Introduction to Approximate Message Passing with Application to LASSO

Yuze Zhou

Apr.30

Review of Iterated Algorithms

- The optimization problem we're dealing with is:

$$\hat{x} = \arg\min_{x} g(x) + \phi(x)$$

- **Proximal Operator:**

$$Prox_f(y, \alpha) = \arg\min_{x} f(x) + \frac{1}{2\alpha} \|y - x\|_2^2$$

Review on ISTA

- The ISTA procedure for solving the optimization problem is:

$$r_k = x_k - \alpha \nabla g(x_k)$$
$$x_{k+1} = Prox_\phi(r_k, \alpha)$$

Review on ADMM

- The original optimization problem could be relaxed to:

$$\min_x g(x) + \phi(y)$$
$$s.t. \quad x = y$$

- The augmented Lagrangian is:

$$L(x, y; s, \alpha) = g(x) + \phi(y) + \alpha s^T (x - y) + \frac{\alpha}{2} \|x - y\|_2^2$$

- The **ADMM** procedure goes as followed:

$$x_{k+1} = Prox_g(y_k - s_k, \alpha^{-1})$$
$$y_{k+1} = Prox_\phi(x_{k+1} + s_k, \alpha^{-1})$$
$$s_{k+1} = s_k + x_{k+1} - y_{k+1}$$

Notations

- The LASSO problem here is formulated as:

$$\hat{x} = \arg\min_{x} \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1$$

  where $y \in \mathbb{R}^n$ and $x \in \mathbb{R}^N$
- Soft threshold function is defined as, let $h(x) = \|x\|_1$:

$$\eta(x, b) = sign(x)(|x| - b)_+ = Prox_h(x, b)$$

ISTA on LASSO

- Given the soft-thresholding function aforementioned, the **ISTA** procedure for LASSO is:

$$
\begin{aligned}
r_k &= x_k + \alpha A^{\tau}(y - Ax_k) \\
x_{k+1} &= \arg\min_x \quad \lambda \|x\|_1 + \frac{1}{2\alpha}\|x - r_k\|_2^2 \\
&= \eta(r_k, \lambda\alpha)
\end{aligned}
$$

ADMM on LASSO

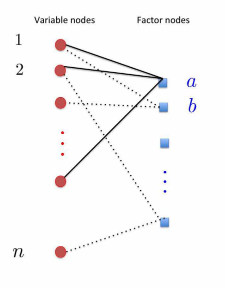
- Let $g(x) = \frac{1}{2}\|y - Ax\|_2^2$ and $h(x) = \lambda\|x\|_1$, the **ADMM** procedure for LASSO is:

$$\begin{aligned}
x_{k+1} &= Prox_g(z_k - s_k, \alpha^{-1}) \\
&= (A^\tau A + \alpha I)^{-1}(A^\tau y + \alpha(z_k - s_k)) \\
z_{k+1} &= Prox_h(x_{k+1} + s_k, \alpha^{-1}) \\
&= \eta(x_{k+1} + s_k; \frac{\lambda}{\alpha}) \\
s_{k+1} &= s_k + x_{k+1} - y_{k+1}
\end{aligned}$$

Intuition and Set-up

- The LASSO problem here is formulated as:
$$\hat{s} = \arg\min_s \tfrac{1}{2}\|y - As\|_2^2 + \lambda\|s\|_1$$
- From simulated annealing, set up the probability density for x:
$$f(s) = \tfrac{1}{Z_\beta} exp(-\beta[\tfrac{1}{2}\|y - As\|_2^2 + \lambda\|s\|_1])$$
- $f(s)$ is most dense around $\hat{s}$ and will concentrate on $\hat{s}$ as $\beta \to \infty$

Belief Propagation on Factor Graph



- $f(x) = \frac{1}{Z} \prod_{j=1}^{m} f_j(x_{S_j}) \prod_{i=1}^{n} g_i(x_i)$
- j are the factor nodes, i are the variable nodes, S_j are the set of variable nodes which are neighbours of factor node j

Belief Propagation on Factor Graph

- $\partial(x_i)$ be the set of factor nodes that are neighbours of variable x_i
- **Belief from factor node j to variable node i:**

$$v_{j\to i}(x_i) \simeq \int_{x_{S_j}/x_i} f_j(x_{S_j}) \prod_{i'\neq i} d\hat{v}_{i'\to j}(x_{i'})$$

- **Belief from variable node i to factor node j:**

$$\hat{v}_{i\to j}(x_i) \simeq g_i(x_i) \prod_{b\in\partial(x_i)/j} v_{b\to i}(x_i)$$

- The products of the beliefs would be the marginal distribution.

