OPT2020: 12th Annual Workshop on Optimization for Machine Learning

Asymptotic Analysis of Sparse Group LASSO via Approximate Message Passing Algorithm

Kan Chen KANCHEN@UPENN.EDU
Shiyun Xu SHIYUNXU@UPENN.EDU
Zhiqi Bu ZBU@UPENN.EDU
University of Pennsylvania, USA

Abstract

Sparse Group LASSO (SGL) is a regularized model for high-dimensional linear regression problems with grouped covariates. SGL applies l_1 and l_2 penalties to the individual predictors and group predictors, respectively, to guarantee sparse effects both on the inter-group and within-group levels. In this paper, we apply the approximate message passing (AMP) algorithm to efficiently solve the SGL problem under Gaussian random designs. We use AMP and a recently developed state evolution analysis for non-separable penalties to derive an asymptotically exact characterization of the SGL solution, which allows us to conduct fine-grained statistical analysis of the solution.

1. Introduction

Suppose we observe an $n \times p$ design matrix $\mathbf{X}$, and the response $\mathbf{y} \in \mathbb{R}^n$ which is modeled by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{w} \tag{1}$$

in which $\boldsymbol{w} \in \mathbb{R}^n$ is a noise vector.

To yield both sparsity of groups and sparsity within each group, [13] introduced the Sparse Group LASSO problem as follows:

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2}\|\mathbf{y} - \sum_{l=1}^{L} \mathbf{X}_l \boldsymbol{\beta}_l\|_2^2 + (1-\gamma)\lambda \sum_{l=1}^{L} \sqrt{p_l}\|\boldsymbol{\beta}_l\|_2 + \gamma\lambda\|\boldsymbol{\beta}\|_1 \tag{2}$$

where $\gamma \in [0, 1]$ refers to the proportion of the LASSO fit in the overall penalty. We assume that p predictors are divided into L groups and denote the size of the l-th group as p_l. If $\gamma = 1$, SGL is purely LASSO, while if $\gamma = 0$, SGL reduces to Group LASSO. We denote the solution to the SGL problem as $\hat{\boldsymbol{\beta}}$.

In this paper, we derive the proximal operator for SGL and establish *approximate message passing* (AMP) [1–3, 6, 7, 10] for SGL from this new approach. We then analyze the algorithmic aspects of SGL via AMP. In general, AMP is a class of computationally efficient gradient-based algorithms originating from graphical models and extensively studied for many compressed sensing problems [9, 11].

We derive, for fixed γ, the SGL AMP as follows: set $\boldsymbol{\beta}^0 = \mathbf{0}$, $\boldsymbol{z}^0 = y$ and for $t > 0$,

$$\boldsymbol{\beta}^{t+1} = \eta_\gamma(\mathbf{X}^\top \mathbf{z}^t + \boldsymbol{\beta}^t, \theta_t) \tag{3}$$

$$\mathbf{z}^{t+1} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{t+1} + \frac{1}{\delta}\mathbf{z}^t \langle \eta'_\gamma(\mathbf{X}^\top \mathbf{z}^t + \boldsymbol{\beta}^t, \theta_t)\rangle. \tag{4}$$

Here the threshold θ_t is carefully designed and can be found in [2]. $\langle \boldsymbol{v} \rangle := \sum_{i=1}^{p} v_p/p$ is the average of vector $\boldsymbol{v}$. Furthermore, η_γ is the proximal operator

$$\eta_\gamma(\boldsymbol{s}, \lambda) := \underset{\boldsymbol{b}}{\operatorname{argmin}} \frac{1}{2}\|\boldsymbol{s} - \boldsymbol{b}\|^2 + (1-\gamma)\lambda \sum_{l=1}^{L} \sqrt{p_l}\|\boldsymbol{b}_l\|_2 + \gamma\lambda\|\boldsymbol{b}\|_1$$

and $\eta'_\gamma := \nabla \circ \eta_\gamma$ is the diagonal of the Jacobian matrix of the proximal operator with respect to its first argument $\boldsymbol{s}$, with $\circ$ being the Hadamard product.

