

18.S096: Synchronization Problems and Alignment

Topics in Mathematics of Data Science (Fall 2015)

Afonso S. Bandeira

bandeira@mit.edu

<http://math.mit.edu/~bandeira>

December 8, 2015

These are lecture notes not in final form and will be continuously edited and/or corrected (as I am sure it contains many typos). Please let me know if you find any typo/mistake. Also, I am posting the open problems on my Blog, see [Ban15c].

10.1 Synchronization-type problems

This section will focus on synchronization-type problems.¹ These are problems where the goal is to estimate a set of parameters from data concerning relations or interactions between pairs of them. A good example to have in mind is an important problem in computer vision, known as structure from motion: the goal is to build a three-dimensional model of an object from several two-dimensional photos of it taken from unknown positions. Although one cannot directly estimate the positions, one can compare pairs of pictures and gauge information on their relative positioning. The task of estimating the camera locations from this pairwise information is a synchronization-type problem. Another example, from signal processing, is multireference alignment, which is the problem of estimating a signal from measuring multiple arbitrarily shifted copies of it that are corrupted with noise.

We will formulate each of these problems as an estimation problem on a graph $G = (V, E)$. More precisely, we will associate each data unit (say, a photo, or a shifted signal) to a graph node $i \in V$. The problem can then be formulated as estimating, for each node $i \in V$, a group element $g_i \in \mathcal{G}$, where the group \mathcal{G} is a group of transformations, such as translations, rotations, or permutations. The pairwise data, which we identify with edges of the graph $(i, j) \in E$, reveals information about the ratios $g_i(g_j)^{-1}$. In its simplest form, for each edge $(i, j) \in E$ of the graph, we have a noisy estimate of $g_i(g_j)^{-1}$ and the synchronization problem consists of estimating the individual group elements $g : V \rightarrow \mathcal{G}$ that are the most consistent with the edge estimates, often corresponding to the Maximum Likelihood (ML) estimator. Naturally, the measure of “consistency” is application specific. While there is a general way of describing these problems and algorithmic approaches to them [BCS15, Ban15a], for the sake of simplicity we will illustrate the ideas through some important examples.

¹And it will follow somewhat the structure in Chapter 1 of [Ban15a]

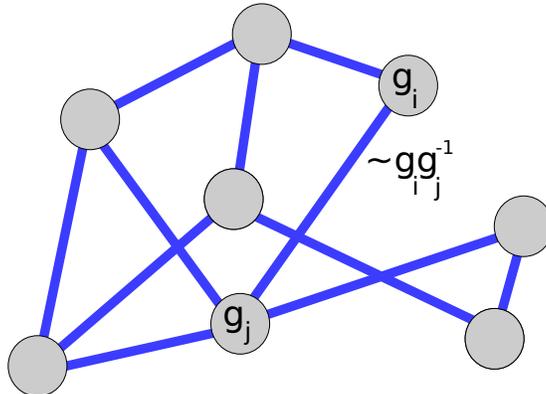


Figure 1: Given a graph $G = (V, E)$ and a group \mathcal{G} , the goal in synchronization-type problems is to estimate node labels $g : V \rightarrow \mathcal{G}$ from noisy edge measurements of offsets $g_i g_j^{-1}$.

10.2 Angular Synchronization

The angular synchronization problem [Sin11, BSS13] consist in estimating n unknown angles $\theta_1, \dots, \theta_n$ from m noisy measurements of their offsets $\theta_i - \theta_j \pmod{2\pi}$. This problem easily falls under the scope of synchronization-type problem by taking a graph with a node for each θ_i , an edge associated with each measurement, and taking the group to be $\mathcal{G} \cong SO(2)$, the group of in-plane rotations. Some of its applications include time-synchronization of distributed networks [GK06], signal reconstruction from phaseless measurements [ABFM12], surface reconstruction problems in computer vision [ARC06] and optics [RW01].

Let us consider a particular instance of this problem (with a particular noise model).

