Multireference Alignment Using Semidefinite Programming

Afonso S. Bandeira
Program in Applied and Computational Math (PACM)
Princeton University
Princeton, New Jersey, USA
ajsb@math.princeton.edu

Moses Charikar
Department of Computer Science
Princeton University
Princeton, New Jersey, USA
moses@cs.princeton.edu

Amit Singer
Department of Mathematics and PACM
Princeton University
Princeton, New Jersey, USA
amits@math.princeton.edu

Andy Zhu
Department of Mathematics
Princeton University
Princeton, New Jersey, USA
azpujps@gmail.com

ABSTRACT

The multireference alignment problem consists of estimating a signal from multiple noisy shifted observations. Inspired by existing Unique-Games approximation algorithms, we provide a semidefinite program (SDP) based relaxation which approximates the maximum likelihood estimator (MLE) for the multireference alignment problem. Although we show this MLE problem is Unique-Games hard to approximate within any constant, we observe that our poly-time approximation algorithm for this problem appears to perform quite well in typical instances, outperforming existing methods. In an attempt to explain this behavior we provide stability guarantees for our SDP under a random noise model on the observations. This case is more challenging to analyze than traditional semi-random instances of Unique-Games: the noise model is on vertices of a graph and translates into dependent noise on the edges.

Interestingly, we show that if certain positivity constraints in the relaxation are dropped, its solution becomes equivalent to performing phase correlation, a popular method used for pairwise alignment in imaging applications. Finally, we describe how symmetry reduction techniques from matrix representation theory can greatly decrease the computational cost of the SDP considered.

Categories and Subject Descriptors

F.2.0 [**Theory of Computation**]: Analysis of Algorithms and Problem Complexity—*General*; G.0 [**Mathematics of Computing**]: General

ITCS'14, January 12–14, 2014, Princeton, New Jersey, USA.

ACM 978-1-4503-2698-8/14/01 ...$15.00.
http://dx.doi.org/10.1145/2554797.2554839.

Keywords

Multireference Alignment; Semidefinite Relaxation; Phase Correlation; Unique-Games

1. INTRODUCTION

The multireference alignment problem consists of estimating an unknown signal x from multiple noisy cyclically-shifted copies. More precisely, we are interested in the problem of estimating an unknown template vector $x \in \mathbb{R}^L$ from N measurements $y_1, \ldots, y_N$ of the form:

$$y_i = R_{l_i} x + \xi_i \in \mathbb{R}^L, \tag{1}$$

where $\xi_i \sim \mathcal{N}(0, \sigma^2 I_L)$ is gaussian white noise with variance σ^2, and R_l denotes the index cyclic shift operator $(x_1, \ldots, x_L) \mapsto (x_{1-l}, \ldots, x_{L-l})$.

The difficulty of this problem resides in the fact that both the template x and the shifts $l_1, \ldots, l_n$ are unknown (moreover, no model is presumed a-priori for their distribution). If the shifts were known, one could easily estimate x by unshifting the observations and averaging them. Motivated by this fact we will focus on the problem of estimating the shifts $l_1, \ldots, l_n$ (up to a global shift).

This problem has a vast list of applications. Alignment is directly used in structural biology [8] [34]; radar [36] [29]; crystalline simulations [32]; and image registration in a number of important contexts, such as in geology, medicine, and paleontology [11] [12]. We will discuss below various methods to solve this problem, some of which are used in these communities (see also Appendix A).

Perhaps the most naïve approach to estimate the shifts in (1) would be to fix one of the observations, say y_i, as a reference template and align every other y_j with it by the shift δ_{ij} minimizing their distance

$$\delta_{ij} = \operatorname*{argmin}_{l \in \mathbb{Z}_L} \|R_l y_j - y_i\|_2. \tag{2}$$

This solution works well at a high signal-to-noise ratio (SNR), but performs poorly at low SNR. A more democratic approach would be to calculate all of the pairwise relative shift estimates δ_{ij} before attempting to recover the shifts

$\{l_i\}$. The shifts could then be estimated by minimizing

$$\min_{l_1,\ldots,l_N\in\mathbb{Z}_L} \sum_{i,j=1}^{N} \left| e\left(\frac{l_i}{L}\right) - e\left(\frac{l_j+\delta_{ij}}{L}\right)\right|^2, \tag{3}$$

where $e(x) = e^{2\pi\imath x}$ denotes the classical Fourier basis function. This approach is known as angular synchronization [31, 3] and (3) can be approximated via a SDP-based relaxation or a spectral method.