Empirically, the simulation results in Table 1 demonstrate the supremacy of AMP convergence speed over the two most well-known proximal gradient descent methods, ISTA and FISTA. We also compare these methods to the Nesterov-accelerated blockwise descent in [13] and in R package `SGL`. We note that the Nesterov-accelerated ISTA (i.e. FISTA) outperforms the accelerated blockwise descent in terms of both the number of iterations and the wall-clock time. This observation suggests that using the proximal operator not only requires fewer iterations but also reduces the complexity of computation at each iteration. We pause to emphasize that, in general, the cost function $\mathcal{C}_{\boldsymbol{X},\boldsymbol{y}}(\boldsymbol{\beta}) := \frac{1}{2}\|\mathbf{y} - \sum_{l=1}^{L} \mathbf{X}_l\boldsymbol{\beta}_l\|_2^2 + (1-\gamma)\lambda\sum_{l=1}^{L}\sqrt{p_l}\|\boldsymbol{\beta}_l\|_2 + \gamma\lambda\|\boldsymbol{\beta}\|_1$ is not strictly convex. We choose the optimization error (mean squared error, or MSE, between $\boldsymbol{\beta}^t$ and $\boldsymbol{\beta}$) as the measure of convergence, as there may exist $\hat{\boldsymbol{\beta}}$ far from $\boldsymbol{\beta}$ for which $\mathcal{C}(\hat{\boldsymbol{\beta}})$ is close to $\mathcal{C}(\boldsymbol{\beta})$.

Number of Iterations				
MSE	10^{-2}	10^{-3}	10^{-4}	10^{-5}
ISTA	309	629	988	1367
FISTA	42	81	158	230
AMP	4	6	14	35

Table 1: $p = 4000, n = 2000, \gamma = 0.5, \boldsymbol{g} = (1, \cdots, 1)$, the entries of $\mathbf{X}$ are i.i.d. $\mathcal{N}(0, 1/n)$, $\lambda = 1$, and the prior $\boldsymbol{\beta}_0$ is $5\times$Bernoulli(0.1)

Our contributions are as follows. We first derive a proximal operator of SGL on which the SGL AMP is based. We prove that the algorithm solves the SGL problem *asymptotically exactly* under i.i.d. Gaussian designs. The proof leverages the recent state evolution analysis [4] for non-separable penalties and shows that the state evolution characterizes the asymptotically exact behaviors of $\hat{\boldsymbol{\beta}}$. Specifically, the distribution of SGL solution is completely specified by a few parameters that are the solution to a certain fixed-point equation asymptotically. As a consequence, we can use the characterization of the SGL solution to analyze the behaviors of the $\hat{\boldsymbol{\beta}}$ precisely. The rest of this paper is divided into four sections. In Section 2, we give some preliminary background of the AMP algorithm. In Section 3, we state our main theorems about the convergence and the characterization. In Section 4, we conclude our paper and list some possible extensions of future work.

2. Algorithm

2.1. Approximate Message Passing

We first list the assumptions.

- (**A1**) The measurement matrix $\mathbf{X}$ has independent entries following $\mathcal{N}(0, \frac{1}{n})$.

- (**A2**) The elements of signal $\boldsymbol{\beta}$ are i.i.d. copies of a random variable Π with $\mathbb{E}(\Pi^2 \max\{0, \log(\Pi)\}) < \infty$. We use $\boldsymbol{\Pi} \in \mathbb{R}^p$ to denote random vector with each component following i.i.d. Π.
- (**A3**) The elements of noise $\boldsymbol{w}$ are i.i.d. W with $\sigma_{\boldsymbol{w}}^2 := \mathbb{E}(W^2) < \infty$.
- (**A4**) The ratio of sample size to feature size $\frac{n}{p}$ approaches a constant $\delta \in (0, \infty)$ as $n, p \to \infty$.

We note that the assumptions are the same as in [5] and the second-moment assumptions **(A2)** and **(A3)** can be relaxed. For example, we can instead assume that $\boldsymbol{w}$ has an empirical distribution that converges weakly to W, with $\|\boldsymbol{w}\|^2/p \to \mathbb{E}(W^2) < \infty$. In general, we may extend assumptions **(A1)** and **(A2)** to a much broader range of cases. Additionally, we need one extra assumption for the group information as follows.