Let $z_1, \dots, z_n \in \mathbb{C}$ satisfying $|z_a| = 1$ be the signal (angles) we want to estimate ($z_a = \exp(i\theta_a)$). Suppose for every pair (i, j) we make a noisy measurement of the angle offset

$$Y_{ij} = z_i \bar{z}_j + \sigma W_{ij},$$

where $W_{ij} \sim \mathcal{N}(0, 1)$. The maximum likelihood estimator for z is given by solving (see [Sin11, BBS14])

$$\max_{|x_i|=1} x^* Y x. \tag{1}$$

There are several approaches to try to solve (1). Using techniques very similar to the study of the spike model in PCA on the first lecture one can (see [Sin11]), for example, understand the performance of the spectral relaxation of (1) into

$$\max_{\|x\|^2=n} x^* Y x. \tag{2}$$

Notice that, since the solution to (2) will not necessarily be a vector with unit-modulus entries, a rounding step will, in general, be needed. Also, to compute the leading eigenvector of A one would

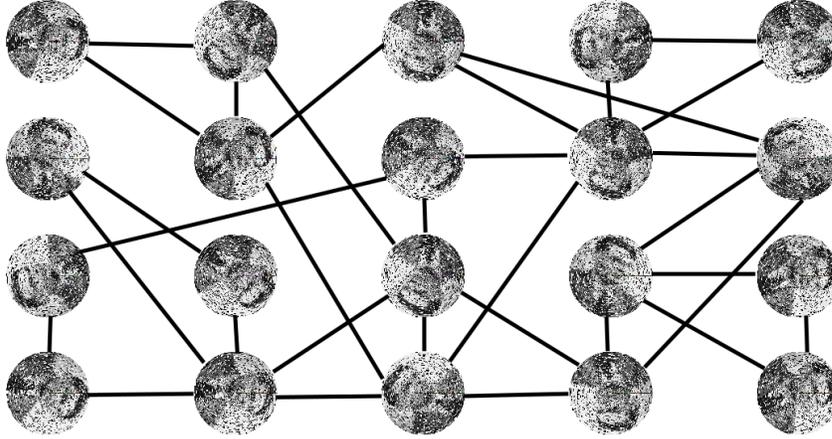


Figure 2: An example of an instance of a synchronization-type problem. Given noisy rotated copies of an image (corresponding to vertices of a graph), the goal is to recover the rotations. By comparing pairs of images (corresponding to edges of the graph), it is possible to estimate the relative rotations between them. The problem of recovering the rotation of each image from these relative rotation estimates is an instance of Angular synchronization.

likely use the power method. An interesting adaptation to this approach is to round after each iteration of the power method, rather than waiting for the end of the process, more precisely:

Algorithm 10.1 *Given Y . Take a original (maybe random) vector $x^{(0)}$. For each iteration k (until convergence or a certain number of iterations) take $x^{(k+1)}$ to be the vector with entries:*

$$\left(x^{(k+1)}\right)_i = \frac{(Yx^{(k)})_i}{|(Yx^{(k)})_i|}.$$

Although this method appears to perform very well in numeric experiments, its analysis is still an open problem.

Open Problem 10.1 *In the model where $Y = zz^* + \sigma W$ as described above, for which values of σ will the Projected Power Method (Algorithm 10.1) converge to the optimal solution of (1) (or at least to a solution that correlates well with z), with high probability?*

We note that Algorithm 10.1 is very similar to the Approximate Message Passing method presented, and analyzed, in [MR14] for the positive eigenvector problem.

Another approach is to consider an SDP relaxation similar to the one for Max-Cut and minimum bisection.

$$\begin{aligned}
\max \quad & \text{Tr}(YX) \\
\text{s.t.} \quad & X_{ii} = 1, \forall_i \\
& X \succeq 0.
\end{aligned} \tag{3}$$

In [BBS14] it is shown that, in the model of $Y = zz^*zz^* + \sigma W$, as long as $\sigma = \tilde{O}(n^{1/4})$ then (3) is tight, meaning that the optimal solution is rank 1 and thus it corresponds to the optimal solution of (1).² It is conjecture [BBS14] however that $\sigma = \tilde{O}(n^{1/2})$ should suffice. It is known (see [BBS14]) that this is implied by the following conjecture:

If x^\natural is the optimal solution to (1), then with high probability $\|Wx^\natural\|_\infty = \tilde{O}(n^{1/2})$. This is the content of the next open problem.

Open Problem 10.2 *Prove or disprove: With high probability the SDP relaxation (3) is tight as long as $\sigma = \tilde{O}(n^{1/2})$. This would follow from showing that, with high probability $\|Wx^\natural\|_\infty = \tilde{O}(n^{1/2})$, where x^\natural is the optimal solution to (1).*

We note that the main difficulty seems to come from the fact that W and x^\natural are not independent random variables.

10.2.1 Orientation estimation in Cryo-EM

A particularly challenging application of this framework is the orientation estimation problem in Cryo-Electron Microscopy [SS11].