The main shortcome of these methods is that the only information they use from the observations $\{y_i\}$ is the best relative shifts δ_{ij}. This means that the performance of a given choice of $\{l_i\}$ can only be evaluated by comparing $l_i - l_j$ with δ_{ij} (in shift space) across pairs (i, j). This does not take into account the cost associated with other possible relative shifts of y_i and y_j. On the other hand, for a candidate solution $\{l_i\}$, relating $R_{-l_i}y_i$ and $R_{-l_i}y_j$ (in signal space) would take into account information about all possible shifts instead of just the best one (2). The quasi maximum likelihood estimator (Section 2) attempts to do exactly that by solving the minimization problem:

$$\min_{l_1,\ldots,l_N\in\mathbb{Z}_L} \sum_{i,j=1}^{N} \left\| R_{-l_i}y_i - R_{-l_j}y_j\right\|^2. \tag{4}$$

Finding the MLE (4) is a non-trivial computational task because its parameter space is of exponential size, and the likelihood function is non-convex. While one can apply optimization methods such as gradient descent, simulated annealing, or expectation-maximization (EM), these are only guaranteed to find local minima of (4), but not the global minimum.

In this paper we take a different approach and propose a semidefinite relaxation for the quasi maximum likelihood problem (4). This particular SDP is inspired by an approximation algorithm designed to solve the Unique Games problem [7] (Section 3).

Convex relaxations of hard combinatorial problems have seen many successes in applied mathematics. They became particularly popular in the last decade with the introduction of Compressed Sensing, in the seminal work of Donoho, Candes, Tao, and others [5, 10]. This idea has since been applied to a vast list of problems. Semidefinite programming (SDP) has served as a convex surrogate for problems arising in applications such as low-rank matrix completion [4], phase retrieval [6], Robust PCA [23], multiple-input multiple-output (MIMO) channel detection [26], and many others. In many of these applications the same phenomenon is present: for typical instances, solving the convex problem is often equivalent to solving the original combinatorial problem [1].

Semidefinite relaxations also play a central role in the design of approximation algorithms in theoretical computer science. Almost two decades ago, Goemans and Williamson [13] proposed an SDP based approximation algorithm for the MAX-CUT problem with approximation ratio $\alpha^{\mathsf{GW}} \approx 0.878$. That is, for any instance of the problem, the computed solution is guaranteed to provide performance (in this case, a cut) at least α^{GW} of the optimum. Many semidefinite relaxations have since been proposed as approximation algorithms for a long list of NP-hard problems [35].

In order to better understand the theoretical limitations of approximation algorithms, substantial work has been done to establish limits on the approximation ratios achievable by poly-time algorithms for certain NP-hard problems (hardness of approximation). The Unique Games conjecture by Khot [16] is central to many recent developments: For $\delta, \varepsilon > 0$, it is impossible for a polynomial-time algorithm to distinguish between δ-satisfiable and $(1-\varepsilon)$-satisfiable Unique-Games instances. A Unique-Games instance consists of a graph along with a permutations for each edge. The problem is to choose the best assignment of labels to each vertex such that as many of the edge permutations are satisfied. The validity of the UGC would imply the optimality of certain poly-time approximation ratios, in particular the Goemans-Williamson constant α^{GW} for the MAX-CUT problem [13, 17].

The best known polynomial time approximation to the unique games problem [7] is based on an SDP relaxation of a quadratic programming formulation that uses indicator variables. We note that this SDP is quite different from the relaxations normally used in applications (such as those described above). In particular, the variable matrix has size $NL \times NL$ and $\Omega\left(N^2L^2\right)$ constraints. The relaxation we propose consists of an adaptation of this SDP to approximate the quasi maximum likelihood problem (4).

As we show that it is Unique-Games hard to approximate (4) within any constant (assuming no noise model on $\{y_i\}$, see Section 2) it is hopeless to aim for good guarantees for general instances. However worst case analysis is often too pessimistic and not indicative of performance observed in practice. In fact, under the random noise model we assume for the observations, numerical simulations suggest that the SDP relaxation performs remarkably well, seeming to outperform existing methods. In an attempt to explain this phenomenon we show that, under our noise model, the SDP is stable at high SNR levels and even tight at extremely high SNR levels. By stability, we mean that with high probability the solution to the SDP does not deviate much from the true solution (see Section 4). The stability for this SDP is particularly interesting as it offers a new challenge in comparison with previously analyzed random instances of Unique-Games [2, 20]. This is because our noise model is on vertices, which translates into dependent noise on the edges. Still, these results fall short of properly explaining the remarkable performance that we see in simulations and more research is needed towards understanding the typical behavior of this relaxation.

