Graph-based Approximate Message Passing Iterations

Cédric Gerbelot[1] and Raphaël Berthier[2]

[1]Laboratoire de Physique de l'Ecole Normale Supérieure, Université PSL, CNRS, Paris, France
[2]Inria - Département d'informatique de l'Ecole Normale Normale Supérieure, Université PSL, Paris, France

Motivation : exactly solvable models

Exactly solvable models are **high dimensional, random design** machine learning models whose properties (generalization error, information theoretic limits, ...) can be entirely described by a set of **low dimensional** parameters.

A typical formulation is the **teacher-student generalized linear model** :

observe the "*teacher*" generative model

$$\mathbf{y} = f_0(\mathbf{A}\mathbf{w}_0) \in \mathbb{R}^N, \quad \mathbf{w}_0 \in \mathbb{R}^d, \quad \mathbf{A} \in \mathbb{R}^{N\times d} \text{ i.i.d. } \mathbf{N}(0, 1/N)$$

learn with the "*student*" model

$$\mathbf{w}^\star \in \underset{\mathbf{w}\in\mathbb{R}^d}{\operatorname{argmin}}\, L(\mathbf{y}, \mathbf{A}\mathbf{w}) + r(\mathbf{w}) \qquad (1)$$

- L, r are given loss and penalty functions
- $N, d \to \infty$ with fixed ratio

Goal : statistical properties/distribution of $\mathbf{w}^\star$

Approximate Message Passing (AMP) iterations

Algorithms inspired by statistical physics involving **random matrices, non-linearities, and a specific correction term.** The distribution of AMP iterates can be **exactly characterized by a low dimensional recursion** at each time step, the **state evolution (SE) equations**. Powerful solver/proof method for exactly solvable models.

Examples of problems solved with an AMP

Spiked matrix recovery:
AMP proposed in, e.g., [RF12], rigorous SE [JM13]. Recover $\mathbf{v}_0 \in \mathbb{R}^N$ from:

$$\mathbf{Y} = \sqrt{\frac{\lambda}{d}}\mathbf{v}_0\mathbf{v}_0^\top + \mathbf{W}$$

where the noise matrix $\mathbf{W} \in GOE(N)$.

Generalized linear modelling (problem (1)):
AMP in, e.g., [Ran11], rigorous SE [BM11], [JM13]. Includes the LASSO, logistic regression, etc . . .

Multilayer generalized linear estimation:
AMP in [MKMZ17], **heuristic SE**. Recover $\mathbf{x}_0 \in \mathbb{R}^{N_1}$ from

$$\mathbf{y} = \phi_L(\mathbf{A}_L\phi_{L-1}(\mathbf{A}_{L-1}(\dots\phi_1(\mathbf{A}_1\mathbf{x}_0)))$$

where, for each layer l, the matrix $\mathbf{A}_l \in \mathbb{R}^{N_{l+1}\times N_l}$ has i.i.d. $\mathbf{N}(0, \frac{1}{N_l})$ with $N_{l+1}/N_l = \delta_l \in [0, 1]$.

Spiked matrix with generative prior:
AMP in [ALM+20], **heuristic SE**. Same notations as before, recover $\mathbf{v}_0$ from

$$\mathbf{Y} = \sqrt{\frac{\lambda}{d}}\mathbf{v}_0\mathbf{v}_0^\top + \mathbf{W}$$

where $\mathbf{v}_0$ has the generative prior:

$$\mathbf{v}_0 = \phi_L(\mathbf{A}_L\phi_{L-1}(\mathbf{A}_{L-1}(\dots\phi_1(\mathbf{A}_1\mathbf{x}_0)))$$

What's the problem ?

- **stat. phys. intuition not grounded in machine learning**
- **SE proofs are tedious and done on a case by case basis**

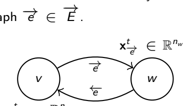
Contributions

- **unifying framework for AMP iterations in the form of an oriented graph**
- **prove SE equations for any graph-based AMP**
- prove recent heuristic SE equations, extend the reach of SE proofs
- offer new design possibilities for AMP iterations

Graph-based AMP iterations

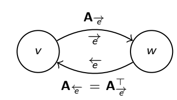
Consider a symmetric finite directed graph $G = (V, \overrightarrow{E})$. We associate an AMP iteration supported by the graph G as follows.