Cryo-EM is a technique used to determine the three-dimensional structure of biological macromolecules. The molecules are rapidly frozen in a thin layer of ice and imaged with an electron microscope, which gives 2-dimensional projections. One of the main difficulties with this imaging process is that these molecules are imaged at different unknown orientations in the sheet of ice and each molecule can only be imaged once (due to the destructive nature of the imaging process). More precisely, each measurement consists of a tomographic projection of a rotated (by an unknown rotation) copy of the molecule. The task is then to reconstruct the molecule density from many such measurements. As the problem of recovering the molecule density knowing the rotations fits in the framework of classical tomography—for which effective methods exist—the problem of determining the unknown rotations, the orientation estimation problem, is of paramount importance. While we will not go into details here, there is a mechanism that, from two such projections, obtains information between their orientation. The problem of finding the orientation of each projection from such pairwise information naturally fits in the framework of synchronization and some of the techniques described here can be adapted to this setting [BCS15].

10.2.2 Synchronization over \mathbb{Z}_2

This particularly simple version already includes many applications of interest. Similarly to before, given a graph $G = (V, E)$, the goal is recover unknown node labels $g : V \rightarrow \mathbb{Z}_2$ (corresponding to

²Note that this makes (in this regime) the SDP relaxation a Probably Certifiably Correct algorithm [Ban15b]

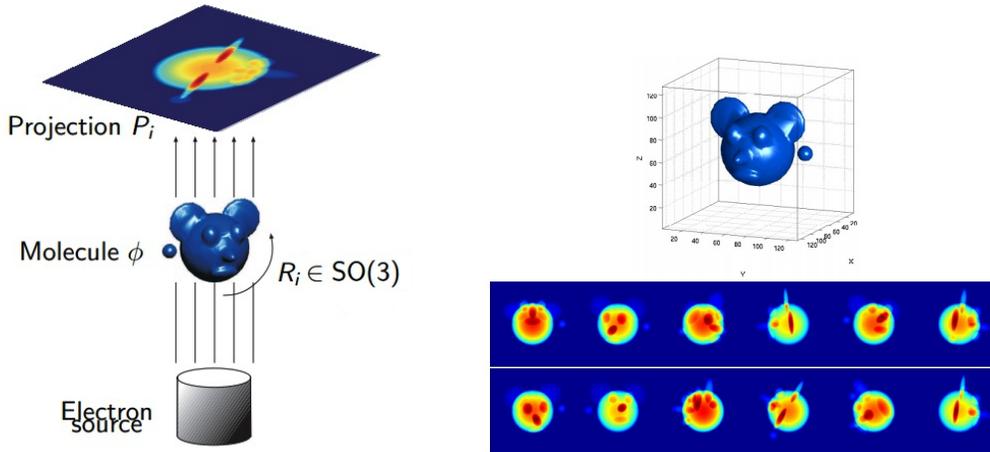


Figure 3: Illustration of the Cryo-EM imaging process: A molecule is imaged after being frozen at a random (unknown) rotation and a tomographic 2-dimensional projection is captured. Given a number of tomographic projections taken at unknown rotations, we are interested in determining such rotations with the objective of reconstructing the molecule density. Images courtesy of Amit Singer and Yoel Shkolnisky [SS11].

memberships to two clusters) from pairwise information. Each pairwise measurement either suggests the two involved nodes are in the same cluster or in different ones (recall the problem of recovery in the stochastic block model). The task of clustering the graph in order to agree, as much as possible, with these measurements is tightly connected to *correlation clustering* [BBC04] and has applications to determining the orientation of a manifold [SW11].

In the case where all the measurements suggest that the involved nodes belong in different communities, then this problem essentially reduces to the **Max-Cut** problem.

10.3 Signal Alignment

In signal processing, the multireference alignment problem [BCSZ14] consists of recovering an unknown signal $u \in \mathbb{R}^L$ from n observations of the form

$$y_i = R_{l_i} u + \sigma \xi_i, \tag{4}$$

where R_{l_i} is a circulant permutation matrix that shifts u by $l_i \in \mathbb{Z}_L$ coordinates, ξ_i is a noise vector (which we will assume standard gaussian i.i.d. entries) and l_i are unknown shifts.