- (**A5**) The relative ratio of each group size, p_l/p, converges to $r_l \in (0, 1)$ as $p \to \infty$.

Now we can write the SGL AMP algorithm based on [7]:

$$\boldsymbol{\beta}^{t+1} = \eta_{\gamma,\boldsymbol{g}}(\mathbf{X}^\top \mathbf{z}^t + \boldsymbol{\beta}^t, \alpha\tau_t) \tag{5}$$

$$\mathbf{z}^{t+1} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{t+1} + \frac{1}{\delta}\mathbf{z}^t \langle \eta'_{\gamma,\boldsymbol{g}}(\mathbf{X}^\top \mathbf{z}^t + \boldsymbol{\beta}^t, \alpha\tau_t)\rangle \tag{6}$$

$$\tau_{t+1}^2 = \sigma_{\boldsymbol{w}}^2 + \lim_{p\to\infty} \frac{1}{\delta p}\mathbb{E}\|\eta_{\gamma,\boldsymbol{g}}(\boldsymbol{\Pi} + \tau_t \boldsymbol{Z}, \alpha\tau_t) - \boldsymbol{\Pi}\|_2^2 \tag{7}$$

where $\boldsymbol{Z}$ is the standard Gaussian $\mathcal{N}(0, \mathcal{I}_p)$ and the expectation is taken with respect to both $\boldsymbol{\Pi}$ and $\boldsymbol{Z}$. We denote $\eta_{\gamma,\boldsymbol{g}}(\boldsymbol{s}, \lambda) : \mathbb{R}^p \times \mathbb{R} \to \mathbb{R}^p$ as the proximal operator for SGL, which we will derive in appendix. We notice that, comparing AMP to the standard proximal gradient descent, the thresholds are related to (α, τ_t) instead of to λ. On one hand, τ_t is derived from equation equation 7, known as the **state evolution**, which relies on α. On the other hand, α corresponds uniquely to λ via equation (8) which is so called **calibration**:

$$\lambda = \alpha\tau_* \left(1 - \lim_{p\to\infty} \frac{1}{\delta}\langle \eta'_{\gamma,\boldsymbol{g}}(\boldsymbol{\Pi} + \tau_* \boldsymbol{Z}, \alpha\tau_*)\rangle\right) \tag{8}$$

in which $\tau_t \to \tau_*$ as $t \to \infty$.

3. Main Result

3.1. State Evolution and Calibration

Notice that in SGL AMP, we use θ_t as the threshold, whose design requires state evolution and calibration. Thus we start with some properties of state evolution recursion (7). To simplify the analysis, we consider the finite approximation of state evolution and present precise conditions which guarantee that the state evolution converges efficiently.

Proposition 1 *Let* $\mathbf{F}_\gamma(\tau_t^2, \alpha\tau_t) = \sigma_{\boldsymbol{w}}^2 + \frac{1}{\delta p}\mathbb{E}\|\eta_\gamma(\boldsymbol{\Pi} + \tau_t \boldsymbol{Z}, \alpha\tau_t) - \boldsymbol{\Pi}\|_2^2$ *and define* $\mathcal{A}(\gamma) = \{\alpha : \delta \geq 2T(\gamma\alpha) - 2(1-\gamma)\alpha\sqrt{2T(\gamma\alpha)} + (1-\gamma)^2\alpha^2\}$ *with* $T(z) = (1+z^2)\Phi(-z) - z\phi(z)$*,* $\phi(z)$ *being the standard Gaussian density and* $\Phi(z) = \int_{-\infty}^{z} \phi(x)dx$*. For any* $\sigma_{\boldsymbol{w}}^2 > 0$*,* $\alpha \in \mathcal{A}(\gamma)$ *, the fixed point equation* $\tau^2 = \mathbf{F}_\gamma(\tau^2, \alpha\tau)$ *admits a unique solution. Denoting the solution as* $\tau_* = \tau_*(\alpha)$*, we have* $\lim_{t\to\infty} \tau_t \to \tau_*(\alpha)$*, where the convergence is monotone under any initial condition. Finally* $\left|\frac{d\mathbf{F}_\gamma}{d\tau^2}\right| < 1$ *at* $\tau = \tau_*$

We note that for all $\gamma < 1$, $\mathcal{A}$ has upper and lower bounds; however, when $\gamma = 1$, i.e. for LASSO, there is no upper bound.

