M., E. van den Berg, C. Sabatti, W. Su, and E. J. Candès. 2015. "SLOPE—adaptive Variable Selection via Convex Optimization." *Ann. Appl. Stat.* 9 (3): 1103–40. doi:10.1214/15-AOAS842.

Bu, Z., J. Klusowski, C. Rush, and W. Su. 2019. "Algorithmic Analysis and Statistical Estimation of Slope via Approximate Message Passing." In *Advances in Neural Information Processing Systems*, 9366–76. http://par.nsf.gov/biblio/10163278.

Donoho, D., A. Maleki, and A. Montanari. 2009. "Message-Passing Algorithms for Compressed Sensing." *Proceedings of the National Academy of Sciences* 106 (45). National Acad Sciences: 18914–9. doi:https://doi.org/10.1073/pnas.0909892106.

Zdeborová, L., and F. Krzakala. 2016. "Statistical Physics of Inference: Thresholds and Algorithms." *Advances in Physics* 65 (5). Taylor & Francis: 453–552. doi:https://doi.org/10.1080/00018732.2016.1211393.