Notes on State Evolution

Alex Wein

November 16, 2017

Here we introduce *state evolution*, a simple heuristic argument for the analysis of *approximate message passing* (AMP) algorithms that has been shown to be asymptotically correct in many settings. The idea of state evolution was first introduced by [DMM09] (in the setting of *compressed sensing*), based on ideas from [Bol12]; it was later proved correct in various settings [BM11, JM13].

We will focus on the Rademacher-spiked Wigner model: we observe

$$Y = \frac{\lambda}{n} x x^\top + \frac{1}{\sqrt{n}} W$$

where $x \in \{\pm 1\}^n$ is the true signal (drawn uniformly at random) and the $n \times n$ noise matrix W is symmetric with the upper triangle drawn i.i.d. as $\mathcal{N}(0,1)$. The parameter $\lambda \geq 0$ is the signal-to-noise ratio. The goal is to (approximately) recover x (up to global sign flip). In this setting, the AMP algorithm and its analysis are due to [DAM16].

The AMP algorithm for this problem takes the form

$$v^{t+1} = Y f(v^t) + [\mathit{Onsager}]$$

where $f(v)$ denotes entrywise application of the function $f(v) = \tanh(\lambda v)$. (Here we abuse notation and let f refer to both the scalar function and its entrywise application to a vector.) The superscript t indexes timesteps of the algorithm (and is not to be confused with an exponent). The Onsager term takes the form $-\langle f'(v^t) \rangle f(v^{t-1})$ where $\langle u \rangle$ denotes the average of the elements of the vector u. AMP (without the Onsager term) can be thought of as a power method with soft projection at each step. One way to derive the Onsager term is via belief propagation on graphical models.

The state evolution heuristic proceeds as follows. Postulate that at timestep t, AMP's iterate v^t is distributed as

$$v^t = \mu_t x + \sigma_t g \quad \text{where } g \sim \mathcal{N}(0, I). \tag{1}$$

This breaks down v^t into a signal term (recall x is the true signal) and a noise term, whose sizes are determined by parameters $\mu_t \in \mathbb{R}$ and $\sigma_t \in \mathbb{R}_{\geq 0}$. The idea of state evolution is to write down a recurrence for how the parameters μ_t and σ_t evolve from one timestep to the next. In performing this calculation we will make two simplifying assumptions that will be justified later: (1) we drop the Onsager term, and (2) we assume the noise W is independent at each timestep (i.e. there is no correlation between W and the noise g in the current iterate). Under these assumptions we have

$$\begin{aligned} v^{t+1} = Y f(v^t) &= \left(\frac{\lambda}{n} x x^\top + \frac{1}{\sqrt{n}} W\right) f(v^t) \\ &= \frac{\lambda}{n}\langle x, f(v^t)\rangle\, x + \frac{1}{\sqrt{n}} W f(v^t) \end{aligned}$$

which takes the form of (1) with a signal term and a noise term. We therefore have

$$\begin{aligned} \mu_{t+1} &= \frac{\lambda}{n}\langle x, f(v^t)\rangle = \frac{\lambda}{n}\langle x, f(\mu_t x + \sigma_t g)\rangle \\ &\approx \lambda\,\mathbb{E}_{X,G}[X f(\mu_t X + \sigma_t G)] \quad \text{with scalars } X \sim \mathrm{Unif}\{\pm 1\}, G \sim \mathcal{N}(0,1) \\ &= \lambda\,\mathbb{E}_G[f(\mu_t + \sigma_t G)] \quad \text{since } f(-v) = -f(v). \end{aligned}$$

For the noise term, think of $f(v^t)$ as fixed and consider the randomness over W. Each entry of the noise term $\frac{1}{\sqrt{n}} W f(v^t)$ has mean zero and variance

$$\begin{aligned} (\sigma^{t+1})^2 = \sum_i \frac{1}{n} f(v_i^t)^2 &= \sum_i \frac{1}{n} f(\mu_t x_i + \sigma_t g_i)^2 \\ &\approx \lambda\,\mathbb{E}_{X,G}[f(\mu_t X + \sigma_t G)^2] \quad \text{with scalars } X, G \text{ as above} \\ &= \mathbb{E}_G[f(\mu_t + \sigma_t G)^2] \quad \text{again by symmetry of } f. \end{aligned}$$

We now have "state evolution" equations for μ_{t+1} and σ_{t+1} in terms of μ_t and σ_t. Since we could arbitrarily scale our iterates v^t without adding or losing information, we really only care about the parameter $\gamma = (\mu/\sigma)^2$. It is possible (see [DAM16]) to reduce the state evolution recurrence to a single parameter:

$$\gamma_{t+1} = \lambda^2 \mathbb{E}_{G\sim\mathcal{N}(0,1)} \tanh(\gamma_t + \sqrt{\gamma_t}\, G) \tag{2}$$

(where we have substituted the actual expression for f).

We can analyze AMP as follows. Choose a small positive initial value γ_0 and iterate (2) until we reach a fixed point γ_∞. We then expect the output v^∞ of AMP to behave like

$$v^\infty = \mu_\infty x + \sigma_\infty g \tag{3}$$

where $g \sim \mathcal{N}(0, I)$, $\mu_\infty = \gamma_\infty/\lambda$, and $\sigma_\infty^2 = \gamma_\infty/\lambda^2$. This has in fact been proven to be correct in the limit $n \to \infty$ [BM11, JM13]. Namely, when we run AMP (with the Onsager term and without fresh noise W at each timestep), the output behaves like (3) in a particular formal sense. The Onsager term is specially designed to cancel out the effects of using the same noise at each timestep.

State evolution reveals a phase transition at $\lambda = 1$: when $\lambda \le 1$ we have $\gamma_\infty = 0$ (so AMP has zero correlation with the truth) and when $\lambda > 1$ we have $\gamma_\infty > 0$ (so AMP achieves nontrivial correlation with the truth). Furthermore, from (3) we can deduce the value of any performance metric (e.g. mean squared error) at any signal-to-noise ratio λ. It has in fact been shown (for Rademacher-spiked Wigner) that the mean squared error achieved by AMP is information-theoretically optimal [DAM16].

References

[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.

[Bol12] Erwin Bolthausen. An iterative construction of solutions of the tap equations for the sherrington-kirkpatrick model. *arXiv preprint arXiv:1201.2891*, 2012.

[DAM16] Yash Deshpande, Emmanuel Abbe, and Andrea Montanari. Asymptotic mutual information for the binary stochastic block model. In *Information Theory (ISIT), 2016 IEEE International Symposium on*, pages 185–189. IEEE, 2016.

[DMM09] 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.

[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.