Angular Synchronization Using Phase Transitions in Spiked Covariance Matrices

Brandon Tran

May 4, 2016

Abstract

We present an exposition on the angular synchronization problem. We use a phase transition result of the eigenvalues of spiked covariance matrices from random matrix theory to present a solution to this problem. Then, we present further directions one can take, including using the theory of semidefinite programming and approximate message passing.

1 Introduction

The angular synchronization problem is to recover n angles $\theta_1, \ldots, \theta_n$ from $m \leq \binom{n}{2}$ noisy measurements δ_{ij} of their offsets $\theta_i - \theta_j$. From a learning perspective, we expect the number of measurements m to be much less than the set of all measurements. By noisy, we mean that with probability p, we receive the actual offset, and with probability $1-p$, we receive a measurement uniform on the interval $[0, 2\pi)$. If we let $G = (V, E)$ be the graph on n vertices with edge set equal to the offset measurements, we note that if G is disconnected, we cannot hope to recover the angles. And even if G were connected, we can only recover the angles up to some global additive constant.

Before the work of [5], the most common approaches to solving synchronization problems was least squares or using a maximum likelihood estimator. But as we will explain in detail later, these do not work for our particular problems. The author introduces a new approach using spectral methods. In particular, he pulls from the theory of random matrices.

Consider a Wigner matrix M, with entries having second moment equal to 1. Then we expect its eigenvalues to follow the semicircle law in $(-2\sqrt{n}, 2\sqrt{n})$ [6]. We can ask the following question: if we add a rank one spike to M, when will its eigenvalue distribution no longer follow this semicircle law? In particular, when does $\frac{\lambda}{n} zz^* + \frac{1}{\sqrt{n}} M$ have an eigenvalue outside $(-2, 2)$? The answer turns out to be whenever $\lambda > 1$ [2] [3]. From this result, we can deduce a solution to the angular synchronization problem.

It turns out that this theory can be applied to other synchronization problems. Here, we can view the real numbers modulo 2π as a group G. Then we

could ask the question: given a general group G and offset measurements $g_i g_j^{-1}$, how can we recover the group G? Whenever this group has a complex or real representation, we can construct a Hermitian matrix whose (i, j) entry is the matrix representation of the measurement, if such a measurement was given. This question when G is $SO(3)$ arises naturally in cryo-electron microscopy, NMR spectroscopy, and localization of sensor networks.

Our exposition will first briefly discuss the random matrix theory behind the aforementioned phase transition in Section 2. Then, we will provide the solution in [5] to the angular synchronization problem in Section 3. Finally, we will conclude in Section 4 with applications and further directions.

2 Spiked Covariance Matrices

In our exposition, we will discuss [2] and only note that their work is extended in [3]. First, we state the main result from that paper in the form we need. We note that the paper does not assume a square matrix and so we are looking at eigenvalues of AA^T.

Theorem 2.1. *Let λ_1 be the largest eigenvalue of the sample covariance matrix constructed from M i.i.d. complex Gaussian sample vectors of N variables. Let $\ell_1, \ldots, \ell_N$ denote the eigenvalues of the covariance matrix of the samples. Suppose that for a fixed r, $\ell_{r+1} = \cdots = \ell_N = 1$. As $M, N \to \infty$, with $M/N = \gamma^2$, the following holds:*

1. *When for some $0 \leq k \leq r, \ell_1 = \cdots = \ell_k = 1 + \gamma^{-1}$, and $\ell_{k+1}, \ldots \ell_r$ are in a compact subset of $(0, 1 + \gamma^{-1})$, $\lambda_1 \to (1 + \gamma^{-1})^2$ in probability.*
2. *When for some $1 \leq k \leq r, \ell_1 = \cdots = \ell_k > 1 + \gamma^{-1}$ and $\ell_{k+1}, \ldots \ell_r$ are in a compact subset of $(0, \ell_1), \lambda_1 \to \ell_1(1 + \frac{\gamma^{-1}}{\ell_1 - 1})$ in probability.*

We will not provide an in-depth proof here, but will instead only outline the ideas. We start by writing the distribution function of the largest eigenvalue λ_1 as a Fredholm determinant.