In an attempt to simplify the SDP we also study a version with fewer constraints. Interestingly, this weaker SDP can be solved explicitly and is equivalent to the pairwise alignment method called phase correlation [14]. This method does not take into account information between all pairs of measurements, which suggests that the full complexity of the Unique-Games SDP [7] is needed to obtain a good approximation to (4).

The fact that a global shift does not affect the solution to (4) creates symmetries in our SDP relaxation. In fact, we leverage such structure, using symmetry reduction techniques to simplify the analysis of the SDP and greatly decrease its computational cost. This is particularly relevant given the high computational cost of state of the art semidefinite programming algorithms.

Contributions: Our main contribution is applying techniques from theoretical computer science to a problem from applied math. We introduce an Unique Games style SDP relaxation for the alignment problem, that is novel for the

applied math community. From the theoretical computer science point of view, we introduce a new problem that has a similar flavor to the Unique Games problem, and show that a worst case version is at least as hard as Unique Games. We introduce a natural average case version of this alignment problem, aligning several shifted copies of a signal corrupted by independent Gaussian noise. Existing analyses of semi-random models of Unique Games do not seem to apply to this problem due to the structure of the noise. We show that for sufficiently high SNR, the SDP solution is close to an integer solution. This is a first step to establishing a signal recovery result which we leave as an open problem. We believe that future investigations into this problem will yield interesting insights into the Unique Games SDP and on dealing with correlated noise in average case analysis.

2. QUASI MAXIMUM LIKELIHOOD ESTIMATOR

The log likelihood function for model (1) is given by

$$\mathcal{L}(x, l_1, \ldots, l_N) = \frac{N \log(2\pi)}{2} - \frac{1}{2\sigma} \sum_{i \in [N]} \|R_{-l_i} y_i - x\|^2. \quad (5)$$

Maximizing $\mathcal{L}$ is equivalent to minimizing the sum of squared residuals $\sum \|R_{-l_i} y_i - x\|^2$. Fixing the l_i's, the minimal value of $\mathcal{L}$ occurs at the average $x = \frac{1}{N} \sum_{i=1}^{N} R_{-l_i} y_i$. Making the tame assumption that $\|x\|^2$ is estimable (indeed the norm is shift-invariant) and thus fixed in (5), maximizing (5) is equivalent to maximizing the sum of the inner products $\langle R_{-l_i} y_i,\, R_{-l_j} y_j \rangle$ across all pairs (i, j). Thus we consider the estimator

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

Unfortunately, the search space for this optimization problem has exponential size. Indeed, assuming no model for the vectors $\{y_i\}$, it is NP-hard to find the shifts which maximize (6), or even estimate it within a close constant factor.

THEOREM 2.1. *Assuming no model on the observatinos $\{y_i\}$, it is NP-hard (under randomized reductions) to find a set of labels approximating (6) within $16/17 + \varepsilon$ of its optimum value. Furthermore, it is UG-hard (under randomized reductions) to approximate (6) within any constant factor.*

PROOF. (outline) We give a randomized reduction from the class of Γ-MAX-2LIN(q) instances consisting of a set of 2 variable linear equations of the form $x_i - x_j \equiv c_{ij} \pmod q$, with the goal of choosing an assignment for the variables which maximizes the number of satisfied equations. We construct a vector y_k for every variable x_k such that shifts of y_k correspond to an assignment to x_k. We pick a random vector z_{ij} corresponding to a constraint on variables x_i, x_j and place a copy of z_{ij} at specific locations in y_i and y_j. Shifts of y_i, y_j corresponding to satisfying assignments of the constraint results in a superposition of the copies of z_{ij}. We choose parameters so that the only non-trivial contributions to the objective function (6) come from such superposition. The value of the objective is (within small error) a scaled version of the number of constraints of the Γ-MAX-2LIN(q) instance satisfied by the assignment corresponding to the shifts. Thus hardness results for Γ-MAX-2LIN(q) directly translate to hardness results for the alignment problem. The details are given in Appendix C. □

