NIH Public Access
Author Manuscript
Appl Comput Harmon Anal. Author manuscript; available in PMC 2012 January 30.

Published in final edited form as:
Appl Comput Harmon Anal. 2011 January 30; 30(1): 20–36. doi:10.1016/j.acha.2010.02.001.

Angular Synchronization by Eigenvectors and Semidefinite Programming

A. Singer*
* Department of Mathematics and PACM, Princeton University, Fine Hall, Washington Road, Princeton NJ 08544-1000 USA, amits@math.princeton.edu

Abstract

The angular synchronization problem is to obtain an accurate estimation (up to a constant additive phase) for a set of unknown angles $\theta_1, \ldots, \theta_n$ from m noisy measurements of their offsets $\theta_i - \theta_j \bmod 2\pi$. Of particular interest is angle recovery in the presence of many outlier measurements that are uniformly distributed in $[0, 2\pi)$ and carry no information on the true offsets. We introduce an efficient recovery algorithm for the unknown angles from the top eigenvector of a specially designed Hermitian matrix. The eigenvector method is extremely stable and succeeds even when the number of outliers is exceedingly large. For example, we successfully estimate $n = 400$ angles from a full set of $m = \binom{400}{2}$ offset measurements of which 90% are outliers in less than a second on a commercial laptop. The performance of the method is analyzed using random matrix theory and information theory. We discuss the relation of the synchronization problem to the combinatorial optimization problem Max-2-Lin mod L and present a semidefinite relaxation for angle recovery, drawing similarities with the Goemans-Williamson algorithm for finding the maximum cut in a weighted graph. We present extensions of the eigenvector method to other synchronization problems that involve different group structures and their applications, such as the time synchronization problem in distributed networks and the surface reconstruction problems in computer vision and optics.

1 Introduction

The angular synchronization problem is to estimate n unknown angles $\theta_1, \ldots, \theta_n \in [0, 2\pi)$ from m noisy measurements δ_{ij} of their offsets $\theta_i - \theta_j \bmod 2\pi$. In general, only a subset of all possible $\binom{n}{2}$ offsets are measured. The set E of pairs $\{i, j\}$ for which offset measurements exist can be realized as the edge set of a graph $G = (V, E)$ with vertices corresponding to angles and edges corresponding to measurements.

When all offset measurements are exact with zero measurement error, it is possible to solve the angular synchronization problem iff the graph G is connected. Indeed, if G is connected, then it contains a spanning tree and all angles are sequentially determined by traversing the tree while summing the offsets modulo 2π. The angles are uniquely determined up to an additive phase, e.g., the angle of the root. On the other hand, if G is disconnected, then it is

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

impossible to determine the offset between angles that belong to disjoint components of the graph.

Sequential algorithms that integrate the measured offsets over a particular spanning tree of the graph are very sensitive to measurement errors, due to accumulation of the errors. It is therefore desirable to integrate all offset measurements in a globally consistent way. The need for such a globally consistent integration method comes up in a variety of applications. One such application is the time synchronization of distributed networks [17,23], where clocks measure noisy time offsets $t_i - t_j$ from which the determination of $t_1, \ldots, t_n \in \mathbb{R}$ is required. Other applications include the surface reconstruction problems in computer vision [13,1] and optics [30], where the surface is to be reconstructed from noisy measurements of the gradient to the surface and the graph of measurements is typically the two-dimensional regular grid. The most common approach in the above mentioned applications for a self consistent global integration is the least squares approach. The least squares solution is most suitable when the offset measurements have a small Gaussian additive error. The least squares solution can be efficiently computed and also mathematically analyzed in terms of the Laplacian of the underlying measurement graph.

There are many possible models for the measurement errors, and we are mainly interested in models that allow many outliers. An outlier is an offset measurement that has a uniform distribution on $[0, 2\pi)$ regardless of the true value for the offset. In addition to outliers that carry no information on the true angle values, there also exist of course good measurements whose errors are relatively small. We have no a-priori knowledge, however, which measurements are good and which are bad (outliers).

In our model, the edges of E can be split into a set of good edges E_{good} and a set of bad edges E_{bad}, of sizes m_{good} and m_{bad} respectively (with $m = |E| = m_{good} + m_{bad}$), such that

$$\begin{array}{ll} \delta_{ij} = \theta_i - \theta_j & \text{for } \{i,j\} \in E_{good} \\ \delta_{ij} \sim Uniform([0, 2\pi)) & \text{for } \{i,j\} \in E_{bad} \end{array} . \tag{1}$$

Perhaps it would be more realistic to allow a small discretization error for the good offsets, for example, by letting them have the wrapped normal distribution on the circle with mean $\theta_i - \theta_j$ and variance σ^2 (where σ is a typical discretization error). This discretization error can be incorporated into the mathematical analysis of Section 4 with a little extra difficulty. However, the effect of the discretization error is negligible compared to that of the outliers, so we choose to ignore it in order to make the presentation as simple as possible.

It is trivial to find a solution to (1) if some oracle whispers to our ears which equations are good and which are bad (in fact, all we need in that case is that E_{good} contains a spanning tree of G). In reality, we have to be able to tell the good from the bad on our own.

The overdetermined system of linear equations (modulo 2π)

$$\theta_i - \theta_j = \delta_{ij} \mod 2\pi, \quad \text{for } \{i,j\} \in E \tag{2}$$

can be solved by the method of least squares as follows. Introducing the complex-valued variables $z_i = e^{\imath\theta_i}$, the system (2) is equivalent to

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$z_i - e^{\iota\delta_{ij}} z_j = 0, \quad \text{for } \{i, j\} \in E, \tag{3}$$

which is an overdetermined system of homogeneous linear equations over $\mathbb{C}$. To prevent the solution from collapsing to the trivial solution $z_1 = z_2 = \cdots = z_n = 0$, we set $z_1 = 1$ (recall that the angles are determined up to a global additive phase, so we may choose $\theta_1 = 0$), and look for the solution $z_2, \ldots, z_n$ of (3) with minimal ℓ_2-norm residual. However, it is expected that the sum of squares errors would be overwhelmingly dominated by outlier equations, making least squares least favorable to succeed if the proportion of bad equations is large (see numerical results involving least squares in Table 3). We therefore seek for a solution method which is more robust to outliers.

Maximum likelihood is an obvious step in that direction. The maximum likelihood solution to (1) is simply the set of angles $\theta_1, \ldots, \theta_n$ that satisfies as many equations of (2) as possible. We may therefore define the self consistency error (SCE) of $\theta_1, \ldots, \theta_n$ as the number of equations not being satisfied

$$SCE(\theta_1, \ldots, \theta_n) = \#\{\{i, j\} \in E : \theta_i - \theta_j \neq \delta_{ij} \mod 2\pi\}. \tag{4}$$

As even the good equations contain some error (due to angular discretization and noise), a more suitable self consistency error is SCE_f that incorporates some penalty function f

$$SCE_f(\theta_1, \ldots, \theta_n) = \sum_{\{i,j\} \in E} f(\theta_i - \theta_j - \delta_{ij}), \tag{5}$$

where f: $[0, 2\pi) \in \mathbb{R}$ is a smooth periodic function with $f(0) = 0$ and $f(\theta) = 1$ for $|\theta| > \theta_0$, where θ_0 is the allowed discretization error. The minimization of (5) is equivalent to maximizing the log likelihood with a different probabilistic error model.

The maximum likelihood approach suffers from a major drawback though. It is virtually impossible to find the global minimizer $\theta_1, \ldots, \theta_n$ when dealing with large scale problems ($n \gg 1$), because the minimization of either (4) or (5) is a non-convex optimization problem in a huge parameter space. It is like finding a needle in a haystack.

In this paper we take a different approach and introduce two different estimators for the angles. The first estimator is based on an eigenvector computation while the second estimator is based on a semidefinite program (SDP) [38]. Our eigenvector estimator $\hat{\theta}_1, \ldots, \hat{\theta}_n$ is obtained by the following two-step recipe. In the first step, we construct an $n \times n$ complex-valued matrix H whose entries are

$$H_{ij} = \begin{cases} e^{\iota\delta_{ij}} & \{i, j\} \in E \\ 0 & \{i, j\} \notin E \end{cases}, \tag{6}$$

where $\iota = \sqrt{-1}$. The matrix H is Hermitian, i.e. $H_{ij} = \bar{H}_{ji}$, because the offsets are skew-symmetric $\delta_{ij} = -\delta_{ji} \mod 2\pi$. As H is Hermitian, its eigenvalues are real. The second step is to compute the top eigenvector v_1 of H with maximal eigenvalue, and to define the estimator in terms of this top eigenvector as

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$e^{\mathrm{i}\widehat{\theta}_i}=\frac{v_1(i)}{|v_1(i)|},\quad i=1,\ldots,n. \tag{7}$$

The philosophy leading to the eigenvector method is explained in Section 2.

The second estimator is based on the following SDP

$$\max_{\Theta\in\mathbb{C}^{n\times n}} \operatorname{trace}(H\Theta) \tag{8}$$

$$s.t.\Theta \succcurlyeq 0 \tag{9}$$

$$\Theta_{ii}=1\ \ i=1,2,\ldots,n, \tag{10}$$

where $\Theta \succcurlyeq 0$ is a shorthand notation for Θ being a Hermitian semidefinite positive matrix. The only difference between this SDP and the Goemans-Williamson algorithm for finding the maximum cut in a weighted graph [18] is that the maximization is taken over all semidefinite positive Hermitian matrices with complex-valued entries rather than just the real-valued symmetric matrices. The SDP-based estimator $\hat{\theta}_1, \ldots, \hat{\theta}_n$ is derived from the normalized top eigenvector v_1 of Θ by the same rounding procedure (7). Our numerical experiments show that the accuracy of the eigenvector method and the SDP method are comparable. Since the eigenvector method is much faster, we prefer using it for large scale problems. The eigenvector method is also numerically appealing, because in the useful case the spectral gap is large, rendering the simple power method an efficient and numerically stable way of computing the top eigenvector. The SDP method is summarized in Section 3.