$$P(\lambda_1 \leq \xi) = det(1 - K_{M,N}|_{(\xi,\infty)}),$$

where $K_{M,N}|_{(\xi,\infty)}$ is the operator with kernel

$$K_{M,N}(x, y) = \frac{M}{(2\pi i)^2} \int_\Gamma dz \int_\Sigma dw e^{-xM(z-q)+yM(w-q)} \frac{1}{w - z} (\frac{z}{w})^M \prod_{k=1}^{N} \frac{\pi_k - w}{\pi_k - z}$$

Here, Γ is any simple contour enclosing $\pi_1, \ldots, \pi_N$ lying right of q, and Σ is any

simple contour enclosing 0. Now we split this up as

$$K_{M,N}(x,y) = \int_0^\infty H(x+a)J(y+a)da$$
$$H(x+a) = \frac{M}{2\pi}\int_\Gamma e^{-(x+a)M(z-q)}z^M \prod_{k=1}^N \frac{1}{\pi_k - z}dz$$
$$J(y+a) = \frac{M}{2\pi}\int_\Sigma e^{(y+a)M(w-q)}w^{-M} \prod_{k=1}^N (\pi_k - w)dw$$

2.1 Basic Ideas for Analysis

We will only give the ideas for the first part of the theorem, since it will suffice to display the kind of arguments used. Recall that π_i^{-1} for i up to k is $1+\gamma^{-1}$, and up to r is in some compact subset. Now if we write ξ as $\mu + \frac{\nu\xi'}{M^\alpha}$ some manipulations show that we need to evaluate

$$H(u) = \frac{\nu M^{1-\alpha}}{2\pi}\int_\Gamma e^{-\nu M^{1-\alpha}u(z-q)}e^{-M\mu(z-q)}\frac{z^M}{(1-z)^{N-r}}\prod_{\ell=1}^r \frac{1}{\pi_\ell - z}dz$$
$$J(v) = \frac{\nu M^{1-\alpha}}{2\pi}\int_\Sigma e^{\nu M^{1-\alpha}v(w-q)}e^{M\mu(w-q)}\frac{(1-w)^{N-r}}{w^M}\prod_{\ell=1}^r (\pi_\ell - w)dw$$

We now take $\alpha = 2/3, \mu = \left(\frac{1+\gamma}{\gamma}\right)^2, \nu = \frac{(1+\gamma)^{4/3}}{\gamma}$. The following analysis will make it clear why these were chosen so.

The term we care about is $e^{Mf(z)}$ (both in this case and the case where some eigenvalues are larger), where

$$f(z) = -\mu(z-q) + \log(z) - \frac{1}{\gamma^2}\log(1-z).$$

To compute the critical point, we need

$$f'(z) = -\mu + \frac{1}{z} - \frac{1}{\gamma^2(z-1)}.$$

We see that our choice of μ creates a double root at $z_0 = \frac{\gamma}{\gamma+1}$. Computing more derivatives, we see that

$$f^{(3)}(z_0) = 2\nu^3 > 0,$$

which explains the choice of ν. This analysis implies that the steepest-descent curve of f comes to z_0 wiht angle $\pm\pi/3$ to the real axis. The idea, then, is to move Γ to approach z_0 with such an angle so that the main contribution to the integral comes from a contour near z_0. The new contours look like

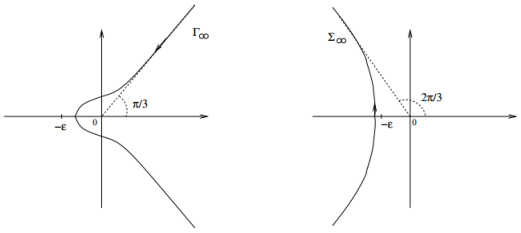

Figure 1: New Contours

To sum up, for H, we now have

$$H(u) \sim \frac{\nu M^{1/3}}{2\pi g(z_0)} \int_\Gamma e^{-\nu M^{1/3} u(z-q)} e^{M(f(z_0)+\frac{f^{(3)}(z_0)}{3!}(z-z_0)^3)} \frac{1}{(z_0-z)^k} dz$$

with

$$g(z) = \frac{1}{(1-z)^r} \prod_{k+1}^{r} (\pi_\ell - z).$$

Thus, we expect

$$H(u) \sim \frac{(-\nu M^{1/3})^k e^{Mf(z_0)}}{2\pi g(z_0)} e^{-\nu M^{1/3} u(z_0-q)} \int_{\Gamma_\infty} e^{-ua+a^3/3} a^{-k} da$$