Before we employ the finite approximation of state evolution to describe the calibration (8), we explain the necessity of calibration by the following lemma.

Lemma 2 *For fixed γ, a stationary point $\hat{\boldsymbol{\beta}}$ with corresponding $\hat{\boldsymbol{z}}$ of the AMP iteration (5), (6) with $\theta_t = \theta_*$ is a minimizer of the SGL cost function in (2) with $\lambda = \theta_* \left(1 - \frac{1}{\delta}\langle \eta_\gamma'(\mathbf{X}^\top \hat{\mathbf{z}} + \hat{\boldsymbol{\beta}}, \theta_*)\rangle\right)$.*

Setting $\theta_* = \alpha\tau_*$, we are now in the position to define the finite approximation of calibration between α and λ by $\lambda = \alpha\tau_* \left(1 - \frac{1}{\delta}\langle \eta_\gamma'(\mathbf{\Pi} + \tau_* \boldsymbol{Z}, \alpha\tau_*)\rangle\right)$. In practice, we need to invert it to input λ and recover $\alpha(\lambda) \in \{a \in \mathcal{A} : \lambda(a) = \lambda\}$. The next proposition and corollary imply that the mapping of $\lambda \to \alpha(\lambda)$ is well-defined and easy to compute.

Proposition 3 *The function $\alpha \to \lambda(\alpha)$ is continuous on $\mathcal{A}(\gamma)$ with $\lambda(\min \mathcal{A}) = -\infty$ and $\lambda(\max \mathcal{A}) = \lambda_{\max}$ for some constant $\lambda_{\max}$ depending on Π and γ. Therefore, the function $\lambda \to \alpha(\lambda)$ satisfying $\alpha(\lambda) \in \{\alpha \in \mathcal{A}(\gamma) : \lambda(\alpha) = \lambda\}$ exists where $\lambda \in (-\infty, \lambda_{\max})$.*

Given λ, Proposition 3 claims that α exists and the following result guarantees its uniqueness.

Corollary 4 *For $\lambda < \lambda_{\max}, \sigma_{\boldsymbol{w}}^2 > 0$, $\exists! \alpha \in \mathcal{A}(\gamma)$ such that $\lambda(\alpha) = \alpha\tau_* \left(1 - \frac{1}{\delta}\langle \eta_\gamma'(\mathbf{\Pi} + \tau_* \boldsymbol{Z}, \alpha\tau_*)\rangle\right)$. Hence the function $\lambda \to \alpha(\lambda)$ is continuous and non-decreasing with $\alpha((-\infty, \lambda_{\max})) = \mathcal{A}(\gamma)$.*

3.2. AMP State Evolution Characterizes SGL Estimate

Having described the state evolution, we now state our main theoretical results. We establish an asymptotic equality between $\hat{\boldsymbol{\beta}}$ and η_γ in pseudo-Lipschitz norm, which allows the fine-grained statistical analysis of the SGL minimizer.

Definition 5 *[4]: For $k \in \mathbb{N}_+$, a function $\phi : \mathbb{R}^d \to \mathbb{R}$ is **pseudo-Lipschitz** of order k, if there exists a constant L such that for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$,*

$$
|\phi(\boldsymbol{a}) - \phi(\boldsymbol{b})| \le L\left(1 + \left(\frac{\|\boldsymbol{a}\|}{\sqrt{d}}\right)^{k-1} + \left(\frac{\|\boldsymbol{b}\|}{\sqrt{d}}\right)^{k-1}\right)\left(\frac{\|\boldsymbol{a} - \boldsymbol{b}\|}{\sqrt{d}}\right). \tag{9}
$$

*A sequence (in p) of pseudo-Lipschitz functions $\{\phi_p\}_{p \in \mathbb{N}_+}$ is **uniformly pseudo-Lipschitz** of order k if, denoting by L_p the pseudo-Lipschitz constant of ϕ_p, $L_p < \infty$ for each p and $\limsup_{p\to\infty} L_p < \infty$.*