In Section 4 we use random matrix theory to analyze the eigenvector method for two different measurement graphs: the complete graph and "small-world" graphs [39]. Our analysis shows that the top eigenvector of H in the complete graph case has a non-trivial correlation with the vector of true angles as soon as the proportion p of good offset measurements becomes greater than $\frac{1}{\sqrt{n}}$. In particular, the correlation goes to 1 as $np^2 \to \infty$, meaning a successful recovery of the angles. Our numerical simulations confirm these results and demonstrate the robustness of the estimator (7) to outliers.

In Section 5 we prove that the eigenvector method is asymptotically nearly optimal in the sense that it achieves the information theoretic Shannon bound up to a multiplicative factor that depends only on the discretization error of the measurements $2\pi/L$, but not on m and n. In other words, no method whatsoever can accurately estimate the angles if the proportion of good measurements is $o(\sqrt{\frac{n}{m}})$. The connection between the angular synchronization problem and Max-2-Lin mod L [3] is explored in Section 6. Finally, Section 7 is a summary and discussion of further applications of the eigenvector method to other synchronization problems over different groups.

2 The Eigenvector Method

Our approach to finding the self consistent solution for $\theta_1, \ldots, \theta_n$ starts with forming the following $n \times n$ matrix H

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$H_{ij} = \begin{cases} e^{\iota\delta_{ij}} & \{i,j\} \in E \\ 0 & \{i,j\} \notin E \end{cases}, \tag{11}$$

where $\iota = \sqrt{-1}$. Since

$$\delta_{ij} = -\delta_{ij}, \quad \text{for all } i, j = 1, \ldots, n, \tag{12}$$

it follows that $H_{ij} = \bar{H}_{ji}$, where for any complex number $z = a + ib$ we denote by $\bar{z} = a - \iota b$ its complex conjugate. In other words, the matrix H is Hermitian, i.e., $H^* = H$. We choose to set the diagonal elements of H to 0 (i.e., $H_{ii} = 0$).

Next, we consider the maximization problem

$$\max_{\theta_1,\ldots,\theta_n \in [0,2\pi)} \sum_{i,j=1}^{n} e^{-\iota\theta_i} H_{ij} e^{\iota\theta_j}, \tag{13}$$

and explain the philosophy behind it. For the correct set of angles $\theta_1, \ldots, \theta_n$, each good edge contributes

$$e^{-\iota\theta_i} e^{\iota(\theta_i - \theta_j)} e^{\iota\theta_j} = 1$$

to the sum in (13). The total contribution of the good edges is just the sum of ones, piling up to be exactly the total number of good edges m_{good}. On the other hand, the contribution of each bad edge will be uniformly distributed on the unit circle in the complex plane. Adding up the terms due to bad edges can be thought of as a discrete planar random walk where each bad edge corresponds to a unit size step at a uniformly random direction. These random steps mostly cancel out each other, such that the total contribution of the m_{bad} edges is only $O(\sqrt{m_{bad}})$. It follows that the objective function in (13) has the desired property of diminishing the contribution of the bad edges by a square root relative to the linear contribution of the good edges.

Still, the maximization problem (13) is a non-convex maximization problem which is quite difficult to solve in practice. We therefore introduce the following relaxation of the problem

$$\max_{\substack{z_1,\ldots,z_n \in \mathbb{C} \\ \sum_{i=1}^{n} |z_i|^2 = n}} \sum_{i,j=1}^{n} \bar{z}_i H_{ij} z_j. \tag{14}$$

That is, we replace the previous n individual constraints for each of the variables $z_i = e^{\iota\theta_i}$ to have a unit magnitude, by a single and much weaker constraint, requiring the sum of squared magnitudes to be n. The maximization problem (14) is that of a quadratic form whose solution is simply given by the top eigenvector of the Hermitian matrix H. Indeed, the spectral theorem implies that the eigenvectors $v_1, v_2, \ldots, v_n$ of H form an orthonormal basis

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

for $\mathbb{C}^n$ with corresponding real eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ satisfying $Hv_i = \lambda_i v_i$. Rewriting the constrained maximization problem (14) as

$$\max_{\left\|z\right\|^2 = n} z^* H z, \tag{15}$$

it becomes clear that the maximizer z is given by $z = v_1$, where v_1 is the normalized top eigenvector satisfying $Hv_1 = \lambda_1 v_1$ and $\|v_1\|^2 = n$, with λ_1 being the largest eigenvalue. The components of the eigenvector v_1 are not necessarily of unit magnitude, so we normalize them and define the estimated angles by

$$e^{\iota\widehat{\theta_i}} = \frac{v_1(i)}{|v_1(i)|}, \quad \text{for } i = 1, \ldots, n \tag{16}$$

(see also equation (7)).

The top eigenvector can be efficiently computed by the power iteration method that starts from a randomly chosen vector b_0 and iterates $b_{n+1} = \frac{Hb_n}{\left\|Hb_n\right\|}$. Each iteration requires just a matrix-vector multiplication that takes $O(n^2)$ operations for dense matrices, but only $O(m)$ operations for sparse matrices, where $m = |E|$ is the number of non-zero entries of H corresponding to edges in the graph. The number of iterations required by the power method decreases with the spectral gap that indeed exists and is analyzed in detail in Section 4.

Note that cycles in the graph of good edges lead to consistency relations between the offset measurements. For example, if the three edges $\{i, j\}$, $\{j, k\}$, $\{k, i\}$ are a triangle of good edges, then the corresponding offset angles δ_{ij}, δ_{jk} and δ_{ki} must satisfy

$$\delta_{ij} + \delta_{jk} + \delta_{ki} = 0 \mod 2\pi, \tag{17}$$

because

$$\delta_{ij} + \delta_{jk} + \delta_{ki} = \theta_i - \theta_j + \theta_j - \theta_k + \theta_k - \theta_i = 0 \mod 2\pi.$$

A closer look into the power iteration method reveals that multiplying the matrix H by itself integrates the information in the consistency relation of triplets, while higher order iterations exploit consistency relations of longer cycles. Indeed,

$$\begin{aligned} H_{ij}^2 = \sum_{k=1}^{n} H_{ik} H_{kj} = \sum_{k:\{i,k\},\{j,k\} \in E} e^{\iota\delta_{ik}} e^{\iota\delta_{kj}} \\ = \#\{k : \{i,k\} \text{ and } \{j,k\} \in E_{good}\} e^{\iota(\theta_i - \theta_j)} + \sum_{k:\{i,k\} \text{ or } \{j,k\} \in E_{bad}} e^{\iota(\delta_{ik} + \delta_{kj})}. \end{aligned} \tag{18}$$

The top eigenvector therefore integrates the consistency relations of all cycles.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

3 The semidefinite program approach

A different natural relaxation of the optimization problem (13) is using SDP. Indeed, the objective function in (13) can be written as

$$\sum_{i,j=1}^{n} e^{-\iota\theta_i} H_{ij} e^{\iota\theta_j} = \mathrm{trace}(H\Theta), \tag{19}$$

where Θ is the $n \times n$ complex-valued rank-one Hermitian matrix

$$\Theta_{ij} = e^{\iota(\theta_i - \theta_j)}. \tag{20}$$

Note that Θ has ones on its diagonal

$$\Theta_{ii} = 1, \quad i = 1, 2, \ldots, n. \tag{21}$$

Except for the non-convex rank-one constraint implied by (20), all other constraints are convex and lead to the natural SDP relaxation (8)–(10). This program is almost identical to the Goemans-Williamson SDP for finding the maximum cut in a weighted graph. The only difference is that here we maximize over all possible complex-valued Hermitian matrices, not just the symmetric real matrices. The SDP-based estimator corresponding to (8)–(10) is then obtained from the best rank-one approximation of the optimal matrix Θ using the Cholesky decomposition.

The SDP method may seem favorable to the eigenvector method as it explicitly imposes the unit magnitude constraint for $e^{\iota\theta_i}$. Our numerical experiments show that the two methods give similar results (see Table 3). Since the eigenvector method is much faster, it is also the method of choice for large scale problems.

4 Connections with random matrix theory and spectral graph theory

In this section we analyze the eigenvector method using tools from random matrix theory and spectral graph theory.

4.1 Analysis of the complete graph angular synchronization problem

We first consider the angular synchronization problem in which all $\binom{n}{2}$ angle offsets are given, so that the corresponding graph is the complete graph K_n of n vertices. We also assume that the probability for each edge to be good is p, independently of all other edges. This probabilistic model for the graph of good edges is known as the Erdős-Rényi random graph $G(n, p)$ [9]. We refer to this model as the complete graph angular synchronization model.

The elements of H in the complete graph angular synchronization model are random variables given by the following mixture model. With probability p the edge $\{i, j\}$ is good and $H_{ij} = e^{\iota(\theta_i - \theta_j)}$, whereas with probability $1 - p$ the edge is bad and $H_{ij} \sim Uniform(S^1)$. It is convenient to define the diagonal elements as $H_{ii} = p$.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The matrix H is Hermitian and the expected value of its elements is

$$\mathbb{E}H_{ij}=pe^{\iota(\theta_i-\theta_j)}. \tag{22}$$

In other words, the expected value of H is the rank-one matrix

$$\mathbb{E}H=npzz^*, \tag{23}$$

where z is the normalized vector ($\|z\| = 1$) given by

$$z_i=\frac{1}{\sqrt{n}}e^{\iota\theta_i}, \quad i=1,\ldots,n. \tag{24}$$

The matrix H can be decomposed as

$$H=npzz^*+R, \tag{25}$$

where $R = H - \mathbb{E}H$ is a random matrix whose elements have zero mean, with $R_{ii} = 0$, and for $i \neq j$

$$R_{ij}=\begin{cases}(1-p)e^{\iota(\theta_i-\theta_j)} & \text{with probability } p\\ e^{\iota\phi}-pe^{\iota(\theta_i-\theta_j)} & \text{w.p.} 1-p \text{ and } \phi\sim Uniform([0,2\pi))\end{cases}. \tag{26}$$

The variance of R_{ij} is

$$\mathbb{E}|R_{ij}|^2=(1-p)^2p+(1+p^2)(1-p)=1-p^2 \tag{27}$$

for $i \neq j$ and 0 for the diagonal elements. Note that for $p = 1$ the variance vanishes as all edges become good.