The discrete optimization problem (6) may be formulated using indicator variables as an integer programming problem

$$\underset{\{u_{ik}\}}{\operatorname{argmax}} \sum_{i,j=1}^{N} \sum_{k,l \in \mathbb{Z}_L} u_{ik} u_{jl} \langle R_{-k} y_i, R_{-l} y_j \rangle, \quad (7)$$

where $u_{ik} \in \{0, 1\}$ and, for each i, $u_{ik} = 1$ for exactly one index k, corresponding to indicator variables $u_{ik} = \delta\{l_i \equiv k\}$. These requirements can be described with quadratic constraints (up to global sign, which cannot be fixed by quadratic constraints)

$$\begin{aligned} \sum_{k,l \in \mathbb{Z}_L} u_{ik} u_{jl} &= 1, \quad i, j \in [N] \\ u_{ik} u_{il} &= 0, \quad i \in [N],\ k \neq l \in \mathbb{Z}_L \\ u_{ik} u_{jl} &\geq 0, \quad i, j \in [N],\ k, l \in \mathbb{Z}_L. \end{aligned} \quad (8)$$

Note that both the objective function (7) and the constraints (8) depend only on products of the form $u_{ik} u_{jl}$. This means that we can write the problem in terms of the Gram matrix $U \in \mathbb{R}^{NL \times NL}$ with entries $U_{ik;jl} = u_{ik} u_{jl}$. To ensure that $U \in \mathbb{R}^{NL \times NL}$ is indeed of the form $U = uu^T$ it suffices to require that $U \succeq 0$ and $\operatorname{rank}(U) = 1$. This means that (7) is equivalent to

$$\begin{aligned} \max_{U \in \mathbb{R}^{NL \times NL}} \quad & \operatorname{tr}(CU) \\ \text{subject to} \quad & \sum_{k,l} U_{ik;jl} = 1 \text{ for } i, j \in [N] \\ & U_{ik;il} = 0 \text{ for } i \in [N],\ k \neq l \in \mathbb{Z}_L \\ & U \geq 0,\ U \succeq 0,\ \operatorname{rank}(U) = 1, \end{aligned} \quad (9)$$

where $C \in \mathbb{R}^{NL \times NL}$ is the data Gram matrix, with entries $C_{ik;jl} = \langle R_{-k} y_i, R_{-l} y_j \rangle$.

Due to the global shift redundancy of the multireference alignment problem, (9) will have L equivalent solutions (corresponding to the L shifts). We will deal with this redundancy by averaging the L solutions. This corresponds to a rank L matrix V satisfying $V_{ik;ik} = 1/L$ for all $i \in [N]$ and $k \in \mathbb{Z}_L$. The L solutions u_{ik} lie in the L dimensional space spanned by the L non-zero eigenvectors of V. Hence the problem can be written as:

$$\begin{aligned} \max_{V \in \mathbb{R}^{NL \times NL}} \quad & \operatorname{tr}(CV) \\ \text{subject to} \quad & \sum_{k,l} V_{ik;jl} = 1 \text{ for } i \neq j \in [N] \\ & V_{ik;ik} = 1/L \text{ for } i \in [N] \\ & V_{ik;il} = 0 \text{ for } i \in [N],\ k \neq l \in \mathbb{Z}_L \\ & V \geq 0,\ V \succeq 0,\ \operatorname{rank}(V) \leq L. \end{aligned} \quad (10)$$

2.1 Data Gram matrix and a spectral method

The data Gram matrix C has special structural properties that may be leveraged. The following lemma describes this structure. Let $\mathcal{F}$ be the normalized Fourier transform $\mathcal{F}(y_i, k) = \frac{1}{\sqrt{L}} \sum_{l=0}^{L-1} e\left(\frac{-kl}{L}\right) (y_i)_l$.

LEMMA 2.2. *The data Gram matrix C with entries $C_{ik;jl} = \langle R_{-k} y_i, R_{-l} y_j \rangle$ satisfies:*

1. *$C \succeq 0$ and has rank L, with non-zero eigenvalues $\lambda_k = L \sum_{i=1}^{N} |\mathcal{F}(y_i, k)|^2$.*

2. *There is a unitary matrix $\mathcal{P}$ for which $\mathcal{P}C\mathcal{P}^* = \mathrm{diag}(\mathcal{C}_0, \ldots, \mathcal{C}_{L-1})$ is block diagonal, where each $\mathcal{C}_k \in \mathbb{C}^{N\times N}$ is a rank 1 matrix,*

where

PROOF. This follows from block circulance of (a permutation of) C. Refer to Lemma B.1. □

Ideally one would be able to read off the solution indicator vectors $\mathbf{1}\{ik : k \equiv l_i\} \in \mathbb{R}^{NL}$ from properties of C. In fact, it is not difficult to see that in the noiseless case ($\sigma = 0$), the span of the top L eigenvectors of C coincides with the span of the indicator vectors corresponding to the L globally shifted solutions.