- The variables $\mathbf{x}^t_{\overrightarrow{e}}$ of the AMP iteration are indexed by the iteration number $t \in \mathbb{N}$ and the oriented edges of the graph $\overrightarrow{e} \in \overrightarrow{E}$.

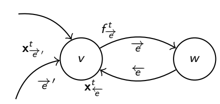


- All variables associated to edges $\overrightarrow{e} = (v, w)$ with end-node $w \in V$ have a same dimension $n_w \in \mathbb{N}_{>0}$, i.e., $\mathbf{x}^t_{\overrightarrow{e}} \in \mathbb{R}^{n_w}$. We define $N = \sum_{(v,w)\in\overrightarrow{E}} n_w$ the sum of the dimensions of all variables.

- Matrices of the AMP iteration are also indexed by the edges of the graph, and all have i.i.d. $\mathbf{N}(0, 1/N)$ elements. If $\overrightarrow{e} = (v, w) \in \overrightarrow{E}$, $\mathbf{A}_{\overrightarrow{e}} \in \mathbb{R}^{n_w\times n_v}$. These matrices must satisfy the symmetry condition $\mathbf{A}_{(v,w)} = \mathbf{A}^\top_{(w,v)}$. In particular, this implies that matrices $\mathbf{A}_{(v,v)} \in \mathbb{R}^{n_v\times n_v}$ associated to loops $(v, v) \in \overrightarrow{E}$ must be symmetric.



- Non-linearities of the AMP iteration are also indexed by the edges of the graph (and possibly by the iteration number t). If $t \geq 0$ and $\overrightarrow{e} = (v, w) \in \overrightarrow{E}$, $f^t_{(v,w)}\left((\mathbf{x}^t_{\overrightarrow{e}'})_{\overrightarrow{e}':\overrightarrow{e}'\to\overrightarrow{e}}\right)$ is a function of all the variables of the edges whose end-node is the starting-node v of $\overrightarrow{e}$, as denoted by the condition $\overrightarrow{e}' \to \overrightarrow{e}$. It is a function from $(\mathbb{R}^{n_v})^{\deg v}$ to $\mathbb{R}^{n_v}$.



Consider a given an arbitrary initial condition $\mathbf{x}^0_{\overrightarrow{e}} \in \mathbb{R}^{n_w}$ for all oriented edges $\overrightarrow{e} \in \overrightarrow{E}$ of the graph. We define recursively the AMP iterates $(\mathbf{x}^t_{\overrightarrow{e}})_{t\geq 0, \overrightarrow{e}\in\overrightarrow{E}}$, by the iteration: for all $t \geq 0$, $\overrightarrow{e} \in \overrightarrow{E}$,

$$\mathbf{x}^{t+1}_{\overrightarrow{e}} = \mathbf{A}_{\overrightarrow{e}}\mathbf{m}^t_{\overrightarrow{e}} - b^t_{\overrightarrow{e}}\mathbf{m}^{t-1}_{\overleftarrow{e}}\,, \qquad (2)$$

$$\mathbf{m}^t_{\overrightarrow{e}} = f^t_{\overrightarrow{e}}\left((\mathbf{x}^t_{\overrightarrow{e}'})_{\overrightarrow{e}':\overrightarrow{e}'\to\overrightarrow{e}}\right)\,, \qquad (3)$$

where $b^t_{\overrightarrow{e}}$ is the so-called *Onsager correction term*

$$b^t_{\overrightarrow{e}} = \frac{1}{N}\operatorname{Tr}\frac{\partial f^t_{\overrightarrow{e}}}{\partial\mathbf{x}_{\overleftarrow{e}}}\left((\mathbf{x}^t_{\overrightarrow{e}'})_{\overrightarrow{e}':\overrightarrow{e}'\to\overrightarrow{e}}\right) \in \mathbb{R}\,. \qquad (4)$$

The above partial derivative makes sense as $\overleftarrow{e} \to \overrightarrow{e}$, thus $\mathbf{x}_{\overleftarrow{e}}$ is a variable of $f^t_{\overrightarrow{e}}$. Note that in (2), the Onsager term multiplies the vector $\mathbf{m}^{t-1}_{\overleftarrow{e}}$ indexed by the symmetric edge $\overleftarrow{e}$ of $\overrightarrow{e}$.