The distribution of the eigenvalues of the random matrix R follows Wigner's semi-circle law [40,41] whose support is $[-2\sqrt{n(1-p^2)}, 2\sqrt{n(1-p^2)}]$. The largest eigenvalue of R, denoted $\lambda_1(R)$, is concentrated near the right edge of the support [2] and the universality of the edge of the spectrum [34] implies that it follows the Tracy-Widom distribution [36] even when the entries of R are non-Gaussian. For our purposes, the approximation

$$\lambda_1(R)\approx 2\sqrt{n(1-p^2)} \tag{28}$$

will suffice, with the probabilistic error bound given in [2].

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The matrix $H = npzz^* + R$ can be considered as a rank-one perturbation to a random matrix. The distribution of the largest eigenvalue of such perturbed random matrices was investigated in [29,11,15] for the particular case where z is proportional to the all-ones vector $(1\ 1\ \cdots\ 1)^T$. Although our vector z given by (24) is different, without loss of generality, we can assume $\theta_1 = \theta_2 = \ldots = \theta_n = 0$, because the matrix zz^* can be reduced to the all-ones matrix by conjugation with the $n \times n$ diagonal matrix Z whose diagonal elements are $Z_{ii} = z_i$, $i = 1, \ldots, n$. Thus, adapting [11, Theorem 1.1] to H gives that for

$$np > \sqrt{n(1-p^2)} \tag{29}$$

the largest eigenvalue $\lambda_1(H)$ jumps outside the support of the semi-circle law and is normally distributed with mean μ and variance σ^2 given by

$$\lambda_1(H) \sim \mathcal{N}(\mu, \sigma^2),\ \ \mu = \frac{np}{\sqrt{1-p^2}} + \frac{\sqrt{1-p^2}}{p},\ \ \sigma^2 = \frac{(n+1)p^2 - 1}{np^2}(1-p^2), \tag{30}$$

whereas for $np < \sqrt{n(1-p^2)}$, $\lambda_1(H)$ still tends to the right edge of the semicircle given at $2\sqrt{n(1-p^2)}$.

Note that the factor of 2 that appears in (28) has disappeared from (29), which is perhaps somewhat non-intuitive: it is expected that $\lambda_1(H) > \lambda_1(R)$ whenever $np > \lambda_1(R)$, but the theorem guarantees that $\lambda_1(H) > \lambda_1(R)$ already for $np > \frac{1}{2}\lambda_1(R)$.

The condition (29) also implies a lower bound on the correlation between the normalized top eigenvector v_1 of H and the vector z. To that end, consider the eigenvector equation satisfied by v_1:

$$\lambda_1(H)v_1 = Hv_1 = (npzz^* + R)v_1. \tag{31}$$

Taking the dot product with v_1 yields

$$\lambda_1(H) = np|\langle z, v_1\rangle|^2 + v_1^* R v_1. \tag{32}$$

From $v_1^* R v_1 \le \lambda_1(R)$ we obtain the lower bound

$$|\langle z, v_1\rangle|^2 \ge \frac{\lambda_1(H) - \lambda_1(R)}{np}, \tag{33}$$

with $\lambda_1(H)$ and $\lambda_1(R)$ given by (28) and (30). Thus, if the spectral gap $\lambda_1(H) - \lambda_1(R)$ is large enough, then v_1 must be close to z, in which case the eigenvector method successfully recovers the unknown angles. Since the variance of the correlation of two random unit

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

vectors in $\mathbb{R}^n$ is $1/n$, the eigenvector method would give above random correlation values whenever

$$\frac{\lambda_1(H) - \lambda_1(R)}{np} > \frac{1}{n}. \tag{34}$$

Replacing in (34) $\lambda_1(H)$ by μ from (30) and $\lambda_1(R)$ by (28) and multiplying by $p\sqrt{n}$ yields the condition

$$\frac{\sqrt{n}p}{\sqrt{1-p^2}} + \frac{\sqrt{1-p^2}}{\sqrt{n}p} - 2\sqrt{1-p^2} > \frac{p}{\sqrt{n}}. \tag{35}$$

Since $\frac{\sqrt{n}p}{\sqrt{1-p^2}} + \frac{\sqrt{1-p^2}}{\sqrt{n}p} \geq 2$, it follows that (35) is satisfied for

$$p > \frac{1}{\sqrt{n}}. \tag{36}$$

Thus, already for $p > \frac{1}{\sqrt{n}}$ we should obtain above random correlations between the vector of angles z and the top eigenvector v_1. We therefore define the threshold probability p_c as

$$p_c = \frac{1}{\sqrt{n}}. \tag{37}$$

When $np \gg \lambda_1(R)$, the correlation between v_1 and z can be predicted by using regular perturbation theory for solving the eigenvector equation (31) in an asymptotic expansion with the small parameter $\varepsilon = \frac{\lambda_1(R)}{np}$. Such perturbations are derived in standard textbooks on quantum mechanics aiming to find approximations to the energy levels and eigenstates of perturbed time-independent Hamiltonians (see, e.g., [20, Chapter 6]). In our case, the resulting asymptotic expansions of the non-normalized eigenvector v_1 and of the eigenvalue $\lambda_1(H)$ are given by

$$v_1 \sim z + \frac{1}{np}[Rz - (z^* Rz)z] + \ldots, \tag{38}$$

and

$$\lambda_1(H) \sim np + z^* Rz + \ldots. \tag{39}$$

Note that the first order term in (38) is perpendicular to the leading order term z, from which it follows that the angle α between the eigenvector v_1 and the vector of true angles z satisfies the asymptotic relation

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\tan^2\alpha \sim \frac{\|Rz\|^2 - (z^*Rz)^2}{(np)^2} + \dots, \tag{40}$$

because $\|Rz - (z^*Rz)z\|^2 = \|Rz\|^2 - (z^*Rz)^2$. The expected values of the numerator terms in (40) are given by

$$\mathbb{E}\|Rz\|^2 = \mathbb{E}\sum_{i=1}^{n}\left|\sum_{j=1}^{n} R_{ij}z_j\right|^2 = \sum_{i,j=1}^{n}\mathrm{Var}(R_{ij}z_j) = \sum_{i=1}^{n}\sum_{j\neq i}|z_j|^2(1-p^2) = (n-1)(1-p^2), \tag{41}$$

and

$$\begin{aligned}\mathbb{E}(z^*Rz)^2 &= \mathbb{E}\left[\sum_{i,j=1}^{n} R_{ij}\bar{z}_i z_j\right]^2 = \sum_{i,j=1}^{n}\mathrm{Var}(R_{ij}\bar{z}_i z_j) = (1-p^2)\sum_{i\neq j}|z_i|^2|z_j|^2 \\ &= (1-p^2)\left[\left(\sum_{i=1}^{n}|z_i|^2\right)^2 - \sum_{i=1}^{n}|z_i|^4\right] = (1-p^2)\left(1-\frac{1}{n}\right),\end{aligned} \tag{42}$$

where we used that R_{ij} are i.i.d zero mean random variables with variance given by (27) and that $|z_i|^2 = \frac{1}{n}$. Substituting (41)–(42) into (40) results in

$$\mathbb{E}\tan^2\alpha \sim \frac{(n-1)^2(1-p^2)}{n^3p^2} + \dots, \tag{43}$$

which for $p \ll 1$ and $n \gg 1$ reads

$$\mathbb{E}\tan^2\alpha \sim \frac{1}{np^2} + \dots. \tag{44}$$

This expression shows that as np^2 goes to infinity, the angle between v_1 and z goes to zero and the correlation between them goes to 1. For $np^2 \gg 1$, the leading order term in the expected squared correlation $\mathbb{E}\cos^2\alpha$ is given by

$$\mathbb{E}\cos^2\alpha = \mathbb{E}\frac{1}{1+\tan^2\alpha} \sim \frac{1}{1+\frac{1}{np^2}} + \dots. \tag{45}$$

We conclude that even for very small p values, the eigenvector method successfully recovers the angles if there are enough equations, that is, if np^2 is large enough.

Figure 1 shows the distribution of the eigenvalues of the matrix H for $n = 400$ and different values of p. The spectral gap decreases as p is getting smaller. From (29) we expect a

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

spectral gap for $p \geq p_c$ where the critical value is $p_c = \frac{1}{\sqrt{400}} = 0.05$. The experimental values of $\lambda_1(H)$ also agree with (30). For example, for $n = 400$ and $p = 0.15$, the expected value of the largest eigenvalue is $\mu = 67.28$ and its standard deviation is $\sigma = 0.93$, while for $p = 0.1$ we get $\mu = 50.15$ and $\sigma = 0.86$; these value are in full agreement with the location of the largest eigenvalues in Figures 1(a)–1(b). Note that the right edge of the semi-circle is smaller than $2\sqrt{n} = 40$, so the spectral gap is significant even when $p = 0.1$.

The skeptical reader may wonder whether the existence of a visible spectral gap necessarily implies that the normalized top eigenvector v_1 correctly recovers the original set of angles $\theta_1, \ldots, \theta_n$ (up to a constant phase). To that end, we compute the following two measures of correlation ρ_1 and ρ_2 for the correlation between the vector of true angles z and the computed normalized top eigenvector v_1:

$$\rho_1 = \left| \frac{1}{n} \sum_{i=1}^{n} e^{-i\theta_i} \frac{v_1(i)}{|v_1(i)|} \right|, \quad \rho_2 = \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e^{-i\theta_i} v_1(i) \right| = \left| \langle z, v_1 \rangle \right|. \tag{46}$$

The correlation ρ_1 takes into account the rounding procedure (16), while ρ_2 is simply the dot product between v_1 and z without applying any rounding. Clearly, $\rho_1, \rho_2 \leq 1$ (Cauchy-Schwartz), and $\rho_1 = 1$ iff the two sets of angles are the same up to a rotation. Note that it is possible to have $\rho_1 = 1$ with $\rho_2 < 1$. This happens when the angles implied by $v_1(i)$ are all correct, but the magnitudes $|v_1(i)|$ are not all the same. Table 1 summarizes the experimentally obtained correlations ρ_1, ρ_2 for different values of p with $n = 100$ (Table 0(a)) and $n = 400$ (Table 0(b)). The experimental results show that for large values of np^2 the correlation is very close to 1, indicating a successful recovery of the angles. The third column, indicating the values of $\left(1 + \frac{1}{np^2}\right)^{-\frac{1}{2}}$ is motivated by the asymptotic expansion (45) and seems to provide a very good approximation for ρ_2 when $np^2 \gg 1$, with deviations attributed to higher order terms of the asymptotic expansion and to statistical fluctuations around the mean value. Below the threshold probability (ending rows of Tables 0(a) and 0(b) with $np^2 < 1$), the correlations take values near $\frac{1}{\sqrt{n}}$, as expected from the correlation of two unit random vectors in $\mathbb{R}^n$ ($\frac{1}{\sqrt{100}} = 0.1$ and $\frac{1}{\sqrt{400}} = 0.05$).