A similar argument gives a similar expression for J. Then we can see that this converges, after multiplying by a constant factor, to

$$H_\infty(u) = \frac{e^{-\varepsilon u}}{2\pi} \int_{\Gamma_\infty} e^{-ua+a^3/3} a^{-k} da$$

What this gives us is that

$$P\left((\lambda_1 - (1+\gamma^{-1})^2) \frac{\gamma}{(1+\gamma)^{4/3}} M^{2/3} \leq x \right)$$

converges to the Fredholm determinant of an operator with kernel

$$\int_0^\infty H_\infty(x+u+y) J_\infty(x+v+y) dy$$

It turns out this last expression can be written as an Airy kernel, but our purposes, it only matters that this exists.

This is very much just a skeleton outline. Refer to [2] for details. But the idea is that we construct a contour integral and want to use a steepest-descent argument. So we plug in values that allow for this, and the threshold is at our desired values.

Also, we remind the reader that the work of [3] extends this to Wigner matrices. With this in mind, we move to applying these results to angular synchronization.

3 Angular Synchronization

First, we formally define the angular synchronization problem.

Problem 3.1. *Suppose there are n unknown angles $\theta_1, \ldots, \theta_n$. We are given m measurements δ_{ij} that are either equal to $\theta_i - \theta_j$ or are uniformly chosen from $[0, 2\pi)$. We wish to produce an algorithm that with high probability recovers θ_i.*

Note that typically, the good measurements should not be exactly $\theta_i - \theta_j$ but rather normally distributed around the difference with some variance σ^2. It is easy to see that the argument for proof of correctness still holds in this case, though.

We briefly mention some standard approaches that don't work for this problem.

3.1 Previous Approaches

The first method we mention is least squares. In applications such as time synchronization and surface resconstruction in computer vision, least squares is a common approach when one needs a gloally consistent integration method to not accumulate too much error. This method does well when the offset measurements have a small Gaussian additive error, and they can be efficiently computed and analyzed in terms of graph Laplacians.

To use this, we write an overdetermined system of linear equations

$$\theta_i - \theta_j = \delta_{ij} \pmod{2\pi}.$$

Writing $z_i = e^{i\theta_i}$, this is

$$z_i - e_{i\delta_{ij}} z_j = 0.$$

To avoid the trivial solution, we force $z_1 = 1$, which we can do since we can only recover up to additive constant. Then, we look for a solution with minimal ℓ_2-norm residual. However, in our model, we expect the error to be dominated by the outlier equations, and so least squares will not work well if the number of bad equations is large.

The next idea to try is maximum likelihood. Here, we wish to find the set of angles that satisfies as many equations as possible. Define the self consistency error as the number of equations not satisfied.

$$SCE(\theta_1, \ldots, \theta_n) = \#\{(i,j) \in E : \theta_i - \theta_j \neq \delta_{ij} \pmod{2\pi}\}.$$

In most situations, due to discretization, even good measurements will contain some small error, so instead we could use

$$SCE_f(\theta_1, \ldots, \theta_n) = \sum_{(i,j)\in E} f(\theta_i - \theta_j - \delta_{ij}),$$

where f is a smooth periodic function with $f(0) = 0$ and $f(\theta) = 1$ when $|\theta|$ is larger than the discretization error. Minimizing SCE_f turns out to be equivalent to maximizing the log likelihood with a different probabilistic error model. The problem, though, is that typically this is a non-convex prolem, and so is difficult to carry out.

Hence, we use a new estimator from spectral theory.

3.2 The Eigenvector Method

First, we form the following $n \times n$ matrix H with entries

$$H_{ij} = \begin{cases} e^{i\delta_{ij}} & \text{if } (i,j) \in E \\ 0 & \text{else.} \end{cases}$$

The idea behind the method is as follows. Consider the maximization problem

$$\max_{\theta_1,\ldots,\theta_n \in [0,2\pi)} \sum_{i,j=1}^{n} e^{-i\theta_i} H_{ij} e^{i\theta_j}.$$

It's easy to see that for the correct angles, good edges contribute 1, and so the total contribution of good edges is m_{good}. The contributions from bad edges are uniform on the unit circle, and so viewing this as a random walk, we get an expected contribution of $O(\sqrt{m_{bad}})$.