Main theorem : state evolution (SE)

Notation: For any matrix $\boldsymbol{\kappa} \in \mathcal{S}^+_q$ and a random matrix $\mathbf{Z} \in \mathbb{R}^{N\times q}$ we write $\mathbf{Z} \sim \mathbf{N}(0, \boldsymbol{\kappa} \otimes \mathbf{I}_N)$ if $\mathbf{Z}$ is a matrix with jointly Gaussian entries such that for any $1 \leqslant i, j \leqslant q$, $\mathbb{E}[\mathbf{Z}^i(\mathbf{Z}^j)^\top] = \boldsymbol{\kappa}_{i,j}\mathbf{I}_N$, where $\mathbf{Z}^i$, $\mathbf{Z}^j$ denote the i-th and j-th columns of $\mathbf{Z}$. For all $v \in V$, $n_v \to \infty$ and n_v/N converges to a well-defined limit $\delta_v \in [0, 1]$. We denote by $n \to \infty$ the limit under this scaling.

Definition (State evolution iterates)

The state evolution iterates are composed of one infinite-dimensional array $(\boldsymbol{\kappa}^{s,r}_{\overrightarrow{e}})_{r,s>0}$ of real values for each edge $\overrightarrow{e} \in \overrightarrow{E}$. These arrays are generated as follows. Define the first state evolution iterates

$$\boldsymbol{\kappa}^{1,1}_{\overrightarrow{e}} = \lim_{n\to\infty}\frac{1}{N}\left\|f^0_{\overrightarrow{e}}\left((\mathbf{x}^0_{\overrightarrow{e}'})_{\overrightarrow{e}':\overrightarrow{e}'\to\overrightarrow{e}}\right)\right\|_2^2\,, \qquad \overrightarrow{e} \in \overrightarrow{E}\,.$$

Recursively, once $(\boldsymbol{\kappa}^{s,r}_{\overrightarrow{e}})_{s,r\leq t, \overrightarrow{e}\in\overrightarrow{E}}$ are defined for some $t \geq 1$, define independently for each $\overrightarrow{e} \in \overrightarrow{E}$, $\mathbf{Z}^0_{\overrightarrow{e}} = \mathbf{x}^0_{\overrightarrow{e}}$ and $(\mathbf{Z}^1_{\overrightarrow{e}}, \dots, \mathbf{Z}^t_{\overrightarrow{e}})$ a centered Gaussian random vector of covariance $(\boldsymbol{\kappa}^{r,s}_{\overrightarrow{e}})_{r,s\leq t} \otimes I_{n_w}$. We then define new state evolution iterates

$$\boldsymbol{\kappa}^{t+1,s+1}_{\overrightarrow{e}} = \boldsymbol{\kappa}^{s+1,t+1}_{\overrightarrow{e}}$$
$$= \lim_{n\to\infty}\frac{1}{N}\mathbb{E}\left[\left\langle f^s_{\overrightarrow{e}}\left((\mathbf{Z}^s_{\overrightarrow{e}'})_{\overrightarrow{e}':\overrightarrow{e}'\to\overrightarrow{e}}\right), f^t_{\overrightarrow{e}}\left((\mathbf{Z}^t_{\overrightarrow{e}'})_{\overrightarrow{e}':\overrightarrow{e}'\to\overrightarrow{e}}\right)\right\rangle\right]$$

for all $s \in \{1, \dots, t\}$, $\overrightarrow{e} \in \overrightarrow{E}$.