From the practical point of view, most important is the fact that the eigenvector method successfully recovers the angles even when a large portion of the offset measurements consists of just outliers. For example, for $n = 400$, the correlation obtained when 85% of the offset measurements were outliers (only 15% are good measurements) was $\rho_1 = 0.97$.

4.2 Analysis of the angular synchronization problem in general

We turn to analyze the eigenvector method for general measurement graphs, where the graph of good measurements is assumed to be connected, while the graph of bad edges is assumed to be made of edges that are uniformly drawn from the remaining edges of the complete graph once the good edges has been removed from it. Our analysis is based on generalizing the decomposition given in (25).

Let A be the adjacency matrix for the set of good edges E_{good}:

$$A_{ij} = \begin{cases} 1 & \{i, j\} \in E_{good} \\ 0 & \{i, j\} \notin E_{good} \end{cases}. \tag{47}$$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

As the matrix A is symmetric, it has a complete set of real eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ and corresponding real orthonormal eigenvectors $\psi_1, \ldots, \psi_n$ such that

$$A = \sum_{l=1}^{n} \lambda_l \psi_l \psi_l^T . \tag{48}$$

Let Z be an $n \times n$ diagonal matrix whose diagonal elements are $Z_{ii} = e^{\iota\theta_i}$. Clearly, Z is a unitary matrix ($ZZ^* = I$). Define the Hermitian matrix B by conjugating A with Z

$$B = ZAZ^* . \tag{49}$$

It follows that the eigenvalues of B are equal to the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A, and the corresponding eigenvectors $\{\varphi_l\}_{l=1}^n$ of B, satisfying $B\varphi_l = \lambda_l \varphi_l$ are given by

$$\varphi_l = Z\psi_l, \quad l = 1, \ldots, n. \tag{50}$$

From (49) it follows that

$$B_{ij} = \begin{cases} e^{\iota(\theta_i - \theta_j)} & \{i, j\} \in E_{good} \\ 0 & \{i, j\} \notin E_{good} \end{cases} . \tag{51}$$

We are now ready to decompose the matrix H defined in (11) as

$$H = B + R, \tag{52}$$

where R is a random matrix whose elements are given by

$$R_{ij} = \begin{cases} e^{\iota\delta_{ij}} & \{i, j\} \in E_{bad} \\ 0 & \{i, j\} \notin E_{bad} \end{cases} , \tag{53}$$

where $\delta_{ij} \sim Uniform([0, 2\pi)$ for $\{i, j\} \in E_{bad}$. The decomposition (52) is extremely useful, because it sheds light into the eigen-structure of H in terms of the much simpler eigen-structures of B and R.

First, consider the matrix B defined in (49), which shares the same spectrum with A and whose eigenvectors $\varphi_1, \ldots, \varphi_n$ are phase modulations of the eigenvectors $\psi_1, \ldots, \psi_n$ of A. If the graph of good measurements is connected, as it must be in order to have a unique solution for the angular synchronization problem (see second paragraph of Section 1), then the Perron-Frobenius theorem (see, e.g., [22, Chapter 8]) for the non-negative matrix A implies that the entries of ψ_1 are all positive

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\psi_1(i)>0, \quad \text{for all } i=1,2,\ldots,n, \tag{54}$$

and therefore the complex phases of the coordinates of the top eigenvector $\varphi_1 = Z\psi_1$ of B are identical to the true angles, that is, $e^{\mathbf{i}\theta_i}=\frac{\varphi_1(i)}{|\varphi_1(i)|}$. Hence, if the top eigenvector of H is highly correlated with the top eigenvector of B, then the angles will be estimated with high accuracy. We will shortly derive the precise condition that guarantees such a high correlation between the eigenvectors of H and B.

The spectral gap Δ_{good} of the good graph is the difference between its first and second eigenvalues, i.e., $\Delta_{good} = \lambda_1(A) - \lambda_2(A)$. The Perron-Frobenius theorem and the connectivity of the graph of good measurements also imply that $\Delta_{good} > 0$.

Next, we turn to analyze the spectrum of the random matrix R given in (53). We assume that the m_{bad} bad edges were drawn uniformly at random from the remaining edges of the complete graph on n vertices that are not already good edges. There are only $2m_{bad}$ nonzero elements in R, which makes R a sparse matrix with an average number of $2m_{bad}/n$ nonzero entries per row. The nonzero entries of R have zero mean and unit variance. The spectral norm of such sparse random matrices was studied in [25,24] where it was shown that with probability 1,

$$\limsup_{n\to\infty}\frac{\sqrt{n}}{\sqrt{2m_{bad}}}\lambda_1 R \le 2$$

as long as $\frac{m_{bad}}{n\log n}\to\infty$ as $n\to\infty$. The implication of this result is that we can approximate $\lambda_1(R)$ with

$$\lambda_1(R) \approx 2\frac{\sqrt{2m_{bad}}}{\sqrt{n}}. \tag{55}$$

Similar to the spectral gap condition (29), requiring

$$\Delta_{good}>\frac{1}{2}\lambda_1(R), \tag{56}$$

ensures that with high probability, the top eigenvector of H would be highly correlated with the top eigenvector of B. Plugging (55) into (56), we get the condition

$$\Delta_{good}>\frac{\sqrt{2m_{bad}}}{\sqrt{n}}. \tag{57}$$

We illustrate the above analysis for the small world graph, starting with a neighborhood graph on the unit sphere S^2 with n vertices corresponding to points on the sphere and m edges, and rewiring each edge with probability $1-p$ at random, resulting shortcut edges. The shortcut edges are considered as bad edges, while unperturbed edges are the good

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

edges. As the original m edges of the small world graph are rewired with probability $1 - p$, the expected number of bad edges $\mathbb{E}m_{bad}$ and the expected number of good edges $\mathbb{E}m_{good}$ are given by

$$\mathbb{E}m_{good}=pm, \quad \mathbb{E}m_{bad}=(1-p)m, \tag{58}$$

with relatively small fluctuations of $O(\sqrt{mp(1-p)})$.

The average degree of the original unperturbed graph is $\bar{d}=\frac{2m}{n}$. Assuming uniform sampling of points on the sphere, it follows that the average area of the spherical cap covered by the neighboring points is $4\pi\frac{\bar{d}}{n}=\frac{8\pi m}{n^2}$. The average opening angle η corresponding to this cap satisfies $2\pi(1-\cos\eta)=\frac{8\pi m}{n^2}$, or $1-\cos\eta=\frac{4m}{n^2}$. Consider the limit $m, n \to \infty$ while keeping the ratio $c = 4m/n^2$ constant. By the law of large numbers, the matrix $\frac{1}{n}A$ converges in this limit to the integral convolution operator $\mathcal{K}$ on S^2 (see, e.g., [7]), given by

$$(\mathcal{K}f)(\beta)=\frac{p}{4\pi}\int_{S^2}\chi_{[1-c,1]}(\langle\beta,\beta'\rangle)f(\beta')dS_{\beta'}, \ \beta\in S^2, \tag{59}$$

where χ_I is the characteristic function of the interval I.

The classical Funk-Hecke theorem (see, e.g., [28, p. 195]) asserts that the spherical harmonics are the eigenfunctions of convolution operators over the sphere, and the eigenvalues λ_l are given by

$$\lambda_l(\mathcal{K})=\frac{p}{2}\int_{-1}^{1}\chi_{[1-c,1]}P_l(t)dt=\frac{p}{2}\int_{1-c}^{1}P_l(t)dt,$$

and have multiplicities $2l + 1$ ($l = 0, 1, 2, \ldots$), where P_l are the Legendre polynomials ($P_0(t) = 1$, $P_1(t) = t$, …). In particular, $\lambda_0(\mathcal{K})=\frac{pc}{2}$, $\lambda_1(\mathcal{K})=\frac{pc}{2}(1-\frac{1}{2}c)$, and the spectral gap of $\mathcal{K}$ is $\Delta(\mathcal{K})=\frac{pc^2}{4}$. The spectral gap of A is approximately

$$\Delta_{good}\approx n\Delta(\mathcal{K})=\frac{npc^2}{4}=\frac{4m^2p}{n^3}. \tag{60}$$

Plugging (58) and (60) into (57) yields the condition

$$\frac{4m^2p}{n^3}>\frac{\sqrt{2(1-p)m}}{\sqrt{n}}, \tag{61}$$

which is satisfied for $p > p_c$, where p_c is the threshold probability

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$p_c \geq \sqrt{\frac{n^5}{8m^3}}. \tag{62}$$

We note that this estimate for the threshold probability is far from being tight and can be improved in principle by taking into account the entire spectrum of the good graph rather than just the spectral gap between the top eigenvalues, but we do not attempt to derive tighter bounds here.

