Note on Approximate Message Passing

Peng Xu

Draft: January 5, 2020

1 Introduction

Generally speaking, the *approximate message passing* (AMP) algorithm is an efficient iterative approach for solving the *linear inverse problems* (LIP). We begin the introduction by reviewing the latter: consider a linear model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ are known, and $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ is a vector of random noises. The goal of LIP is to recover $\boldsymbol{\beta} \in \mathbb{R}^p$ by optimizing certain criteria. Linear regression, our favorite linear model, is an example of the unconstrained LIP. The objective of linear regression is

$$\underset{\boldsymbol{\beta}}{\text{minimize}} \quad \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 \tag{1}$$

Assuming $\boldsymbol{X}$ is of full-rank, a closed form solution can be easily derived: $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$. In practice, however, we often have some prior knowledge or assumptions on the structure of $\boldsymbol{\beta}$. A popular way to incorporate these structures is through the *regularized least-squares*:

$$\underset{\boldsymbol{\beta}}{\text{minimize}} \quad \min_{\boldsymbol{\beta}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + R(\boldsymbol{\beta}), \tag{2}$$

where $R(\boldsymbol{\beta})$ is some regularization function. For example, when $R(\boldsymbol{\beta}) = \lambda\|\boldsymbol{\beta}\|_1$ the problem is known as LASSO, and more generally, if $R(\boldsymbol{\beta}) = \lambda\|\boldsymbol{D}\boldsymbol{\beta}\|_1$ for certain penalty matrix $\boldsymbol{D} \in \mathbb{R}^{m\times p}$, the problem is known as the generalized LASSO [cite], and has useful applications such as fused LASSO [cite], trend filtering [cite], and wavelet smoothing [cite] as special cases.

Next we turn our eyes to the problem of *compressed sensing* (CS). This is a very important type of LIP that instigated the recent interest in AMP. The object of CS is

$$\begin{aligned} \underset{\boldsymbol{\beta}}{\text{minimize}} \quad & \|\boldsymbol{\beta}\|_0, \\ \text{subject to} \quad & \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}, \end{aligned} \tag{3}$$

where we define $\|\boldsymbol{x}\|_0 \coloneqq \#\{i \mid x_i = 0\}$ (not really a norm) and usually assume $n \ll p$. Of course, the ℓ_0-norm is computationally intractable so in practice it is surrogated by the so-called *basis pursuit*:

$$\begin{aligned} \underset{\boldsymbol{\beta}}{\text{minimize}} \quad & \|\boldsymbol{\beta}\|_1, \\ \text{subject to} \quad & \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}. \end{aligned} \tag{4}$$

This is a convex optimization problem and can be reformulated as a linear program (LP): let $\boldsymbol{X}^* \coloneqq \begin{pmatrix}\boldsymbol{X} & -\boldsymbol{X}\end{pmatrix}$. Then the solution $\boldsymbol{z}^*$ of

$$\begin{aligned} \underset{\boldsymbol{z}}{\text{minimize}} \quad & \mathbf{1}^\top\boldsymbol{z}, \\ \text{subject to} \quad & \boldsymbol{X}^*\boldsymbol{z} = \boldsymbol{y}, \\ & z_i > 0 \quad i = 1, \ldots, 2p \end{aligned} \tag{5}$$

is a vector of length $2p$ and can be partitioned in half as $\boldsymbol{z}^* = \begin{pmatrix}\boldsymbol{u}^* & \boldsymbol{v}^*\end{pmatrix}$. It can be shown then $\boldsymbol{x}^* = \boldsymbol{u}^* - \boldsymbol{v}^*$ solves problem (4). While linear program solvers are readily available and very efficient, for some large-scale problem the performance can still be less than satisfactory. For example, in image compression problem [GP19], $\boldsymbol{\beta}$ represents a vectorized image. Hence for a image of size 1000×1000, the resulting LP can have 2 millions of variables. This is where the AMP come into play.

2 Algorithm

As stated in the beginning, AMP is an iterative algorithm, its update scheme [DMM09] is

$$\begin{aligned}\boldsymbol{\beta}^{t+1} &= \eta_t\big(\boldsymbol{\beta}^t + \boldsymbol{X}^\top \boldsymbol{z}^t\big),\\ \boldsymbol{z}^t &= \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^t + \frac{\boldsymbol{z}^{t-1}}{n}\sum_{j=1}^{p} \eta_t'\big(\boldsymbol{X}_{\cdot,j}^\top \boldsymbol{z}^{t-1} + \boldsymbol{\beta}_j^{t-1}\big),\end{aligned}$$

where we set $\boldsymbol{\beta}^0 = \mathbf{0}$, and $\boldsymbol{X}_{\cdot,j}^\top$ denotes the j-th row of $\boldsymbol{X}^\top$. Here $\eta_t(x)$ are some component-wise scalar threshold functions and $\eta_t'(x)$ are their derivatives. At each step, η_t denoises the *effective observation* $\boldsymbol{\beta}^t + \boldsymbol{X}^\top \boldsymbol{z}^t$. The correction term $\boldsymbol{z}^t$, also known as the *modified residual*, ensures that for large enough p, $\boldsymbol{\beta}^t + \boldsymbol{X}^\top \boldsymbol{z}^t$ is close to the solution $\boldsymbol{\beta}$ plus a Gaussian noise. It is derived from the theory of belief propagation in graphical models.