Theorem 6 *Under the assumptions (**A1**)-(**A5**), for any uniformly pseudo-Lipschitz sequence of function $\varphi_p : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ and for $\mathbf{Z} \sim \mathcal{N}(0, \mathcal{I}_p)$, $\mathbf{\Pi} \sim p_\Pi$, $\lim_{p\to\infty} \varphi_p(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}) = \lim_t \lim_p \mathbb{E}[\varphi_p(\eta_\gamma(\mathbf{\Pi} + \tau_t \boldsymbol{Z}; \alpha\tau_t), \mathbf{\Pi})]$.*

Essentially, up to a uniformly pseudo-Lipschitz loss, we can replace $\hat{\boldsymbol{\beta}}$ by η_γ in the large system limit. The distribution of η_γ is explicit, thus allowing the analysis of certain quantities. For instance, the true positive rate (or recall) $\mathbb{P}(\hat{\beta}_i \neq 0 | \beta_i \neq 0)$ can be well-approximated by $\mathbb{P}(\eta_\gamma(\Pi^* + \tau_* Z; \alpha\tau_*) \neq 0)$, where Π^* is the random variable of signals conditioned on being non-zero.

Specifically, if we use $\varphi(\boldsymbol{a}, \boldsymbol{b}) = \|\boldsymbol{a} - \boldsymbol{b}\|_2^2$, the MSE between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}$ can be characterized by τ.

Corollary 7 *Under the assumptions **(A1)**-**(A5)**, then almost surely,* $\lim_{p\to\infty}\frac{1}{p}\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_2^2=\delta(\tau_*^2-\sigma_{\boldsymbol{w}}^2)$.

Now that we have demonstrated the usefulness of our main theoretical result. We first show the convergence of $\boldsymbol{\beta}^t$ to $\hat{\boldsymbol{\beta}}$, i.e. the AMP iterates converge to the true minimizer.

Theorem 8 *Under assumptions **(A1)**-**(A5)**, for the output of the AMP algorithm in (5) and the Sparse Group LASSO estimator given by the solution of (2),* $\lim_{p\to\infty}\frac{1}{p}\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^t\|_2^2=k_t$, *where* $\lim_{t\to\infty}k_t=0$.

In addition to Theorem 8, we borrow the state evolution analysis from [4] Theorem 14 to complete the proof of Theorem 6.

Lemma 9 *[4] Under assumptions **(A1)** - **(A5)**, given that **(S1)** and **(S2)** are satisfied, consider the recursion equation 5 and equation 6. For any uniformly pseudo-Lipschitz sequence of functions* $\phi_n:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$ *and* $\varphi_p:\mathbb{R}^p\times\mathbb{R}^p\to\mathbb{R}$,

$$\phi_n(\mathbf{z}^t,\mathbf{w})\xrightarrow{\mathbb{P}}\mathbb{E}\left[\phi_n(\mathbf{w}+\sqrt{\tau_t^2-\sigma_w^2}\mathbf{Z}',\mathbf{w})\right] \tag{10}$$

$$\varphi_p(\boldsymbol{\beta}^t+\mathbf{X}^\top\mathbf{z}^t,\Pi)\xrightarrow{\mathbb{P}}\mathbb{E}\left[\varphi_p(\Pi+\tau_t\mathbf{Z},\Pi)\right] \tag{11}$$

where τ_t *is defined in equation 7,* $\mathbf{Z}'\sim\mathcal{N}(\mathbf{0},\mathcal{I}_n)$ *and* $\mathbf{Z}\sim\mathcal{N}(\mathbf{0},\mathcal{I}_p)$.

To see that Theorem 1 holds, we combine $\hat{\boldsymbol{\beta}}\approx\boldsymbol{\beta}^t$ from Theorem 8 and $\boldsymbol{\beta}^t+\boldsymbol{X}^\top\boldsymbol{z}^t\approx\boldsymbol{\Pi}+\tau_t\boldsymbol{Z}$ from Lemma 9 to obtain that $\hat{\boldsymbol{\beta}}=\eta_\gamma(\boldsymbol{\beta}^t+\mathbf{X}^\top\mathbf{z}^t,\alpha\tau_t)$ within uniformly pseudo-Lipschitz loss, asymptotically, in large system limit.

In order to apply Lemma 9, we need to check the proximal operator of SGL satisfies the following properties.

