Soo Min Kwon

Towards Understanding Spectral Initialization for Phase Retrieval

The purpose of this set of notes is to try and unpack what's going on in the initialization stage of Truncated Wirtinger Flow (TWF). TWF is an algorithm that optimally solves the phase retrieval problem. Recall that in phase retrieval, the objective is to recover $\mathbf{x}$ given sampling matrices $\mathbf{a}_i$ and

$$y_i = \langle \mathbf{a}_i, \mathbf{x} \rangle^2,\ i = 1, \ldots, m. \tag{1}$$

Due to the non-convex nature of this problem, we need to initialize $\mathbf{x}^0$ via a spectral method to guarantee that the initial $\mathbf{x}_0$ is close to the true $\mathbf{x}^*$ with high probability. To do this, we will assume the Gaussian measurement setting, where $\mathbf{a}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathbf{I})$. In TWF, we take the top eigenvector of the matrix

$$\mathbf{Y} = \sum_{i=1}^{m} y_i^2 \mathbf{a}_i \mathbf{a}_i^\top \mathbf{1}_{\{|y_i|^2 \leq \alpha^2 \lambda^2\}}, \tag{2}$$

where

$$\lambda = \sqrt{\frac{1}{m} \sum_{i=1}^{m} y_i}. \tag{3}$$

The truncation is here is saying to only take the measurements of y_i that are not too far away from some magnitude of its mean. We will show why this truncation is useful from a mathematical point of view. To show this, let's ignore the truncation in $\mathbf{Y}$ and call the un-truncated matrix $\mathbf{Z}$. Why is this matrix $\mathbf{Z}$ important? Well, under the Gaussian measurement assumption, a moments calculation[1] shows that

$$\mathbf{Z}^* = \mathbb{E}[\mathbf{Z}] = 2\mathbf{x}^* \mathbf{x}^{*\top} + \|\mathbf{x}^*\|_2^2 \mathbf{I}. \tag{4}$$

We can show that the top eigenvalue of $\mathbf{Z}^*$ is

$$\begin{aligned}
\mathbf{Z}^* \mathbf{u} &= \lambda \mathbf{u} && (5)\\
(2\mathbf{x}^* \mathbf{x}^{*\top} + \|\mathbf{x}^*\|_2^2 \mathbf{I}) \mathbf{u} &= \lambda \mathbf{u} && (6)\\
\mathbf{x}^{*\top} (2\mathbf{x}^* \mathbf{x}^{*\top} + \|\mathbf{x}^*\|_2^2 \mathbf{I}) \mathbf{u} &= \mathbf{x}^{*\top} \lambda \mathbf{u} && (7)\\
2\mathbf{x}^{*\top} \mathbf{x}^* \mathbf{x}^{*\top} \mathbf{u} + \mathbf{x}^{*\top} \mathbf{x}^{*\top} \mathbf{x}^* \mathbf{u} &= \lambda \mathbf{x}^{*\top} \mathbf{u} && (8)\\
2\mathbf{x}^{*\top} \mathbf{x}^* (\mathbf{x}^{*\top} \mathbf{u}) + \mathbf{x}^{*\top} \mathbf{x}^* (\mathbf{x}^{*\top} \mathbf{u}) &= \lambda (\mathbf{x}^{*\top} \mathbf{u}) && (9)\\
3\|\mathbf{x}^*\|_2^2 &= \lambda. && (10)
\end{aligned}$$

Clearly, it follows that the leading eigenvector with $\lambda = 3\|\mathbf{x}^*\|_2^2$ is equivalent to $\mathbf{u}_1 = \frac{\mathbf{x}^*}{\|\mathbf{x}^*\|_2}$. Note that by the law of large numbers,

$$\frac{1}{m} \sum_{i=1}^{m} y_i \implies \mathbb{E}[y_i], \tag{11}$$

and $\mathbb{E}[(\mathbf{a}^\top \mathbf{x}^*)^2] = \|\mathbf{x}^*\|_2^2$ (also by a moments calculation). Thus, scaling the leading eigenvector $\mathbf{u}_1$ with λ gives us the initial spectral estimate

$$\mathbf{x}^0 = \left(\frac{1}{m} \sum_{i=1}^{m} y_i \right)^{1/2} \mathbf{u}_1. \tag{12}$$

[1] We can see this by looking at every (i, j)-th element of $\mathbf{Z}$.