We end this section by describing the results of a few numerical experiments. Figure 2 shows the histogram of the eigenvalues of the matrix H for small-world graphs on S^2. Each graph was generated by sampling n points $\beta_1, \ldots, \beta_n$ on the unit sphere S^2 in $\mathbb{R}^3$ from the uniform distribution as well as n random rotation angles $\theta_1, \ldots, \theta_n$ uniformly distributed in $[0, 2\pi)$. An edge between i and j exists iff $\langle \beta_i, \beta_j \rangle > 1 - \varepsilon$, where ε is a small parameter that determines the connectivity (average degree) of the graph. The resulting graph is a neighborhood graph on S^2. The small world graphs were obtained by randomly rewiring the edges of the neighborhood graph. Every edge is rewired with probability $1 - p$, so that the expected proportion of good edges is p.

The histograms of Figure 2 for the eigenvalues of H seem to be much more exotic than the ones obtained in the complete graph case shown in Figure 1. In particular, there seems to be a long tail of large eigenvalues, rather than a single eigenvalue that stands out from all the others. But now we understand that these eigenvalues are nothing but the top eigenvalues of the adjacency matrix of the good graph, related to the spherical harmonics. This behavior is better visible in Figure 3.

The experimental correlations given in Table 2 indicate jumps in the correlation values that occur between $p = 0.15$ and $p = 0.2$ for $n = 100$ and between $p = 0.1$ and $p = 0.12$ for $n = 400$. The experimental threshold values seem to follow the law $p_c \approx \sqrt{\frac{n}{2m}}$ that holds for the complete graph case (36) with $m = \binom{n}{2}$. As mentioned earlier, (62) is a rather pessimistic estimate of the threshold probability.

Also evident from Table 2 is that the correlation goes to 1 as $2mp^2/n \to \infty$. We remark that using regular perturbation theory and the relation of the eigenstructure of B to the spherical harmonics, it should be possible to obtain an asymptotic series for the correlation in terms of the large parameter $2mp^2/n$, similar to the asymptotic expansion (45).

The comparison between the eigenvector and SDP methods (as well as the least squares method of Section 1) is summarized in Table 3 showing the numerical correlations for $n = 200$, $\varepsilon = 0.3$ (number of edges $m \approx 3000$) and for different values of p. Although the SDP is slightly more accurate, the eigenvector method runs faster.

5 Information Theoretic Analysis

The optimal solution to the angular synchronization problem can be considered as the set of angles that maximizes the log-likelihood. Unfortunately, the log-likelihood is a non-convex function and the maximum likelihood cannot be found in a polynomial time. Both the eigenvector method and the SDP method are polynomial-time relaxations of the maximum log-likelihood problem. In the previous section we showed that the eigenvector method fails to recover the true angles when p is below the threshold probability p_c^{eig}. It is clear that even the maximum likelihood solution would fail to recover the correct set of angles below some

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

(perhaps lower) threshold. It is therefore natural to ask if the threshold value of the polynomial eigenvector method gets close to the optimal threshold value of the exponential-time maximum likelihood exhaustive search. In this section we provide a positive answer to this question using the information theoretic Shannon bound [8]. Specifically, we show that the threshold probability for the eigenvector method is asymptotically larger by just a multiplicative factor compared to the threshold probability of the optimal recovery algorithm. The multiplicative factor is a function of the angular discretization resolution, but not a function of n and m. The eigenvector method becomes less optimal as the discretization resolution improves.

We start the analysis by recalling that from the information theoretic point of view, the uncertainty in the values of the angles is measured by their entropy. The noisy offset measurements carry some bits of information on the angle values, therefore decreasing their uncertainty, which is measured by the conditional entropy that we need to estimate.

The angles $\theta_1, \ldots, \theta_n$ can take any real value in the interval $[0, 2\pi)$. However, an infinite number of bits is required to describe real numbers, and so we cannot hope to determine the angles with an arbitrary precision. Moreover, the offset measurements are often also discretized. We therefore seek to determine the angles only up to some discretization precision $\frac{2\pi}{L}$, where L is the number of subintervals of $[0, 2\pi)$ obtained by dividing the unit circle is into L equally sized pieces.

Before observing any of the offset measurements, the angles are uniformly distributed on $\{0, 1, \ldots, L-1\}$, that is, each of them falls with equal probability $1/L$ to any of the L subintervals. It follows that the entropy of the i'th angle θ_i is given by

$$H(\theta_i) = -\sum_{l=0}^{L-1} \frac{1}{L} \log_2 \frac{1}{L} = \log_2 L, \quad \text{for } i = 1, 2, \ldots, n. \tag{63}$$

We denote by $\boldsymbol{\theta}^n = (\theta_1, \ldots, \theta_n)$ the vector of angles. Since $\theta_1, \ldots, \theta_n$ are independent, their joint entropy $H(\boldsymbol{\theta}^n)$ is given by

$$H(\theta^n) = \sum_{i=1}^{n} H(\theta_i) = n \log_2 L, \tag{64}$$

reflecting the fact that the configuration space is of size $L^n = 2^{n \log_2 L}$.

Let δ_{ij} be the random variable for the outcome of the noisy offset measurement of θ_i and θ_j. The random variable δ_{ij} is also discretized and takes values in $\{0, 1, \ldots, L-1\}$. We denote by $\boldsymbol{\delta}^m = (\delta_{i_1 j_1}, \ldots, \delta_{i_m j_m})$ the vector of all offset measurements. Conditioned on the values of θ_i and θ_j, the random variable δ_{ij} has the following conditional probability distribution

$$\Pr\{\delta_{ij} | \theta_i, \theta_j\} = \begin{cases} \frac{1-p}{L} & \delta_{ij} \neq \theta_i - \theta_j \mod L, \\ p + \frac{1-p}{L} & \delta_{ij} = \theta_i - \theta_j \mod L, \end{cases} \tag{65}$$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

because with probability $1 - p$ the measurement δ_{ij} is an outlier that takes each of the L possibilities with equal probability $\frac{1}{L}$, and with probability p it is a good measurement that equals $\theta_i - \theta_j$. It follows that the conditional entropy $H(\delta_{ij} \mid \theta_i, \theta_j)$ is

$$H(\delta_{ij}|\theta_i,\theta_j) = -(L-1)\frac{1-p}{L}\log_2\frac{1-p}{L} - \left(p+\frac{1-p}{L}\right)\log_2\left(p+\frac{1-p}{L}\right). \tag{66}$$

We denote this entropy by $H(L, p)$ and its deviation from $\log_2 L$ by $I(L, p)$, that is,

$$H(L,p) \equiv -(L-1)\frac{1-p}{L}\log_2\frac{1-p}{L} - \left(p+\frac{1-p}{L}\right)\log_2\left(p+\frac{1-p}{L}\right). \tag{67}$$

and

$$I(L,p) \equiv \log_2 L - H(L,p). \tag{68}$$

Without conditioning, the random variable δ_{ij} is uniformly distributed on $\{0, \ldots, L-1\}$ and has entropy

$$H(\delta_{ij}) = \log_2 L. \tag{69}$$

It follows that the mutual information $I(\delta_{ij}; \theta_i, \theta_j)$ between the offset measurement δ_{ij} and the angle values θ_i and θ_j is

$$I(\delta_{ij};\theta_i,\theta_j) = H(\delta_{ij}) - H(\delta_{ij}|\theta_i,\theta_j) = \log_2 L - H(L,p) = I(L,p). \tag{70}$$

This mutual information measures the reduction in the uncertainty of the random variable δ_{ij} from knowledge of θ_i and θ_j. Due to the symmetry of the mutual information,

$$I(\delta_{ij};\theta_i,\theta_j) = H(\delta_{ij}) - H(\delta_{ij}|\theta_i,\theta_j) = H(\theta_i,\theta_j) - H(\theta_i,\theta_j|\delta_{ij}), \tag{71}$$

the mutual information is also the reduction in uncertainty of the angles θ_i and θ_j given the noisy measurement of their offset δ_{ij}. Thus,

$$H(\theta_i,\theta_j|\delta_{ij}) = H(\theta_i,\theta_j) - I(\delta_{ij};\theta_i,\theta_j). \tag{72}$$

Similarly, given all m offset measurements $\boldsymbol{\delta}^m$, the uncertainty in $\boldsymbol{\theta}^n$ is given by

$$H(\theta^n|\delta^m) = H(\theta^n) - I(\delta^m;\theta^n), \tag{73}$$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

with

$$I(\delta^m;\theta^n)=H(\delta^m) - H(\delta^m|\theta^n). \tag{74}$$

A simple upper bound for this mutual information is obtained by explicit evaluation of the conditional entropy $H(\boldsymbol{\delta}^m|\boldsymbol{\theta}^n)$ combined with a simple upper bound on the joint entropy term $H(\boldsymbol{\delta}^m)$. First, note that given the values of $\theta_1, \ldots, \theta_n$, the offsets become independent random variables. That is, knowledge of $\delta_{i_1 j_1}$ (given $\theta_{i_1}, \theta_{j_1}$) does not give any new information on the value of $\delta_{i_2 j_2}$ (given $\theta_{i_2}, \theta_{j_2}$). The conditional probability distribution of the offsets is completely determined by (65), and the conditional entropy is therefore the sum of m identical entropies of the form (66)

$$H(\delta^m|\theta^n)=mH(L, p). \tag{75}$$

Next, bounding the joint entropy $H(\boldsymbol{\delta}^m)$ by the logarithm of its configuration space size L^m yields

$$H(\delta^m) \leq m\log_2 L. \tag{76}$$

Note that this simple upper bound ignores the dependencies among the offsets which we know to exist, as implied, for example, by the triplet consistency relation (17). As such, (76) is certainly not a tight bound, but still good enough to prove our claim about the nearly optimal performance of the eigenvector method.

Plugging (75) and (76) in (74) yields the desired upper bound on the mutual information

$$I(\delta^m;\theta^n) \leq m\log_2 L - mH(L, p)=mI(L, p). \tag{77}$$

Now, substituting the bound (77) and the equality (64) in (73) gives a lower bound for the conditional entropy

$$H(\theta^n|\delta^m) \geq n\log_2 L - mI(L, p). \tag{78}$$

We may interpret this bound in the following way. Before seeing any offset measurement the entropy of the angles is $n \log_2 L$, and each of the m offset measurements can decrease the conditional entropy by at most $I(L, p)$, the information that it carries.