If the shifts were known, the estimation of the signal u would reduce to a simple denoising problem. For that reason, we will focus on estimating the shifts $\{l_i\}_{i=1}^n$. By comparing two observations y_i and y_j we can obtain information about the relative shift $l_i - l_j \pmod L$ and write this problem as a Synchronization problem

10.3.1 The model bias pitfall

In some of the problems described above, such as the multireference alignment of signals (or the orientation estimation problem in Cryo-EM), the alignment step is only a subprocedure of the estimation of the underlying signal (or the 3d density of the molecule). In fact, if the underlying signal was known, finding the shifts would be nearly trivial: for the case of the signals, one could simply use match-filtering to find the most likely shift l_i for measurement y_i (by comparing all possible shifts of it to the known underlying signal).

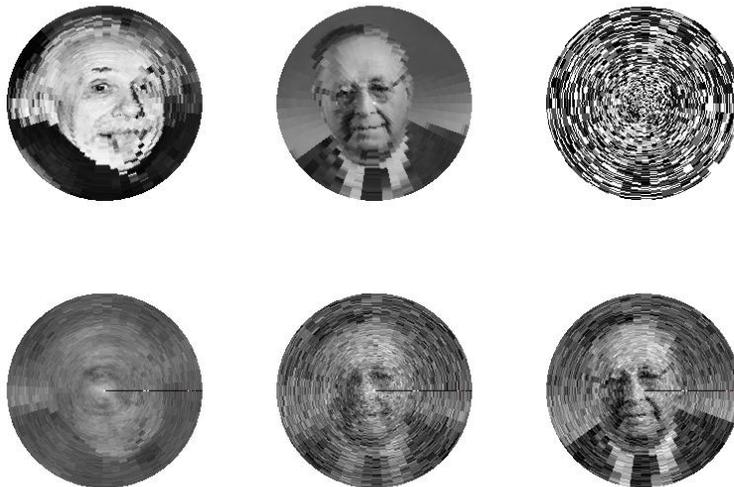


Figure 4: A simple experiment to illustrate the model bias phenomenon: Given a picture of the mathematician Hermann Weyl (second picture of the top row) we generate many images consisting of random rotations (we considered a discretization of the rotations of the plane) of the image with added gaussian noise. An example of one such measurements is the third image in the first row. We then proceeded to align these images to a reference consisting of a famous image of Albert Einstein (often used in the model bias discussions). After alignment, an estimator of the original image was constructed by averaging the aligned measurements. The result, first image on second row, clearly has more resemblance to the image of Einstein than to that of Weyl, illustration the model bias issue. One the other hand, the method based on the synchronization approach produces the second image of the second row, which shows no signs of suffering from model bias. As a benchmark, we also include the reconstruction obtained by an oracle that is given the true rotations (third image in the second row).

When the true signal is not known, a common approach is to choose a reference signal z that is not the true template but believed to share some properties with it. Unfortunately, this creates a high risk of model bias: the reconstructed signal \hat{u} tends to capture characteristics of the reference z that are not present on the actual original signal u (see Figure 10.3.1 for an illustration of this phenomenon).

This issue is well known among the biological imaging community [SHBG09, Hen13] (see, for example, [Coh13] for a particularly recent discussion of it). As the experiment shown on Figure 10.3.1 suggests, the methods treated in this paper, based solely on pairwise information between observations, do not suffer from model bias as they do not use any information besides the data itself.

In order to recover the shifts l_i from the shifted noisy signals (4) we will consider the following estimator

$$\operatorname{argmin}_{l_1, \dots, l_n \in \mathbb{Z}_L} \sum_{i, j \in [n]} \|R_{-l_i} y_i - R_{-l_j} y_j\|^2, \quad (5)$$

which is related to the maximum likelihood estimator of the shifts. While we refer to [Ban15a] for a derivation we note that it is intuitive that if l_i is the right shift for y_i and l_j for y_j then $R_{-l_i} y_i - R_{-l_j} y_j$ should be random gaussian noise, which motivates the estimator considered.

Since a shift does not change the norm of a vector, (5) is equivalent to

$$\operatorname{argmax}_{l_1, \dots, l_n \in \mathbb{Z}_L} \sum_{i, j \in [n]} \langle R_{-l_i} y_i, R_{-l_j} y_j \rangle, \quad (6)$$

we will refer to this estimator as the quasi-MLE.

It is not surprising that solving this problem is NP-hard in general (the search space for this optimization problem has exponential size and is nonconvex). In fact, one can show [BCSZ14] that, conditioned on the Unique Games Conjecture, it is hard to approximate up to any constant.