Hence, one may try to recover a solution to (7) by examining the eigenvectors associated with the top eigenvalues in the spectrum of C. For random signals x, Lemma 2.2 indicates that the spectral gap between the top L eigenvalues and the remaining eigenvalues of C will be large, on the order $\min_k \sum_i |\mathcal{F}(y_i, k)|^2 = \Omega(LN)$. This suggests that a rounding procedure from the eigenvalues of C could be robust to noise, and motivate the application of a spectral method for the noisy case.

Unfortunately, this spectral gap will be rather small for a large class of signals. As long as a single power spectra $|\mathcal{F}(x, k)|^2$ of x is near zero, the corresponding eigenvalue λ_k will separate less from the small eigenvalues of C. Hence with additional noise, it would be less reliable to try and recover the solution indicator vector from the top L eigenvectors of C. This suggests that in this context, the SNR should be defined with respect to the spectral gap $\min_k |\mathcal{F}(x, k)|^2$. Furthermore, our simulations suggest that recovery from a spectral relaxation performs worse than the semidefinite relaxation we are about to propose.

3. SEMIDEFINITE RELAXATION

In (10), the only non-convex constraint imposed on V is that of rank deficiency, which obstructs the use of convex programming techniques. Removing this rank constraint yields a semidefinite program:

$$\begin{aligned} \max_{V\in\mathbb{R}^{NL\times NL}} \quad & \mathrm{tr}(CV) && (11)\\ \text{subject to} \quad & \sum_{k,l} V_{ik;jl} = 1 \text{ for } i \neq j \in [N] \\ & V_{ik;ik} = 1/L,\ V_{ik;il} = 0 \text{ for } k \neq l, && (12)\\ & V \geq 0,\ V \succeq 0. \end{aligned}$$

This SDP is extremely similar to and motivated by SDPs commonly used to approximate solutions to certain constraint satisfaction problems (CSPs), notably Unique-Games instances. An Unique-Games instance consists of a graph $G = ([N], E)$, a label set $\mathbb{Z}_L$, and a set of permutations $\pi_{ij} : \mathbb{Z}_L \to \mathbb{Z}_L$. The problem is to choose the best assignment of labels to each vertex such that as many of the permutations $(\pi_{ij})_{(i,j)\in E}$ as possible are satisfied. Γ-MAX-2LIN(L) is a special case where the permutations π_{ij} are cyclic. In our notation, the SDP studied for Unique-Games is usually of the form

$$\begin{aligned} \max_{V} \quad & \frac{1}{2}\mathrm{tr}(\tilde{C}V) && (13)\\ \text{subject to} \quad & \tilde{C}, V \in \mathbb{R}^{NL\times NL},\ \tilde{C}_{ik;jl} = \delta\{l = \pi_{ij}(k)\}, \\ & \sum_k V_{ik;ik} = 1,\ V_{ik;il} = 0 \text{ for } k \neq l, && (14)\\ & V \geq 0,\ V \succeq 0. \end{aligned}$$

This formulation attempts to count the number of satisfied edge constraints for an Unique-Games instance. In this context, the matrix $\tilde{C}$ is dubbed the label-extended adjacency matrix [18]. One can treat the SDP (11) as an instance of Γ-MAX-2LIN(L) on a weighted complete graph, with each cyclic permutation weighted by $C_{ik;jl} = \langle R_{-l_i} y_i, R_{-l_j} y_j \rangle$. Thus the significant body of literature conducted on Unique-Games may be useful in understanding (11). Another common feature of the aligment SDP with Γ-MAX-2LIN(L) instances is that the assigned labels may be chosen up to cyclic symmetry. This induces a block circulant symmetry in the semidefinite program. For example, it was this symmetry that allowed us to obtain the constraints $V_{ik;ik} = 1/L$ from the constraints $\sum_k V_{ik;ik} = 1$.

A major feature of (11) which distinguishes it from the general Unique-Games problem is the structure of the data coefficient matrix C (see Lemma 2.2). While Unique-Games specifies constraints on edges of a graph (there are N^2 pieces of information), the alignment problem only specifies information on its vertices (N pieces of information). While this does assist our understanding of the semidefinite program, since it enables us to apply more symmetry conditions, it also will complicate some of our analysis (see Section 4).