The bound (78) demonstrates, for example, that for fixed n, p and L, the conditional entropy is bounded from below by a linear decreasing function of m. It follows that unless m is large enough, the uncertainty in the angles would be too large. Information theory says that a successful recovery of all $\theta_1, \ldots, \theta_n$ is possible only when their uncertainty, as expressed by the conditional entropy, is small enough. The last statement can be made precise by Fano's inequality and Wolfowitz' converse, also known as the weak and strong converse theorems to the coding theorem that provide a lower bound for the probability of the error probability

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

in terms of the conditional entropy, see, e.g., [8, Chapter 8.9, pages 204–207] and [16, Chapter 5.8, pages 173–176].

In the language of coding, we may think of $\boldsymbol{\theta}^n$ as a codeword that we are trying to decode from the noisy vector of offsets $\boldsymbol{\delta}^m$ which is probabilistically related to $\boldsymbol{\theta}^n$. The codeword $\boldsymbol{\theta}^n$ is originally uniformly distributed on $\{1, 2, \ldots, 2^{n \log_2 L}\}$ and from $\boldsymbol{\delta}^m$ we estimate $\boldsymbol{\theta}^n$ as one of the $2^{n \log_2 L}$ possibilities. Let the estimate be $\hat{\boldsymbol{\theta}}^n$ and define the probability of error as $P_e = \Pr\{\hat{\boldsymbol{\theta}}^n \neq \boldsymbol{\theta}^n\}$. Fano's inequality [8, Lemma 8.9.1, page 205] gives the following lower bound on the error probability

$$H(\theta^n|\delta^m) \leq 1 + P_e n \log_2 L. \tag{79}$$

Combining (79) with the lower bound for the conditional entropy (78) we obtain a weak lower bound on the error probability

$$P_e \geq 1 - \frac{m}{n}\frac{I(L,p)}{\log_2 L} - \frac{1}{n \log_2 L}. \tag{80}$$

This lower bound for the probability of error is applicable to all decoding algorithms, not just for the eigenvector method. For large n, we see that for any $\beta < 1$,

$$\frac{m}{n}\frac{I(L,p)}{\log_2 L} < \beta \Rightarrow P_e \geq 1 - \beta + o(1). \tag{81}$$

We are mainly interested in the limit $m, n \to \infty$ and $p \to 0$ with L being fixed. The Taylor expansion of $I(L, p)$ (given by (67)–(68)) near $p = 0$ reads

$$I(L,p) = \frac{1}{2}(L-1)p^2 + O(p^3). \tag{82}$$

Combining (81) and (82) we obtain that

$$p = \sqrt{\frac{n}{m}\frac{2\log_2 L}{(L-1)}\beta} \Rightarrow P_e \geq 1 - \beta + o(1), \quad \text{as } n, m \to \infty, n/m \to 0. \tag{83}$$

Note that $n/m \to 0$, because $m \geq n \log n$ in order to ensure with high probability the connectivity of the measurement graph G. The bound (83) was derived using the weak converse theorem (Fano's inequality). It is also possible to show that the probability of error goes exponentially to 1 (using the Wolfowitz' converse and Chernoff bound, see [16, Theorem 5.8.5, pages 173–176]).

The above discussion shows that there does not exist a decoding algorithm with a small probability for the error for values of p below the threshold probability p_c^{inf} given by

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$p_c^{inf} = \sqrt{\frac{n}{m}\frac{2\log_2 L}{L-1}}. \tag{84}$$

Note that for $L = 2$, the threshold probability $p_c^{eig} = \frac{1}{\sqrt{n}}$ of the eigenvector method in the complete graph case for which $m = \binom{n}{2}$ is 2 times smaller than p_c^{inf}. This is not a violation of information theory, because the fact that the top eigenvector has a non-trivial correlation with the vector of true angles does not mean that all angles are recovered correctly by the eigenvector.

We turn to shed some light on why it is possible to partially recover the angles below the information theoretic bound. The main issue here is that it is perhaps too harsh to measure the success of the decoding algorithm by $P_e = \Pr\{\hat{\boldsymbol{\theta}}^n \neq \boldsymbol{\theta}^n\}$. For example, when the decoding algorithm decodes 999 angles out of $n = 1000$ correctly while making just a single mistake, we still count it as a failure. It may be more natural to consider the probability of error in the estimation of the individual angles. We proceed to show that this measure of error leads to a threshold probability which is smaller than (84) by just a constant factor.

Let $P_e^{(1)} = \Pr\{\widehat{\theta_1} \neq \theta_1\}$ be the probability of error in the estimation of θ_1. Again, we want to use Fano's inequality to bound the probability of the error by bounding the conditional entropy. A simple lower bound to the conditional entropy $H(\theta_1|\boldsymbol{\delta}^m)$ is obtained by conditioning on the remaining $n - 1$ angles

$$H(\theta_1|\delta^m) \geq H(\theta_1|\delta^m, \theta_2, \theta_3, \ldots, \theta_n). \tag{85}$$

Suppose that there are d_1 noisy offset measurements of the form $\theta_1 - \theta_j$, that is, d_1 is the degree of node 1 in the measurement graph G. Let the neighbors of node 1 be $j_1, j_2, \ldots, j_{d_1}$ with corresponding offset measurements $\delta_{1j_1}, \ldots, \delta_{1j_{d_1}}$. Given the values of all other angles $\theta_2, \ldots, \theta_n$, and in particular the values of $\theta_{j_1}, \ldots, \theta_{j_{d_1}}$, these d_1 equations become noisy equations for the single variable θ_1. We denote these transformed equations for θ_1 alone by $\tilde{\delta}_1, \ldots, \tilde{\delta}_{d_1}$. All other $m - d_1$ equations do not involve θ_1 and therefore do not carry any information on its value. It follows that

$$H(\theta_1|\delta^m, \theta_2, \theta_3, \ldots, \theta_n) = H(\theta_1|\tilde{\delta}_1, \ldots, \tilde{\delta}_{d_1}). \tag{86}$$

We have

$$H(\tilde{\delta}_1, \ldots, \tilde{\delta}_{d_1}|\theta_1) = d_1 H(L, p), \tag{87}$$

because given θ_1 these d_1 equations are i.i.d random variables with entropy $H(L, p)$. Also, a simple upper bound on the d_1 equations (without conditioning) is given by

$$H(\tilde{\delta}_1, \ldots, \tilde{\delta}_{d_1}) \leq d_1 \log_2 L, \tag{88}$$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

ignoring possible dependencies among the outcomes. From (87)–(88) we get an upper bound for the mutual information between θ_1 and the transformed equations

$$I(\theta_1;\tilde{\delta}_1,\ldots,\tilde{\delta}_{d_1}) \le d_1[\log_2 L - H(L,p)] = d_1 I(L,p). \tag{89}$$

Combining (85),(86) (89) and (63) we get

$$\begin{aligned} H(\theta_1|\delta^m) &\ge H(\theta_1|\delta^m,\theta_2,\theta_3,\ldots,\theta_n) \\ &= H(\theta_1|\tilde{\delta}_1,\ldots,\tilde{\delta}_{d_1}) \\ &= H(\theta_1) - I(\theta_1;\tilde{\delta}_1,\ldots,\tilde{\delta}_{d_1}) \\ &\ge \log_2 L - d_1 I(L,p). \end{aligned} \tag{90}$$

This lower bound on the conditional entropy translates, via Fano's inequality, to a lower bound on the probability of error $P_e^{(1)}$, and it follows that

$$d_1 I(L,p) > \log_2 L \tag{91}$$

is a necessary condition for having a small $P_e^{(1)}$. Similarly, the condition for a small probability of error in decoding θ_i is

$$d_i I(L,p) > \log_2 L, \tag{92}$$

where d_i is the degree of vertex i in the measurement graph. This condition suggests that we should have more success in decoding angles of high degree. The average degree $\bar{d}$ in a graph with n vertices and m edges is $\bar{d} = \frac{2m}{n}$. The condition for successful decoding of angles with degree $\bar{d}$ is

$$\frac{2m}{n} I(L,p) > \log_2 L. \tag{93}$$

In particular, this would be the condition for all vertices in a regular graph, or in a graph whose degree distribution is concentrated near $\bar{d}$.

Substituting the Taylor expansion (82) into (93) results in the condition

$$p > \sqrt{\frac{n}{m}\frac{\log_2 L}{L-1}}. \tag{94}$$

This means that successful decoding of the individual angles may be possible already for $p > p_c^{ind}$, where

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$p_c^{ind} = \sqrt{\frac{n}{m}\frac{\log_2 L}{L-1}}, \tag{95}$$

but the estimation of the individual angles must contain some error when $p < p_c^{ind}$. Note that $p_c^{ind} < p_c^{inf}$, so while for p values between p_c^{ind} and p_c^{inf} it is impossible to successfully decode all angles, it may still be possible to decode some angles.

In the complete graph case, comparing the threshold probability of the eigenvector method $p_c^{eig} = \frac{1}{\sqrt{n}}$ given by (36) and the information theoretic threshold probability p_c^{ind} (95) below which no algorithm can successfully recover individual angles, we find that their ratio is asymptotically independent of n and m:

$$\frac{p_c^{eig}}{p_c^{ind}} = \sqrt{\frac{L-1}{2\log_2 L}} + o(1). \tag{96}$$

Note that the threshold probability p_c^{eig} is smaller than p_c^{ind} for $L \leq 6$. Thus, we may regard the eigenvector method as a very successful recovery algorithm for offset equations with a small modulo L.

For $L \geq 7$, equation (96) implies a gap between the threshold probabilities p_c^{eig} and p_c^{ind}, suggesting that the exhaustive exponential search for the maximum likelihood would perform better than the polynomial time eigenvector method. Note, however, that the gap would be significant only for very large values of L that correspond to very fine angular resolutions. For example, even for $L = 100$ the threshold probability of the eigenvector method would only be $\sqrt{\frac{99}{2\log_2 100}} \approx 2.73$ times larger than that of the maximum likelihood. The exponential complexity of $O(mL^n)$ of the exhaustive search for the maximum likelihood makes it impractical even for moderate-scale problems. On the other hand, the eigenvector method has a polynomial running time and it can handle large scale problems with relative ease.