Okay, so this is nice so far. Netrapalli et al. [?] shows that bounding $\|\mathbf{Z} - \mathbf{Z}^*\|_{\mathsf{F}}$ yields that $\mathbf{x}$ is also close to the true $\mathbf{x}^*$.[2] Now, we want to show what the truncation buys us. To see this, we need to look at the lower bound of $\|\mathbf{Z} - \mathbf{Z}^*\|_{\mathsf{F}}$. Some algebra shows that

$$\begin{aligned}
\|\mathbf{Z}\| &\geq \frac{\mathbf{a}_j^\top \mathbf{Z} \mathbf{a}_j}{\|\mathbf{a}_j\|_2^2} \quad \text{(By Rayleigh–Ritz)} && (13)\\
&= \frac{1}{m} \sum_{i=1}^{m} y_i \frac{(\mathbf{a}_i^\top \mathbf{a}_j)^2}{\|\mathbf{a}_j\|_2^2} \quad \text{(By the definition of } \mathbf{Z}\text{)} && (14)\\
&\geq \frac{1}{m} \left(\max_i y_i \right) \|\mathbf{a}_{i^*}\|_2^2, \quad \text{(By replacing the sum with the max)} && (15)
\end{aligned}$$

where $\mathbf{a}_{i^*}$ refers to the index i that maximizes y_i. Note that under the Gaussian measurement assumption, $\frac{y_i}{\|\mathbf{x}^*\|_2^2} = \frac{(\mathbf{a}_i^\top \mathbf{x})^2}{\|\mathbf{x}^*\|_2^2}$ forms a collection of $\mathcal{X}^2$ random variables with 1 degree of freedom! Then, searching the internet for concentration of the maximum of sub-exponential random variables, we have that

$$\begin{aligned}
\max_{1 \leq i \leq m} \frac{y_i}{\|\mathbf{x}^*\|_2^2} &\approx \sqrt{2 \log m} + 2 \log m && (16)\\
\implies \max_{1 \leq i \leq m} y_i &= (\sqrt{2 \log m} + 2 \log m) \|\mathbf{x}^*\|_2^2 && (17)
\end{aligned}$$

and

$$\|\mathbf{a}_i\|_2^2 \approx n. \tag{18}$$

Plugging these results into the lower bound of $\|\mathbf{Z}\|$, we have

$$\|\mathbf{Z}\| \geq \frac{n(\sqrt{2 \log m} + 2 \log m)}{m} \|\mathbf{x}^*\|_2^2. \tag{19}$$

Some more algebra shows us that

$$\begin{aligned}
\|\mathbf{Z} - \mathbf{Z}^*\| &\geq \|\mathbf{Z}\| - \|\mathbf{Z}^*\| && (20)\\
&= \|\mathbf{Z}\| - 3\|\mathbf{x}^*\|_2^2 && (21)\\
&\geq \frac{n \log m}{m} \|\mathbf{x}^*\|_2^2 - 3\|\mathbf{x}^*\|_2^2 && (22)\\
&\gg \|\mathbf{Z}^*\|. && (23)
\end{aligned}$$

If we have $m \ll n \log m$, which is very possible in the under-determined setting, this means that there exists a top eigenvalue (and hence eigenvector) that is much closer to $\mathbf{Z}$ than $\mathbf{Z}^*$ and that the deviation between these two matrices is not well-controlled. This suggests a natural remedy, which is to truncate the values of y_i that are much larger than its mean. This would give us that $\|\mathbf{Y} - \mathbf{Z}^*\|$ is much more well-controlled.

[2] We will show this in the next set of notes.