10.3.2 The semidefinite relaxation

We will now present a semidefinite relaxation for (6) (see [BCSZ14]).

Let us identify R_l with the $L \times L$ permutation matrix that cyclicly permutes the entries of a vector by l_i coordinates:

$$R_l \begin{bmatrix} u_1 \\ \vdots \\ u_L \end{bmatrix} = \begin{bmatrix} u_{1-l} \\ \vdots \\ u_{L-l} \end{bmatrix}.$$

This corresponds to an L -dimensional representation of the cyclic group. Then, (6) can be rewritten:

$$\begin{aligned} \sum_{i, j \in [n]} \langle R_{-l_i} y_i, R_{-l_j} y_j \rangle &= \sum_{i, j \in [n]} (R_{-l_i} y_i)^T R_{-l_j} y_j \\ &= \sum_{i, j \in [n]} \operatorname{Tr} \left[(R_{-l_i} y_i)^T R_{-l_j} y_j \right] \\ &= \sum_{i, j \in [n]} \operatorname{Tr} \left[y_i^T R_{-l_i}^T R_{-l_j} y_j \right] \\ &= \sum_{i, j \in [n]} \operatorname{Tr} \left[(y_i y_j^T)^T R_{l_i} R_{l_j}^T \right]. \end{aligned}$$

We take

$$X = \begin{bmatrix} R_{l_1} \\ R_{l_2} \\ \vdots \\ R_{l_n} \end{bmatrix} [R_{l_1}^T \quad R_{l_2}^T \quad \cdots \quad R_{l_n}^T] \in \mathbb{R}^{nL \times nL}, \quad (7)$$

and can rewrite (6) as

$$\begin{aligned} \max \quad & \text{Tr}(CX) \\ \text{s. t.} \quad & X_{ii} = I_{L \times L} \\ & X_{ij} \text{ is a circulant permutation matrix} \\ & X \succeq 0 \\ & \text{rank}(X) \leq L, \end{aligned} \quad (8)$$

where C is the rank 1 matrix given by

$$C = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} [y_1^T \quad y_2^T \quad \cdots \quad y_n^T] \in \mathbb{R}^{nL \times nL}, \quad (9)$$

with blocks $C_{ij} = y_i y_j^T$.

The constraints $X_{ii} = I_{L \times L}$ and $\text{rank}(X) \leq L$ imply that $\text{rank}(X) = L$ and $X_{ij} \in O(L)$. Since the only doubly stochastic matrices in $O(L)$ are permutations, (8) can be rewritten as

$$\begin{aligned} \max \quad & \text{Tr}(CX) \\ \text{s. t.} \quad & X_{ii} = I_{L \times L} \\ & X_{ij} \mathbf{1} = \mathbf{1} \\ & X_{ij} \text{ is circulant} \\ & X \succeq 0 \\ & X \succeq 0 \\ & \text{rank}(X) \leq L. \end{aligned} \quad (10)$$

Removing the nonconvex rank constraint yields a semidefinite program, corresponding to (??),

$$\begin{aligned} \max \quad & \text{Tr}(CX) \\ \text{s. t.} \quad & X_{ii} = I_{L \times L} \\ & X_{ij} \mathbf{1} = \mathbf{1} \\ & X_{ij} \text{ is circulant} \\ & X \succeq 0 \\ & X \succeq 0. \end{aligned} \quad (11)$$

Numerical simulations (see [BCSZ14, BKS14]) suggest that, below a certain noise level, the semidefinite program (11) is tight with high probability. However, an explanation of this phenomenon remains an open problem [BKS14].

Open Problem 10.3 *For which values of noise do we expect that, with high probability, the semidefinite program (11) is tight? In particular, is it true that for any σ by taking arbitrarily large n the SDP is tight with high probability?*

10.3.3 Sample complexity for multireference alignment

Another important question related to this problem is to understand its sample complexity. Since the objective is to recover the underlying signal u , a larger number of observations n should yield a better recovery (considering the model in (??)). Another open question is the consistency of the quasi-MLE estimator, it is known that there is some bias on the power spectrum of the recovered signal (that can be easily fixed) but the estimates for phases of the Fourier transform are conjecture to be consistent [BCSZ14].