6 Connection with Max-2-Lin mod *L* and Unique Games

The angular synchronization problem is related to the combinatorial optimization problem Max-2-Lin mod L for maximizing the number of satisfied linear equations mod L with exactly 2 variables in each equation, because the discretized offset equations $\theta_i - \theta_j = \delta_{ij}$ mod L are exactly of this form. Max-2-Lin mod L is a problem mainly studied in theoretical computer science, where we prefer using the notation "mod L" instead of the more common "mod p", to avoid confusion between the size of the modulus and the proportion of good measurements.

Note that a random assignment of the angles would satisfy a $\frac{1}{L}$ fraction of the offset equations. Andersson, Engebretsen, and Håstad [3] considered SDP based algorithms for Max-2-Lin mod L, and showed that they could obtain an $\frac{1}{L}(1+\kappa(L))$-approximation algorithm, where $\kappa(L) > 0$ is a constant that depends on L. In particular, they gave a very weak proven performance guarantee of $\frac{1}{L}(1+10^{-8})$, though they concluded that it is most likely that their bounds can be improved significantly. Moreover, for $L = 3$ they numerically find the approximation ratio to be $\frac{1}{1.27} \approx 0.79$, and later Goemans and Williamson [19] proved

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

a 0.793733-approximation. The SDP based algorithms in [3] are similar in their formulation to the SDP based algorithm of Frieze and Jerrum for Max-k-Cut [14], but with a different rounding procedure. In these SDP models, L vectors are assigned to each of the n angle variables, so that the total number of vectors is nL. The resulting $nL \times nL$ matrix of inner products is required to be semidefinite positive, along with another set of $O(n^2L^2)$ linear and inequality constraints. Due to the large size of the inner product matrix and the large number of constraints, our numerical experiments with these SDP models were limited to relatively small size problems (such as $n = 20$ and $L = 7$) from which it was difficult to get a good understanding of their performance. In the small scale problems that we did manage to test, we did not find any supporting evidence that these SDP algorithms perform consistently better than the eigenvector method, despite their extensive running times and memory requirements. For our SDP experiments we used the software SDPT3 [35,37] and SDPLR [5] in MATLAB. In [3] it is also shown that it is NP-hard to approximate Max-2-Lin mod L within a constant ratio, independent of L. Thus, we should expect an L-dependent gap similar to (96) for any polynomial time algorithm, not just for the eigenvector method.

Max-2-Lin is an instance of what is known as unique games [10], described below. One distinguishing feature of the offset equations is that every constraint corresponds to a bijection between the values of the associated variables. That is, for every possible value of θ_i, there is a unique value of θ_j that satisfies the constraint $\theta_i - \theta_j = \delta_{ij}$. Unique games are systems of constraints, a generalization of the offset equations, that have this uniqueness property, so that every constraint corresponds to some permutation.

As in the setting of offset equations, instances of unique games where all constraints are satisfiable are easy to handle. Given an instance where $1 - \varepsilon$ fraction of constraints are satisfiable, the Unique Games Conjecture (UGC) of Khot [26] says that it is hard to satisfy even a $\gamma > 0$ fraction of the constraints. The UGC has been shown to imply a number of inapproximability results for fundamental problems that seem difficult to obtain by more standard complexity assumptions. Note that in our angular synchronization problem the fraction of constraints that are satisfiable is $1 - \varepsilon = p + \frac{1-p}{L}$.

Charikar, Makarychev and Makarychev [6] presented improved approximation algorithms for unique games. For instances with domain size L where the optimal solution satisfies $1 - \varepsilon$ fraction of all constraints, their algorithms satisfy roughly $L^{-\varepsilon/(2-\varepsilon)}$ and $1 - O(\sqrt{\varepsilon \log L})$ fraction of all constraints. Their algorithms are based on SDP, also with an underlying inner products matrix of size $nL \times nL$, but their constraints and rounding procedure are different than those of [3]. Given the results of [27], the algorithms in [6] are near optimal if the UGC is true, that is, any improvement (beyond low order terms) would refute the conjecture. We have not tested their SDP based algorithm in practice, because, like the SDP of [3] it is also expected to be limited to relatively small scale problems.

7 Summary and Further Applications

In this paper we presented an eigenvector method and an SDP approach for solving the angular synchronization problem. We used random matrix theory to prove that the eigenvector method finds an accurate estimate for the angles even in the presence of a large number of outlier measurements.

The idea of synchronization by eigenvectors can be applied to other problems exhibiting a group structure and noisy measurements of ratios of group elements. In this paper we specialized the synchronization problem over the group SO(2). In the general case we may consider a group G other than SO(2) for which we have good and bad measurements g_{ij} of ratios between group elements

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$g_{ij} = g_i g_j^{-1}, \quad g_i, g_j \in G. \tag{97}$$

For example, in the general case, the triplet consistency relation (17) simply reads

$$g_{ij} g_{jk} g_{ki} = g_i g_j^{-1} g_j g_k^{-1} g_k g_i^{-1} = e, \tag{98}$$

where e is the identity element of G.

Whenever the group G is compact and has a complex or real representation (for example, the rotation group SO(3) has a real representation using 3×3 rotation matrices), we may construct an Hermitian matrix that is a matrix of matrices: the ij element is either the matrix representation of the measurement g_{ij} or the zero matrix if there is no direct measurement for the ratio of g_i and g_j. Once the matrix is formed, one can look for its top eigenvectors (or SDP) and estimate the group elements from them.

In some cases the eigenvector and the SDP methods can be applied even when there is only partial information for the group ratios. This problem arises naturally in the determination of the three-dimensional structure of a macromolecule in cryo-electron microscopy [12]. In [32] we show that the common lines between projection images give partial information for the group ratios between elements in SO(3) that can be estimated accurately using the eigenvector and SDP methods. In [33] we explore the close connection between the angular synchronization problem and the class averaging problem in cryo-electron microscopy [12]. Other possible applications of the synchronization problem over SO(3) include the distance geometry problem in NMR spectroscopy [42,21] and the localization of sensor networks [4,31].

The eigenvector method can also be applied to non-compact groups that can be "compactified". For example, consider the group of real numbers $\mathbb{R}$ with addition. One may consider the synchronization problem of clocks that measure noisy time differences of the form

$$t_i - t_j = t_{ij}, \quad t_i, t_j \in \mathbb{R}. \tag{99}$$

We compactify the group $\mathbb{R}$ by mapping it to the unit circle $t \mapsto e^{\imath\omega t}$, where $\omega \in \mathbb{R}$ is a parameter to be chosen not too small and not too large, as we now explain. There may be two kinds of measurement errors in (99). The first kind of error is a small discretization error (e.g., a small Gaussian noise) of typical size Δ. The second type of error is a large error that can be regarded as an outlier. For example, in some practical application an error of size 10Δ may be considered as an outlier. We therefore want ω to satisfy $\omega \gg (1/10)\Delta^{-1}$ (not too small) and $\omega \ll \Delta^{-1}$ (not too large), so that when constructing the matrix

$$H_{ij} = \begin{cases} e^{\imath\omega t_{ij}} & \{i,j\} \in E, \\ 0 & \{i,j\} \notin E, \end{cases} \tag{100}$$

each good equation will contribute approximately 1, while the contribution of the bad equations will be uniformly distributed on the unit circle. One may even try several different

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

values for the "frequency" ω in analogy to the Fourier transform. An overdetermined linear system of the form (99) can also be solved using least squares, which is also the maximum likelihood estimator if the measurement errors are Gaussian. However, in the many outliers model, the contribution of outlier equations will dominate the sum of squares error. For example, each outlier equation with error 10Δ contributes to the sum of squares error the same as 100 good equations with error Δ. The compactification of the group combined with the eigenvector method has the appealing effect of reducing the impact of the outlier equations. This may open the way for the eigenvector method based on (100) to be useful for the surface reconstruction problems in computer vision [13,1] and optics [30] in which current methods succeed only in the presence of a limited number of outliers.

Acknowledgments

The author would like to thank Yoel Shkolniskly, Fred Sigworth and Ronald Coifman for many stimulating discussions regarding the cryo-electron microscopy problem; Boaz Barak for references to the vast literature on Max-2-Lin mod *L* and unique games; Amir Bennatan for pointers to the weak and strong converse theorems to the coding theorem; Robert Ghrist and Michael Robinson for valuable discussions at UPenn and for the reference to [17]; and Steven (Shlomo) Gortler, Yosi Keller and Ben Sonday for reviewing an earlier version of the manuscript and for their helpful suggestions.

The project described was supported by Award Number DMS-0914892 from the NSF, by Award Number FA9550-09-1-0551 from AFOSR, and by Award Number R01GM090200 from the National Institute of General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

References

1. Agrawal, AK.; Raskar, R.; Chellappa, R. What is the range of surface reconstructions from a gradient field?. Computer Vision – ECCV 2006: 9th European Conference on Computer Vision; Graz, Austria. May 7–13, 2006; 2006. Proceedings, Part IV (Lecture Notes in Computer Science); p. 578-591.
2. Alon N, Krivelevich M, Vu VH. On the concentration of eigenvalues of random symmetric matrices. Israel Journal of Mathematics 2002;131(1):259–267.
3. Andersson, G.; Engebretsen, L.; Håstad, J. A new way to use semidefinite programming with applications to linear equations mod. Proceedings 10th annual ACM-SIAM symposium on Discrete algorithms; 1999. p. 41-50.
4. Biswas P, Liang TC, Toh KC, Wang TC, Ye Y. Semidefinite programming approaches for sensor network localization with noisy distance measurements. IEEE Transactions on Automation Science and Engineering 2006;3(4):360–371.
5. Burer S, Monteiro RDC. A Nonlinear Programming Algorithm for Solving Semidefinite Programs Via Low-Rank Factorization. Mathematical Programming (series B) 2003;95 (2):329–357.
6. Charikar, M.; Makarychev, K.; Makarychev, Y. Near-Optimal Algorithms for Unique Games. Proceedings 38th annual ACM symposium on Theory of computing; 2006. p. 205-214.
7. Coifman RR, Lafon S. Diffusion maps. Applied and Computational Harmonic Analysis 2006;21(1): 5–30.
8. Cover, TM.; Thomas, JA. Elements of Information Theory. Wiley; New York: 1991.
9. Erdős P, Rényi A. On random graphs. Publicationes Mathematicae 1959;6:290–297.
10. Feige, U.; Lovász, L. Two-prover one round proof systems: Their power and their problems. Proceedings of the 24th ACM Symposium on Theory of Computing; 1992. p. 733-741.
11. Féral D, Péché S. The Largest Eigenvalue of Rank One Deformation of Large Wigner Matrices. Communications in Mathematical Physics 2007;272(1):185–228.
12. Frank, J. Three-Dimensional Electron Microscopy of Macro-molecular Assemblies: Visualization of Biological Molecules in Their Native State. Oxford: 2006.
13. Frankot RT, Chellappa R. A method for enforcing integrability in shape from shading algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 1988;10(4):439–451.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