Theorem (Informal)

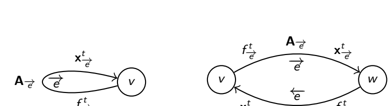
Under mild regularity assumptions, for any sequence of uniformly (in n) pseudo-Lipschitz function $\Phi : \mathbb{R}^{(t+1)N} \to \mathbb{R}$,

$$\Phi\left((\mathbf{x}^s_{\overrightarrow{e}})_{0\leq s\leq t, \overrightarrow{e}\in\overrightarrow{E}}\right) \overset{\text{P}}{\simeq} \mathbb{E}\left[\Phi\left((\mathbf{Z}^s_{\overrightarrow{e}})_{0\leq s\leq t, \overrightarrow{e}\in\overrightarrow{E}}\right)\right]$$

Included extensions : matrix-valued variables, structured/correlated and spatially coupled random (Gaussian) matrices.

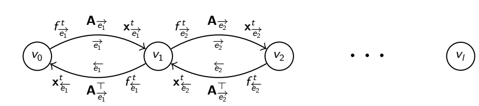
Recovering existing AMP/ rigorous SE equations

Spiked matrix recovery (left) and generalized linear modelling (right)

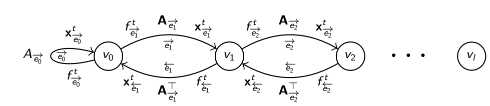


Proving heuristic SE equations

Multilayer generalized linear estimation:



Spiked matrix with generative prior:



A recent application of SE equations

Classifying a high-dimensional Gaussian mixture [LSG+21]

Generative model ("*teacher*")

$$\mathbf{x} \in \mathbb{R}^d, \mathbf{y} \in \mathbb{R}^K \quad P(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^K y_k\rho_k\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\,,$$

"*Student*"

$$\mathbf{W}^\star \in \min_{\mathbf{W}\in\mathbb{R}^{d\times K}} L(\mathbf{Y}, \mathbf{X}\mathbf{W}) + r(\mathbf{W}) \qquad (5)$$

Learn K separating hyperplanes, i.e. a matrix $\mathbf{W} \in \mathbb{R}^{d\times K}$

How to obtain the statistical properties of $\mathbf{W}^\star$?

- design an AMP s.t. its fixed point matches the optimality condition of (5)
- find a converging trajectory (convexity helps)
- statistical properties are then given by the **fixed point of the SE equations** (see main theorem of [LSG+21])

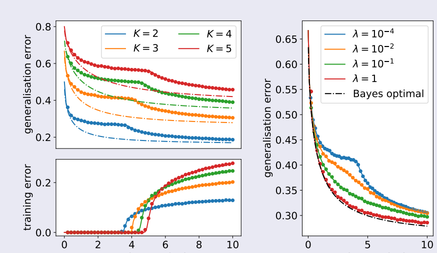


Figure: Training and generalization error for ridge penalized logistic regression on K Gaussian clusters, $\Sigma_k = \Delta Id$. (Left) Sample complexity (Right) Regularization

Future directions

- many more graphs : loops, highly connected nodes, etc . . .
- how to systematically design an AMP for a given problem
- universality and finite size rates
- other sources of randomness : randomly initialized algorithms, ...

Acknowledgements

We thank F. Krzakala and L. Zdeborova for suggesting this problem, insightful discussions and organising the Ecole des Houches Summer Workshop on Statistical Physics and Machine Learning where this work was initiated.

References

[ALM+20] Benjamin Aubin, Bruno Loureiro, Antoine Maillard, Florent Krzakala, and Lenka Zdeborová. The spiked matrix model with generative priors. *IEEE Transactions on Information Theory*, 2020.

[BM11] 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.

[JM13] Adel Javanmard and Andrea Montanari. State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Information and Inference: A Journal of the IMA*, 2(2):115–144, 2013.

[LSG+21] Bruno Loureiro, Gabriele Sicuro, Cédric Gerbelot, Alessandro Pacco, Florent Krzakala, and Lenka Zdeborová. Learning gaussian mixtures with generalised linear models: Precise asymptotics in high-dimensions. *arXiv preprint arXiv:2106.03791*, 2021.

[MKMZ17] Andre Manoel, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. Multi-layer generalized linear estimation. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 2098–2102. IEEE, 2017.

[Ran11] 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.

[RF12] Sundeep Rangan and Alyson Fletcher. Iterative estimation of constrained rank-one matrices in noise. In *2012 IEEE International Symposium on Information Theory Proceedings*, pages 1246–1250. IEEE, 2012.

Contact: cedric.gerbelot@ens.fr