- Open Problem 10.4** 1. *Is the quasi-MLE (or the MLE) consistent for the Multireference alignment problem? (after fixing the power spectrum appropriately).*
2. *For a given value of L and σ , how large does n need to be in order to allow for a reasonably accurate recovery in the multireference alignment problem?*

Remark 10.2 *One could design a simpler method based on angular synchronization: for each pair of signals take the best pairwise shift and then use angular synchronization to find the signal shifts from these pairwise measurements. While this would yield a smaller SDP, the fact that it is not using all of the information renders it less effective [BCS15]. This illustrates an interesting trade-off between size of the SDP and its effectiveness. There is an interpretation of this through dimensions of representations of the group in question (essentially each of these approaches corresponds to a different representation), we refer the interested reader to [BCS15] for more on that.*

References

- [ABFM12] B. Alexeev, A. S. Bandeira, M. Fickus, and D. G. Mixon. Phase retrieval with polarization. *available online*, 2012.
- [ARC06] A. Agrawal, R. Raskar, and R. Chellappa. What is the range of surface reconstructions from a gradient field? In A. Leonardis, H. Bischof, and A. Pinz, editors, *Computer Vision – ECCV 2006*, volume 3951 of *Lecture Notes in Computer Science*, pages 578–591. Springer Berlin Heidelberg, 2006.
- [Ban15a] A. S. Bandeira. *Convex relaxations for certain inverse problems on graphs*. PhD thesis, Program in Applied and Computational Mathematics, Princeton University, 2015.
- [Ban15b] A. S. Bandeira. A note on probably certifiably correct algorithms. *Available at arXiv:1509.00824 [math.OC]*, 2015.
- [Ban15c] A. S. Bandeira. Relax and Conquer BLOG: Ten Lectures and Forty-two Open Problems in Mathematics of Data Science. 2015.
- [BBC04] N. Bansal, A. Blum, and S. Chawla. Correlation clustering. *Machine Learning*, 56(1-3):89–113, 2004.
- [BBS14] A. S. Bandeira, N. Boumal, and A. Singer. Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. *Available online at arXiv:1411.3272 [math.OC]*, 2014.

- [BCS15] A. S. Bandeira, Y. Chen, and A. Singer. Non-unique games over compact groups and orientation estimation in cryo-em. *Available online at arXiv:1505.03840 [cs.CV]*, 2015.
- [BCSZ14] A. S. Bandeira, M. Charikar, A. Singer, and A. Zhu. Multireference alignment using semidefinite programming. *5th Innovations in Theoretical Computer Science (ITCS 2014)*, 2014.
- [BKS14] A. S. Bandeira, Y. Khoo, and A. Singer. Open problem: Tightness of maximum likelihood semidefinite relaxations. In *Proceedings of the 27th Conference on Learning Theory*, volume 35 of *JMLR W&CP*, pages 1265–1267, 2014.
- [BSS13] A. S. Bandeira, A. Singer, and D. A. Spielman. A Cheeger inequality for the graph connection Laplacian. *SIAM J. Matrix Anal. Appl.*, 34(4):1611–1630, 2013.
- [Coh13] J. Cohen. Is high-tech view of HIV too good to be true? *Science*, 341(6145):443–444, 2013.
- [GK06] A. Giridhar and P.R. Kumar. Distributed clock synchronization over wireless networks: Algorithms and analysis. In *Decision and Control, 2006 45th IEEE Conference on*, pages 4915–4920. IEEE, 2006.
- [Hen13] R. Henderson. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences*, 110(45):18037–18041, 2013.
- [MR14] A. Montanari and E. Richard. Non-negative principal component analysis: Message passing algorithms and sharp asymptotics. *Available online at arXiv:1406.4775v1 [cs.IT]*, 2014.
- [RW01] J. Rubinstein and G. Wolansky. Reconstruction of optical surfaces from ray data. *Optical Review*, 8(4):281–283, 2001.
- [SHBG09] M. Shatsky, R. J. Hall, S. E. Brenner, and R. M. Glaeser. A method for the alignment of heterogeneous macromolecules from electron microscopy. *Journal of Structural Biology*, 166(1), 2009.
- [Sin11] A. Singer. Angular synchronization by eigenvectors and semidefinite programming. *Appl. Comput. Harmon. Anal.*, 30(1):20 – 36, 2011.
- [SS11] A. Singer and Y. Shkolnisky. Three-dimensional structure determination from common lines in Cryo-EM by eigenvectors and semidefinite programming. *SIAM J. Imaging Sciences*, 4(2):543–572, 2011.
- [SW11] A. Singer and H.-T. Wu. Orientability and diffusion maps. *Appl. Comput. Harmon. Anal.*, 31(1):44–58, 2011.