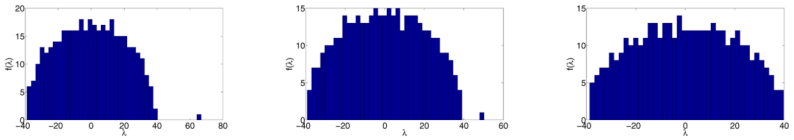
14. Frieze A, Jerrum M. Improved Approximation Algorithms for MAX k-CUT and MAX BISECTION. Algorithmica 1997;18(1):67–81.
15. Füredi Z, Komlós J. The eigenvalues of random symmetric matrices. Combinatorica 1981;1:233–241.
16. Gallager, RG. Information Theory and Reliable Communication. Wiley; New York: 1968.
17. Giridhar, A.; Kumar, PR. Distributed Clock Synchronization over Wireless Networks: Algorithms and Analysis. 45th IEEE Conference on Decision and Control 2006; 2006. p. 4915-4920.
18. Goemans MX, Williamson DP. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. Journal of the ACM (JACM) 1995;42(6): 1115–1145.
19. Goemans, MX.; Williamson, DP. Approximation algorithms for Max-3-Cut and other problems via complex semidefinite programming. Proceedings 33rd annual ACM symposium on Theory of computing; 2001. p. 443-452.
20. Griffiths, DJ. Introduction to Quantum Mechanics. Prentice Hall; NJ: 1994. 416 Pages
21. Havel TF, Wuthrich K. An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformation in solution. J Mol Biol 1985;182:281–294. [PubMed: 2582141]
22. Horn, RA.; Johnson, CR. Matrix Analysis. Cambridge University Press; 1990. 575 pages
23. Karp, R.; Elson, J.; Estrin, D.; Shenker, S. Technical Report. Center for Embedded Networked Sensing, University of California; Los Angeles: 2003. Optimal and global time synchronization in sensornets.
24. Khorunzhiy O. Rooted trees and moments of large sparse random matrices. Discrete Mathematics and Theoretical Computer Science AC 2003:145–154.
25. Khorunzhy A. Sparse Random Matrices: Spectral Edge and Statistics of Rooted Trees. Advances in Applied Probability 2001;33(1):124–140.
26. Khot, S. On the power of unique 2-prover 1-round games. Proceedings of the ACM Symposium on the Theory of Computing; 2002. p. 767-775.
27. Khot S, Kindler G, Mossel E, O'Donnell R. Optimal inapproximability results for MAX-CUT and other two-variable CSPs? SIAM Journal of Computing 2007;37(1):319–357.
28. Natterer, F. The Mathematics of Computerized Tomography. SIAM: Society for Industrial and Applied Mathematics, Classics in Applied Mathematics; 2001.
29. Péché S. The largest eigenvalues of small rank perturbations of Hermitian random matrices. Prob Theo Rel Fields 2006;134 (1):127–174.
30. Rubinstein J, Wolansky G. Reconstruction of optical surfaces from ray data. Optical Review 2001;8(4):281–283.
31. Singer A. A Remark on Global Positioning from Local Distances. Proceedings of the National Academy of Sciences 2008;105 (28):9507–9511.
32. Singer A, Shkolnisky Y. Three-Dimensional Structure Determination from Common Lines in Cryo-EM by Eigenvectors and Semidefinite Programming. (submitted for publication).
33. Singer A, Shkolnisky Y, Hadani R. Viewing Angle Classification of Cryo-Electron Microscopy Images using Eigenvectors. (submitted for publication).
34. Soshnikov A. Universality at the edge of the spectrum in Wigner random matrices. Comm Math Phys 1999;207:697–733.
35. Toh KC, Todd MJ, Tutuncu RH. SDPT3 — a Matlab software package for semidefinite programming. Optimization Methods and Software 1999;11:545–581.
36. Tracy CA, Widom H. Level-spacing distributions and the Airy kernel. Communications in Mathematical Physics 1994;159(1):151–174.
37. Tutuncu RH, Toh KC, Todd MJ. Solving semidefinite-quadratic-linear programs using SDPT3. Mathematical Programming Ser B 2003;95:189–217.
38. Vandenberghe L, Boyd S. Semidefinite programming. SIAM Review 1996;38(1):49–95.
39. Watts DJ, Strogatz SH. Collective dynamics of small-world networks. Nature 1998;393:440–442. [PubMed: 9623998]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

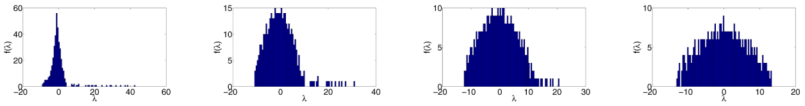
40. Wigner EP. Characteristic vectors of bordered matrices with infinite dimensions. Annals of Mathematics 1955;62:548–564.
41. Wigner EP. On the distribution of tile roots of certain symmetric matrices. Annals of Mathematics 1958;67:325–328.
42. Wuthrich K. NMR studies of structure and function of biological macromolecules (Nobel Lecture). J Biomol NMR 2003;27:13–39. [PubMed: 15143746]



(a) $p = 0.15$ (b) $p = 0.1$ (c) $p = 0.05$

Figure 1.
Histogram of the eigenvalues of the matrix H in the complete graph model for $n = 400$ and different values of p.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

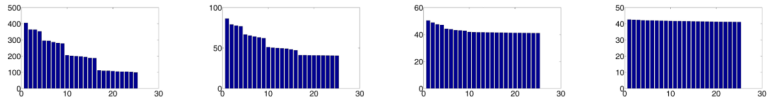


(a) $p = 1$ (b) $p = 0.7$ (c) $p = 0.4$ (d) $p = 0.1$

Figure 2.

Histogram of the eigenvalues of the matrix H in the small-world model for $n = 400$, $\varepsilon = 0.2$, $m \approx 8000$, and different values of p.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript



(a) $p = 1$ (b) $p = 0.2$ (c) $p = 0.1$ (d) $p = 0.05$

Figure 3.

Bar plot of the 25 largest eigenvalues of the matrix H in the small-world model for $n = 4000$, $\varepsilon = 0.2$, $m \approx 8 \cdot 10^5$, and different values of p. The multiplicities 1, 3, 5, 7, 9 corresponding to the spherical harmonics are evident as long as p is not too small. As p decreases, the high-oscillatory spherical harmonics are getting "swallowed" by the semi-circle.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Correlations between the top eigenvector v_1 of H and the vector z of true angles for different values of p in the complete graph model.

(a) $n = 100$					(b) $n = 400$				
p	np^2	$\left(1+\frac{1}{np^2}\right)^{-\frac{1}{2}}$	ρ_1	ρ_2	p	np^2	$\left(1+\frac{1}{np^2}\right)^{-\frac{1}{2}}$	ρ_1	ρ_2
0.4	16	0.97	0.99	0.98	0.2	16	0.97	0.99	0.97
0.3	9	0.95	0.97	0.95	0.15	9	0.95	0.97	0.95
0.2	4	0.89	0.90	0.88	0.1	4	0.89	0.90	0.87
0.15	2.25	0.83	0.75	0.81	0.075	2.25	0.83	0.77	0.76
0.1	1	0.71	0.34	0.35	0.05	1	0.71	0.28	0.32
0.05	0.25	0.45	0.13	0.12	0.025	0.25	0.45	0.06	0.07

Appl Comput Harmon Anal. Author manuscript; available in PMC 2012 January 30.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 2

Correlations between the top eigenvector of H and the vector of true angles for different values of p in the small-world S^2 model.

(a) $n = 100$, $\varepsilon = 0.3$, $m \approx 750$			(b) $n = 400$, $\varepsilon = 0.2$, $m \approx 8000$		
p	$\frac{2mp^2}{n}$	ρ_1	p	$\frac{2mp^2}{n}$	ρ_1
0.8	9.6	0.923	0.8	26	0.960
0.6	5.4	0.775	0.4	6.4	0.817
0.4	2.4	0.563	0.3	3.6	0.643
0.3	1.4	0.314	0.2	1.6	0.282
0.2	0.6	0.095	0.1	0.4	0.145

Appl Comput Harmon Anal. Author manuscript; available in PMC 2012 January 30.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Comparison between the correlations obtained by the eigenvector method ρ_{eig}, by the SDP method ρ_{sdp} and by the least squares method ρ_{lsqr} for different values of p (small world graph on S^2, $n = 200$, $\varepsilon = 0.3$, $m \approx 3000$). The SDP tends to find low-rank matrices despite the fact that the rank-one constraint on Θ is not included in the SDP. The rightmost column gives the rank of the Θ matrices that were found by the SDP. To solve the SDP (8)–(10) we used SDPLR, a package for solving large-scale SDP problems [5]. The least squares solution was obtained using MATLAB's lsqr function. As expected, the least squares method yields poor correlations compared to the eigenvector and the SDP methods.

p	ρ_{lsqr}	ρ_{eig}	ρ_{sdp}	rank Θ
1	1	1	1	1
0.7	0.787	0.977	0.986	1
0.4	0.046	0.839	0.893	3
0.3	0.103	0.560	0.767	3
0.2	0.227	0.314	0.308	4
0.15	0.091	0.114	0.102	5