This maximization problem has the same problems as our maximum likelihood estimator, which is that it is non convex. Thus, we must relax it to

$$\max_{\sum_{i=1}^{n} |z_i|^2 = n} \sum_{i,j=1}^{n} z_i^* H_{ij} z_j.$$

It's easy to see that the solution to this is just the normalized top eigenvector of H. Once we have a solution, we can take $e^{i\theta_i} = \frac{z_i}{|z_i|}$. This problem has an efficient solution, since finding the top eigenvector is fast by the power iteration method. Of course, we have to show that the relaxation gives a good solution to the original problem.

3.3 Analysis in the Complete Graph Case

We will first provide an analysis of why our method works in the case that all $\binom{n}{2}$ measurements are given. Also, we assume that whether an edge is good or bad is decided with probability p or $1-p$. Thus, H_{ij} is $e^{i\theta_i - i\theta_j}$ with probability p and uniformly chosen at random from the circle with probability $1-p$. We also let the diagonal entries of H be p.

Now, we can compute the expected value

$$\mathbf{E} H_{ij} = p e^{i\theta_i - i\theta_j}$$

so that

$$\mathbf{E}H = npzz^*,$$

where z is our vector of true angles with entries $z_i = \frac{1}{\sqrt{n}}e^{i\theta_i}$. Since we wish to appeal to spiked covariance results, it is natural to write

$$H = npzz^* + R,$$

where R is the matrix of the difference of H from its expectation, with entries

$$R_{ij} = \begin{cases} 0 & \text{if } i = j \\ (1-p)e^{i(\theta_i - \theta_j)} & \text{w.p. } p \\ e^{i\phi} - pe^{i(\theta_i - \theta_j)} & \text{w.p.} 1-p \end{cases},$$

where ϕ is uniform in $[0, 2\pi)$. We can compute the variance and see, then, that $R/\sqrt{n(1-p^2)}$ has eigenvalues following the semicircle law in $(-2, 2)$. Thus, we can apply the results from the previous section to see that as long as

$$np > \sqrt{n(1-p^2)},$$

the largest eigenvalue of H will be outside the support of the semicircle law. In fact, we know that

$$\begin{aligned} \lambda_1(H) &\sim N(\mu, \sigma^2) \\ \mu &= \frac{np}{\sqrt{1-p^2}} + \frac{\sqrt{1-p^2}}{p} \\ \sigma^2 &= \frac{(n+1)p^2 - 1}{np^2}(1-p^2) \end{aligned}$$

We need a little more than just that the eigenvalue pops out, though. In particular, we also need a correlation between the largest eigenvector v_1 and z. To this end,

$$\begin{aligned} \lambda_1(H)v_1 &= (npzz^* + R)v_1 \\ \Rightarrow \lambda_1(H) &= np| < z, v_1 > |^2 + v_1^* R v_1 \\ \Rightarrow | < z, v_1 > |^2 &\geq \frac{\lambda_1(H) - \lambda_1(R)}{np} \end{aligned}$$

We want this to be better than random, i.e. greater than $\frac{1}{n}$. The threshold probability, then, is $p_c = \frac{1}{\sqrt{n}}$. It is straightforward from a Taylor expansion and approximation that as np^2 gets large, the correlation between z and v_1 increases.

3.4 General Analysis

Now, we are ready to move to the general case, the idea being to generalize the decomposition in the complete graph case. We separate our edge set into

E_{good} and E_{bad}. Let A be the adjacency matrix for E_{good}, and suppose its decomposition into eigenvectors is $A = \sum_{i=1}^{n} \lambda_i \psi_i \psi_i^T$.

Let Z be the diagonal matrix with entries $Z_{ii} = e^{i\theta_i}$, and $B = ZAZ^*$. It has the same eigenvalues as A, and its eigenvectors are $\phi_i = Z\psi_i$. Moreover, it's easy to see that

$$B_{ij} = \begin{cases} e^{i(\theta_i - \theta_j)} & \text{if } (i,j) \in E_{good} \\ 0 & \text{else.} \end{cases}$$

We decompose $H = B + R$ where R now has entries

$$R_{ij} = \begin{cases} e^{i\delta_{ij}} & \text{if } (i,j) \in E_{bad} \\ 0 & \text{else.} \end{cases}$$

In this case, the δ_{ij} are uniformly chosen at random from $[0, 2\pi)$. Now, the Perron-Frobenius theorem states that the entries of ψ_1 are all positive, and this implies that $\frac{\phi_i}{|\phi_i|}$ gives us the true angles as $e^{i\theta_i}$. Thus, what we need to show is that the top eigenvector of H corresponds with that of B.