Belief Propagation on Factor Graph

- In most cases, bipartite graphs are loopy, beliefs could not be obtained easily.
- **Belief from factor node j to variable node i:**

$$v_{j\to i}^{t+1}(x_i) \simeq \int_{x_{S_j}/x_i} f_j(x_{S_j}) \prod_{i'\neq i} d\hat{v}_{i'\to j}^{t}(x_{i'})$$

- **Belief from variable node i to factor node j:**

$$\hat{v}_{i\to j}^{t+1}(x_i) \simeq g_i(x_i) \prod_{b\in\partial(x_i)/j} v_{b\to i}^{t}(x_i)$$

Approximate Message Passing for LASSO

- Density Set-up for LASSO:

$$f(s) \propto \prod_{i=1}^{N} exp(-\beta\lambda|s_i|) \prod_{a=1}^{n} exp(-\frac{\beta}{2}(y_a - (As)_a)^2)$$

- **Belief from factor node** a **to variable node** i:

$$v_{a\to i}^{t+1}(x_i) \simeq \int exp(-\frac{\beta}{2}(y_a - (As)_a)^2) \prod_{j\neq i} d\hat{v}_{j\to a}^{t}(x_j)$$

- **Belief from variable node** i **to factor node** j:

$$\hat{v}_{i\to a}^{t+1}(x_i) \simeq exp(-\beta\lambda|s_i|) \prod_{b\in\partial(x_i)/a} v_{b\to i}^{t}(x_i)$$

Approximate Message Passing for LASSO

- Let $x^t_{j\to a}$ and $\tau^t_{j\to a}$ be the mean and variance of $\hat{v}^t_{j\to a}$
- Let $z^t_{a\to i} = y_a - \sum_{j\neq i} A_{aj} x^t_{j\to a}$ and $\hat{\tau}^t_{a\to i} = \sum_{j\neq i} A^2_{aj}\tau^t_{j\to a}$
- Define the density function $\phi^t_{a\to i}(s_i)$ as:

$$\phi^t_{a\to i}(s_i) = \sqrt{\frac{\beta A^2_{ai}}{2\pi(1+\hat{\tau}^t_{a\to i})}} exp(-\frac{\beta}{2(1+\hat{\tau}^t_{a\to i})}(A_{ai} - z^t_{a\to i})^2)$$

Theorem (Donoho.D.L et.al. 2009)

There exists a constant C_t such that the KS distance between $\phi^t_{a\to i}$ and $v^t_{a\to i}$ satisfies:

$$\sup_{s_i\in\mathbb{R}} |\phi^t_{a\to i}(s_i) - v^t_{a\to i}(s_i)| \leq \frac{C_t}{N(\hat{\tau}^t_{a\to i})^{3/2}}$$

Approximate Message Passing for LASSO

- Define the family of distributions $f_\beta(s;x,b)$ as:

$$f(s;x,b) \propto exp(-\beta|s| - \frac{\beta}{2b}(s-x)^2)$$

- $F_\beta(x;b)$ and $G_\beta(x;b)$ are defined as the mean and variance of the distribution above
- The bounds on the KS distance between $\phi^t_{a\to i}(s_i)$ and $v^t_{a\to i}(s_i)$ suggests that we could replace the mean and variance of the latter distribution by the previous one, which gives rise to the following message passing algorithm:

Approximate Message Passing for LASSO

Algorithm 1: Message Passing for LASSO

for $0 \le t \le n-1$ **do**

$$x_{i\to a}^{t+1} = \frac{1}{\lambda} F_\beta(\lambda \sum_{b\neq a} A_{bi} z_{b\to i}^t; \lambda^2(1+\hat{\tau}^t));$$

$$\hat{\tau}^{t+1} = \frac{\beta}{\lambda^2 n} \sum_{i=1}^{N} G_\beta(\lambda \sum_{b} A_{bi} z_{b\to i}^t; \lambda^2(1+\hat{\tau}^t));$$

$$z_{a\to i}^{t+1} = y_a - \sum_{j\neq i} A_{aj} x_{j\to a}^{t+1}$$

end

Approximate Message Passing for LASSO

- Recall the definition for the soft-thresholding function $\eta(x, b)$:

$$\lim_{\beta \to \infty} F_\beta(x, b) = \eta(x, b)$$
$$\lim_{\beta \to \infty} \beta G_\beta(x, b) = b\eta^{'}(x, b)$$

- With these two limits
  above, let $\delta = n/N$ we could improve the message passing algorithm as:

Algorithm 2: Message Passing for LASSO with large β limits

for $0 \leq t \leq n-1$ **do**

$$x_{i\to a}^{t+1} = \eta(\lambda \sum_{b\neq a} A_{bi} z_{b\to i}^t; \lambda(1+\hat{\tau}^t));$$
$$\hat{\tau}^{t+1} = \frac{1+\hat{\tau}^t}{N\delta} \sum_{i=1}^N \eta^{'}(\lambda \sum_b A_{bi} z_{b\to i}^t; \lambda(1+\hat{\tau}^t));$$
$$z_{a\to i}^{t+1} = y_a - \sum_{j\neq i} A_{aj} x_{j\to a}^{t+1}$$

end

Approximate Message Passing for LASSO

- The Message Passing Algorithm is still computationally expensive, if we further assume that:

$$
\begin{aligned}
x_{i\to a}^t &= x_i^t + \delta x_{i\to a}^t + \mathcal{O}(1/N) \\
z_{a\to i}^t &= z_a^t + \delta z_{a\to i}^t + \mathcal{O}(1/N)
\end{aligned}
$$

  where $\delta x_{i\to a}^t$ and $\delta z_{a\to i}^t$ are both of order $\mathcal{O}(1/\sqrt{N})$
- $x_{i\to a}^t$ and $z_{a\to i}^t$ could be approximate by one step from x_i^t and z_a^t

Approximate Message Passing for LASSO

Let $\gamma^t = \lambda \hat{\tau}^t$ and $< \cdot >$ be the averaging operator:

Algorithm 3: Approximate Message Passing for LASSO

for $0 \leq t \leq n-1$ **do**

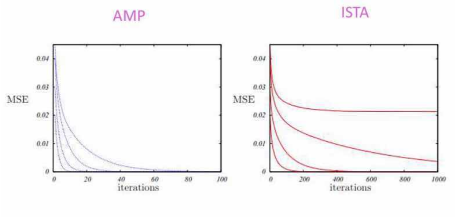
$$x^{t+1} = \eta(x^t + A^* z^t; \lambda + \gamma^t);$$

$$\gamma^{t+1} = \frac{\lambda + \gamma^t}{\delta} < \eta^{'}(Az^t + x^t; \lambda + \gamma^t) >;$$

$$z^{t+1} = y - Ax^{t+1} + \frac{1}{\delta} z^t < \eta^{'}(Az^t + x^t; \lambda + \gamma^t) >$$

end

Empirical Results



- For the synthesized dataset $N = 8000$, $n = 1600$, and A_{ai} is sampled from $N(0, 1/N)$, AMP converge with much fewer iterations compared to ISTA

Future Work

- AMP algorithm is unstable
- No theoretical guarantees yet
- Genralization to other sparse learning problems

References I

- Aarts.E, Korst.J(1988)
  Simulated Annealing and Boltzmann Machine
- Donoho.D.L, Maleki.A & Montanari.A(2009)
  Message Passing Algorithms for Compressed Sensing: I.Motivation and Construction
- Bayati.M, Donoho.D.L., Javanmard.A, Johnstone.I, Lelarge.M, Maleki.A & Montanari.A(2012)
  Approxiamte Message Passing
- Rangan.S, Fletcher.A.K & Schniter.P(2018)
  Approximate Message Passing Tutorial
- Donoho.D.L, Maleki.A. & Montanari.A(2011)
  How to Design Approximate Message Passing Algorithms for Compressed Sensing
- Freeman.B & Torralba.A(2010)
  Graphical Models and Belief Propagation