The number of constraints in our SDP (11) is dominated by the positivity constraints $V \geq 0$. One source of intuition behind it is that it enforces triangle inequality constraints $\|v_{ik} - \mathbf{0}\| + \|\mathbf{0} - v_{jl}\| \geq \|v_{ik} - v_{jl}\|$ [7]. From the next section, as well as from empirical results, we see that the constraint $V \geq 0$ seems to cause the SDP solution to be more stable around integral instances. As a trade-off, the computational cost of solving the SDP scales with the number of constraints, so we investigated the effect of removing the positivity constraints in (11). Interestingly, without this positivity constraint, the SDP (11) can be solved in closed form, and is effectively equivalent to applying phase correlation (see Appendix A) to each pair of the observations.

THEOREM 3.1. *There is a matrix $V \in \mathbb{R}^{NL\times NL}$ composed of circulant $L \times L$ blocks which solves the program (11), excluding the positivity constraint $V \geq 0$. V has rank L, corresponding to one eigenspace of eigenvalue N/L. This eigenspace contains the vector $v^{\mathsf{phase}} \in \mathbb{R}^{NL}$ satisfying*

$$v^{\mathsf{phase}}_{ik} = \frac{1}{\sqrt{L}} \sum_{l=0}^{L-1} e\left(\tfrac{-kl}{L}\right) \cdot \frac{\mathcal{F}(y_1, l)\mathcal{F}^*(y_i, l)}{|\mathcal{F}(y_1, l)\mathcal{F}^*(y_i, l)|}$$

which is the concatenation of phase correlation vectors between y_1 and y_i.

PROOF. (outline) The existence of a solution V composed of circulant blocks follows from symmetry. Hence, as in Lemma 2.2, V may block diagonalized as $\mathcal{P}V\mathcal{P}^* = \mathrm{diag}(\mathcal{V}_0, \ldots, \mathcal{V}_{L-1})$ for $\mathcal{V}_k \in \mathbb{C}^{N\times N}$. When the SDP constraints are written with respect to the entries of the $\mathcal{V}_k$'s, one resulting constraint is that the magnitude of each entry

of $\mathcal{V}_k$ is bounded. As a consequence, the objective function (11) would be maximized when the entries of $\mathcal{V}_k$ have maximal magnitudes, and have the same complex phases as the corresponding entries in $\mathcal{C}_k$. This choice of V indeed lies in the SDP feasibility region (ignoring the positivity constraints). Analogous to phase correlation, the information used from the data matrix is entirely phase information. The details are given in the proof of Theorem B.3. □

From the SDP solution, one must round back to a solution in the original search space. There is a significant body of literature on the topic of rounding the solutions to various SDPs for Unique Games. The analysis and guarantees for these rounding schemes are usually expressed in terms of percentage of constraints satisfied by the SDP solution and do not immediately give a result about signal recovery in our setting. For example, in studying semi-random instances of Unique-Games, the authors of [19] give a rounding technique that uses both SDP and LP solutions when the SDP is known to be somewhat sparse. This is similar to a condition we obtain in the following stability section. Exploiting these ideas to establish an exact signal recovery guarantee is an interesting open problem.

4. STABILITY

For simplicity, we will assume, without loss of generality, that all the ground truth shifts correspond to the zero shift. Hence $y_i = x + \xi_i$, where $\xi_i \sim \mathcal{N}(0, \sigma^2 I_L)$ i.i.d.. The ground truth integral instance for the SDP will be (up to block cyclic symmetry) the indicator matrix $V^{int} \in \mathbb{R}^{NL \times NL}$, defined as $V^{int}_{ik;jl} = \delta_{k=l}/L$. It is not difficult to see that, if there is a set of labels $\{l_i\}$ which are pairwise optimal (in the sense that $\langle R_{-l_i} y_i, R_{-l_j} y_j \rangle = \max_k \langle y_i, R_{-k} y_j \rangle$ for all pairs (i,j)) the SDP (11) will return the integral instance. While this may be likely in a very high signal-to-noise setting, it is a rather stringent condition. We relax this condition and show that the SDP solution must still resemble the integral instance at a reasonably high SNR. The exact definition for the SNR will be deferred for later, but will be characterized in terms of the gap between the correct offset and incorrect shifts

$$\Delta = \|x\|^2 - \max_{l \neq 0} \langle x, R_l x \rangle, \tag{15}$$

and the noise level.

For any $V \in \mathbb{R}^{NL \times NL}$ lying in the SDP feasibility region, we can characterize the distance of V from the integral instance by the differences

$$D_{ij} = \sum_{k \neq l \in \mathbb{Z}_L} V_{ik;jl} = 1 - \sum_{k \in \mathbb{Z}_L} V_{ik;jk} \in [0,1].$$

D_{ij} is a measure of how much the SDP would weight shift preferences other than the ground truth. Note that D_{ij} is always non-negative and moreover, when all $D_{ij} = 0$ we obtain the integral instance.