(S1) For each t, the proximal operators are uniformly Lipschitz (i.e. uniformly pseudo- Lipschitz of order $k=1$).

(S2) For any s,t with $(\mathbf{Z},\mathbf{Z}')$ a pair of length p vectors such that (Z_i,Z_i') are i.i.d $\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma})$ for $i\in\{1,2,\cdots,p\}$ where $\boldsymbol{\Sigma}$ is any 2×2 covariance matrix, the following limits exist and are finite:

$$\lim_{p\to\infty}\frac{1}{p}\|\boldsymbol{\beta}\|_2,\ \lim_{p\to\infty}\frac{1}{p}\mathbb{E}\left(\boldsymbol{\beta}^T\eta_\gamma(\boldsymbol{\beta}+\mathbf{Z},\alpha\tau_t)\right),\ \lim_{p\to\infty}\frac{1}{p}\mathbb{E}\left(\eta_\gamma(\boldsymbol{\beta}+\mathbf{Z}',\alpha\tau_s)^\top\eta_\gamma(\boldsymbol{\beta}+\mathbf{Z},\alpha\tau_t)\right) \tag{12}$$

4. Discussion and Future Work

Our work suggests several possible future research. In one direction, it is promising to extend the proximal algorithms (especially AMP) to a broader class of models with structured sparsity, such as the sparse linear regression with overlapping groups, Group SLOPE and the sparse group logistic regression. On a different road, although AMP is robust in distributional assumptions in the sense of fast convergence under i.i.d. non-Gaussian measurements, multiple variants of AMP may be applied to adapt to real-world data. To name a few, one may look into SURE-AMP [8], EM-AMP [14, 15] and VAMP [12] to relax the known signal assumption and non-i.i.d. measurement assumption.

References

[1] Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2): 764–785, 2011.

[2] Mohsen Bayati and Andrea Montanari. The lasso risk for gaussian matrices. *IEEE Transactions on Information Theory*, 58(4):1997–2017, 2011.

[3] Mohsen Bayati, Murat A Erdogdu, and Andrea Montanari. Estimating lasso risk and noise level. In *Advances in Neural Information Processing Systems*, pages 944–952, 2013.

[4] Raphael Berthier, Andrea Montanari, and Phan-Minh Nguyen. State evolution for approximate message passing with non-separable functions. *arXiv preprint arXiv:1708.03950*, 2017.

[5] Zhiqi Bu, Jason Klusowski, Cynthia Rush, and Weijie Su. Algorithmic analysis and statistical estimation of slope via approximate message passing. *arXiv preprint arXiv:1907.07502*, 2019.

[6] David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

[7] David L Donoho, Arian Maleki, and Andrea Montanari. How to design message passing algorithms for compressed sensing. *preprint*, 2011.

[8] Chunli Guo and Mike E Davies. Near optimal compressed sensing without priors: Parametric sure approximate message passing. *IEEE Transactions on Signal Processing*, 63(8):2130–2141, 2015.

[9] Florent Krzakala, Marc Mézard, Francois Sausset, Yifan Sun, and Lenka Zdeborová. Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(08): P08009, 2012.

[10] Andrea Montanari, YC Eldar, and G Kutyniok. Graphical models concepts in compressed sensing. *Compressed Sensing: Theory and Applications*, pages 394–438, 2012.

[11] Sundeep Rangan. Generalized approximate message passing for estimation with random linear mixing. In *2011 IEEE International Symposium on Information Theory Proceedings*, pages 2168–2172. IEEE, 2011.

[12] Sundeep Rangan, Philip Schniter, and Alyson K Fletcher. Vector approximate message passing. *IEEE Transactions on Information Theory*, 2019.

[13] Noah Simon, Jerome Friedman, Trevor Hastie, and Robert Tibshirani. A sparse-group lasso. *Journal of Computational and Graphical Statistics*, 22(2):231–245, 2013.

[14] Jeremy Vila and Philip Schniter. Expectation-maximization bernoulli-gaussian approximate message passing. In *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pages 799–803. IEEE, 2011.

[15] Jeremy P Vila and Philip Schniter. Expectation-maximization gaussian-mixture approximate message passing. *IEEE Transactions on Signal Processing*, 61(19):4658–4672, 2013.