Rigorous asymptotic analysis [BM11] has been given under mild assumptions: assuming $\boldsymbol{X}_{i,j} \sim N(0, 1/n)$ and $\varepsilon \sim N(0, \sigma^2)$ i.i.d. If $Z \sim N(0,1)$, and X is some random variable such that the empirical distribution of the entries of $\boldsymbol{\beta}$ coincides with X. Then under some further assumption on moments of X, we have

$$\lim_{n\to\infty} \sum_{i=1}^{n} \psi(\boldsymbol{\beta}_i^{t+1}, \boldsymbol{\beta}_i) = \mathrm{E}[\psi(\eta_t(X + \tau_t Z), X)]$$

almost surely, where $\psi\colon \mathbb{R}^2 \to \mathbb{R}$ is any pseudo-Lipschitz function of certain order and τ_t can be computed via state evolution. In term of MSE $\psi(x, y) = (x - y)^2$, this implies

$$\lim_{n\to\infty} \frac{1}{n}\|\boldsymbol{\beta}^t - \boldsymbol{\beta}\|^2 = (\tau_t^2 - \sigma^2)\delta$$

almost surely, where $\delta = n/p$ is the *sampling ratio*.

2.1 TODO: Belief Propagation, ISTA, State Evolution of MSE

References

[BM11] M. Bayati and A. Montanari. “The Dynamics of Message Passing on Dense Graphs, with Applications to Compressed Sensing”. In: *IEEE Transactions on Information Theory* 57.2 (Feb. 2011), pp. 764–785. ISSN: 1557-9654. DOI: 10.1109/TIT.2010.2094817.

[BMN19] Raphaël Berthier, Andrea Montanari, and Phan-Minh Nguyen. “State evolution for approximate message passing with non-separable functions”. In: *Information and Inference: A Journal of the IMA* (Jan. 2019). iay021. ISSN: 2049-8772. DOI: 10.1093/imaiai/iay021. eprint: http://oup.prod.sis.lan/imaiai/advance-article-pdf/doi/10.1093/imaiai/iay021/27503802/iay021.pdf. URL: https://doi.org/10.1093/imaiai/iay021.

[DMM09] David L. Donoho, Arian Maleki, and Andrea Montanari. “Message-passing algorithms for compressed sensing”. In: *Proceedings of the National Academy of Sciences* 106.45 (2009), pp. 18914–18919. ISSN: 0027-8424. DOI: 10.1073/pnas.0909892106. eprint: https://www.pnas.org/content/106/45/18914.full.pdf. URL: https://www.pnas.org/content/106/45/18914.

[Don06] David L. Donoho. “Compressed sensing”. In: *IEEE Transactions on Information Theory* 52.4 (Apr. 2006), pp. 1289–1306. ISSN: 1557-9654. DOI: 10.1109/TIT.2006.871582.

[GP19] S. K. Gunasheela and H. S. Prasantha. “Compressed Sensing for Image Compression: Survey of Algorithms”. In: *Emerging Research in Computing, Information, Communication and Applications*. Ed. by N. R. Shetty et al. Singapore: Springer Singapore, 2019, pp. 507–517. ISBN: 978-981-13-6001-5. DOI: 10.1007/978-981-13-6001-5_42.

[Mal18] Arian Maleki. “Message passing and approximate message passing”. Columbia STAT GR6104 - Computational Statistics. 2018. URL: http://www.stat.columbia.edu/~liam/teaching/compstat-spr19/AMPlecture.pdf.

[RFS18] Sundeep Rangan, Alyson K. Fletcher, and Phil Schniter. “Approximate Message Passing Tutorial”. Institut d’Études Scientifiques de Cargèse Statistical Physics And Machine Learning Workshop. 2018. URL: https://krzakala.github.io/cargese.io/AMP_Tutorial_18.pdf.

[Rus18] Cynthia Rush. “Finite Sample Analysis of AMP”. Institut d’Études Scientifiques de Cargèse Statistical Physics And Machine Learning Workshop. 2018. URL: https://krzakala.github.io/cargese.io/Cynthia.pdf.

[Wai08] Martin J. Wainwright. “Message passing and approximate message passing”. 2008 International Conference on Machine Learning Tutorial. 2008. URL: https://people.eecs.berkeley.edu/~wainwrig/icml08/tutorial_icml08.html.