For convenience, we make the definitions $\xi_0 = 2x$ and $\eta_{ij} = 2 \max_l |\langle R_l \xi_i, \xi_j \rangle| \geq 0$ for $i, j = 0, 1, \ldots, N$. The following Lemma provides a control on D_{ij} from the difference of the signal term Δ and the noise terms η_{ij}.

LEMMA 4.1. *If* $\operatorname{tr}(CV) \geq \operatorname{tr}(CV^{int})$, *then* $\sum_{i \neq j} (\Delta - \eta_{ij} - \eta_{j0}) D_{ij} \leq 0$.

PROOF. We can find a block circulant matrix which attains the same SDP objective value as V, so without loss of generality presume V is block circulant. Hence $\sum_k V_{ik;j0} = 1/L$ and $D_{ij} = 1 - L V_{i0;j0}$. Expanding,

$$\begin{aligned}
0 &\geq \operatorname{tr}(CV^{int}) - \operatorname{tr}(CV) \\
&= \sum_{i,j} \sum_k \langle R_{-k} y_i, y_j \rangle \left(\delta_{k=0} - L V_{ik;j0} \right) \\
&\geq \sum_{i,j} \left(\langle y_i, y_j \rangle (1 - L V_{i0;j0}) - \max_{l \neq 0} \langle R_l y_i, y_j \rangle \sum_{k \neq 0} L V_{ik;j0} \right) \\
&= \sum_{i,j} \left(\langle y_i, y_j \rangle - \max_{l \neq 0} \langle R_l y_i, y_j \rangle \right) D_{ij}.
\end{aligned}$$

The second inequality requires the SDP constraint $V \geq 0$. Expanding the inner products and applying pessimistic bounds,

$$\langle y_i, y_j \rangle - \max_{l \neq 0} \langle R_l y_i, y_j \rangle \geq \Delta - (\eta_{i0}/2 + \eta_{j0}/2 + \eta_{ij}),$$

and rearrangement gives the desired inequality. □

THEOREM 4.2. *With probability* $1 - e^{-N + o(N)}$, *the solution to the SDP satisfies*

$$\sum_{i,j} D_{ij} \leq \frac{(\|x\| + \sigma^2 \sqrt{L}) \cdot 12 \log eL}{\Delta} \cdot N^2.$$

PROOF. For sufficiently high SNR, we would expect that the inequality in Lemma 4.1 would fail to hold. Indeed, by Lemma D.2, the inequality

$$\sum_{i \neq j} (\eta_{ij} + \eta_{j0}) D_{ij} \leq \mathcal{O}(\log L) \cdot \left(2\|x\| + \frac{1}{N} \sum_i \|\xi_i\| \right) N^2,$$

holds with probability at least $1 - e^{-N + \mathcal{O}(\log N)}$. It arises from tail bounds on the sum of slightly dependent random variables, and is independent of the structure of the SDP (as opposed to Lemma 4.1). Combined with Lemma 4.1, we can obtain a guarantee on the deviation between the SDP solution and the integral instance. The full proof may be found following Theorem D.3. □

Theorem 4.2 indicates that at a sufficiently high SNR, the UG-based SDP will produce a matrix V, of which each $L \times L$ block has most weight concentrated on its main diagonal. A rounding scheme would likely interpret this as the identity shift being the optimal shift. This motivates us to choose a definition for the SNR to be along the lines of $SNR = \Delta / [(\|x\| + \sigma^2 \sqrt{L}) \log L]$ (for reference, for random signals x, we would expect that $\Delta = L - \mathcal{O}(\sqrt{L})$). This characterization of SNR is significantly more lenient than that for spectral relaxation, in Section 2. For example, if the signal x was a sum of a few sinusoids, it would have several small power spectra, but as long as the least common multiple of their periods does not divide L, then Δ will still be large. Note further that we have more flexibility when defining Δ: for example, suppose there is a set of shifts $\boldsymbol{\ell}^*$ (including the identity element) of size $|\boldsymbol{\ell}^*| = \mathcal{O}(1)$ for which $\max_{l \in \boldsymbol{\ell}^*} \langle x, R_l x \rangle > \Delta + \min_{l \notin \boldsymbol{\ell}^*} \langle x, R_l x \rangle$. The above results may be modified to describe the concentration of the SDP solution on the entries of the shifts in $\boldsymbol{\ell}^*$ (not just along the identity shift) for each pair of observations y_i, y_j.