We turn to look at the matrix R. Suppose that there are m_{bad} edges and they were chosen at random from the edges not in E_{good}. Then R has $2m_{bad}$ entries, with an average of $2m_{bad}/n$ per row. These entries have 0 mean and variance 1. From [4], with probability 1,

$$\limsup_{n\to\infty} \frac{\sqrt{n}}{\sqrt{2m_{bad}}} \leq 2$$

so we can approximate

$$\lambda_1(R) \approx 2\frac{\sqrt{2m_{bad}}}{\sqrt{n}}$$

Note, however, that B is not a rank one perturbation. Thus, what we need is the spectral gap $\Delta_{good} = \lambda_1(A) - \lambda_2(A)$ satisfies

$$\Delta_{good} > \frac{\sqrt{2m_{bad}}}{\sqrt{n}}$$

We have now shown that our algorithm recovers the angles with better accuracy as n increases. We conclude with further directions and applications.

4 Further Directions and Applications

Looking back to our objective function, we see a very natural relation to semidefinite programming. In particular, our problem is trying to maximize the follow-

ing SDP:

$$\begin{aligned} \max_{\Theta \in \mathbb{C}^{n\times n}} & \; trace(H^*\Theta) \\ \Theta &\succeq 0 \\ \Theta_{ii} &= 1 \; i = 1, \ldots, n \\ \text{rank}(\Theta) &\leq 1 \end{aligned}$$

In order to get this into the form of a usual SDP, we must relax the rank constraint. The issue, then, is showing that the solution to the relaxed problem gives the solution to the original problem. In fact, we have the following theorem [1]:

Theorem 4.1. *Consider $M = zz^* + \sigma W$, where W has $N(0,1)$ entries off the diagonal. Then if $\sigma = O(n^{1/4})$, the solution to the SDP relaxation is also a solution to the original problem. That is, the rank of the solution is* 1.

However, we expect the actual answer should be $\sqrt{n}$. In particular, we have the following conjecture:

Conjecture 4.2. *Let M be as aove. Then if $\sigma = O(\sqrt{n})$, the solution to the SDP relaxation is rank* 1.

Some progress has been made on this front, but we still don't have a proof.

We can also ask about generalizing the synchronization problems. To do this, view the real numbers modulo 2π as a group under addition. We can then ask

Problem 4.3. *Given a group G and n unknown elements $g_1, \ldots, g_n$, suppose we are given offset measurements in the form of $g_i g_j^{-1}$. Recover the group elements.*

This has many applications, for example to Cryo-Electron Microscopy. Cryo-EM is a technique used to determine the three-dimensional structure of biological molecules. These molecules are frozen in a layer of ice and imaged with an electron microscope, which gives a two-dimensional projection of the molecule. Of course, we don't know the orientation the molecule was at when we image. There exists a mechanism by which looking at two projections provides some information between their orientation. Thus, we can view this is as a synchronization problem with $G = SO(3)$. In general, spectral techniques can be used when the group G is compact.

Recently, a new approach has been suggested using Approximate Message Passing. This has not yet been written up and is still being worked on by Alex Wein and Will Perry.

References

[1] Badeira, A. (2015) MIT lecture notes, http://math.mit.edu/ bandeira/TenLecturesFortyTwoProblems.pdf

[2] Baik, J., Ben Arous, G., Peche, S. (2004) Phase transition of the largest eigenvalue for non-null complex sample covariance matrices, http://arxiv.org/abs/math/0403022

[3] Feral, P., Peche, S. (2007) The largest eigenvalue of rank one deformations of large Wigner matrices, Communications in Mathematical Physics 272 (1):185-228

[4] Khorunzhiy, O. (2003) Rooted trees and moments of large sparse random matrices, Discrete Mathematics and Theoretical Computer Science AC, pp. 145-154.

[5] Singer, A. (2009) Angular Synchronization by Eigenvectors and Semidefinite Programming, arXiv:0905.3174

[6] Wigner, E. P. (1958) On the distribution of tile roots of certain symmetric matrices, Annals of Mathematics 67 pp. 325-328