This result may be strengthened in other manners. For constants $0 < \delta, \varepsilon \ll 1$, we can attain a tighter concentration

condition of the form

$$\sum D_{ij} \geq (1+\delta)^2 N \sqrt{\sum D_{ij}^2}/SNR + 2\varepsilon N^2 \qquad (16)$$

instead of the current condition $\sum D_{ij} \geq N^2/SNR$. The argument is based on the analysis of the SDP for adversarial semi-random Unique-Games instances by [19]. However, the proof must be more nuanced in our case, due to correlations in the noise model of C, caused by the smaller source of randomness available to us. Hence we omit the full details, but provide a sketch below.

Any SDP feasible matrix V can alternatively be represented as $V_{ik;jl} = \langle v_{ik}, v_{jl} \rangle / L$ for a set of unit vectors $v_{ik} \in \mathbb{R}^{NL}$. With this notation, $2D_{ij} = 2(1 - LV_{i0;j0}) = \|v_{i0} - v_{j0}\|^2$. The space of these unit vectors can be approximately discretized by a random projection due to the Johnson-Lindenstrauss lemma. More precisely, Lemma D.4 provides a set of unit vectors $\mathfrak{N}$ of size $|\mathfrak{N}| = \exp\left(\mathcal{O}\left(\delta^{-2}\log^2(1/\varepsilon)\right)\right)$ such that the set of unit vectors $\{v_{i0}\}_{i=1}^N$ can be approximated under a random projection φ by an N-tuple in $\mathfrak{N}$. Specifically, if $D_{ij}^{\varphi} = \|\varphi(v_{i0}) - \varphi(v_{j0})\|^2/2$, then the D_{ij}'s lie within an α_{JL}-ball of D_{ij}^{φ}, this ball being defined as $\{D : \alpha_{JL}^{-1}(D^{\varphi}) \leq D \leq \alpha_{JL}(D^{\varphi})\}$ for $\alpha_{JL}(D) = (1+\delta)D + \varepsilon$.

Instead of finding a global tail bound in Lemma D.2, we can derive a local tail bound for each α_{JL}-ball of D_{ij}^{φ}'s, each tail bound holding with probability at least $1 - 2Ne^{-t^2 N}$. For $0 < \delta, \varepsilon = \varepsilon' \ll 1$ constants, the number of N-tuples of vectors in $\mathfrak{N}$ is of size $\exp(\mathcal{O}(N))$. With a sufficiently large constant t, the local tail bound may be union bounded across all balls of vectors in $\mathfrak{N}^N$, and thus will hold for all SDP-feasible V. With some care, this argument would yield a concentration condition of the form (16).

5. NUMERICAL RESULTS

We implemented several baseline methods for multireference alignment, and plotted their average error performance across 500 iterations in Figure 5. For each iteration, we chose a signal x randomly from the distribution $\mathcal{N}(0, I_L)$, as well as N i.i.d. noise vectors $\xi_i \sim \mathcal{N}(0, \sigma^2 I_L)$, and applied each of our methods. These simulations confirm our intuition that the UG-based SDP performs better than other benchmark techniques. In particular, they suggest that the UG-based SDP is highly stable around integral instances.

The implementation using bispectrum-like invariants is discussed in Appendix A. For each of the other procedures, we construct a $NL \times L$ matrix W which records alignment preferences of shifts of y_i with y_1. For cross- or phase- correlation, $W_{ik;l}$ is the $(k-l)$th entry of the cross- or phase-correlation vector between y_1 and y_i. For the spectral rounding off the Gram matrix C, and the solution of the UG-SDP, W is formed by the top L eigenvectors of the respective matrix. The shifts are read off this matrix, and the un-shifted y_i are averaged to produce an estimate for x. The first plot shows the difference between this estimate and x.

A second measurement of performance we looked at is how easy it would be to determine the best shifts from W. A natural method is to identify an indicator vector $\mathbf{1}\{ik : l_i \equiv k\}$ lying close to the column span of W, for some labelling $\{l_i\}$. The heuristic we implemented was to apply a linear transformation to W such that the top $L \times L$ block of W is the identity matrix I_L (although there is no reason to believe this rounding is robust, it is sufficient for the purpose of having an easy benchmark between multiple methods). Then, for each i, the maximal entry in the first column of W gives the choice of shift for y_i. The second plot shows the distance this indicator vector lies from the column span of W.

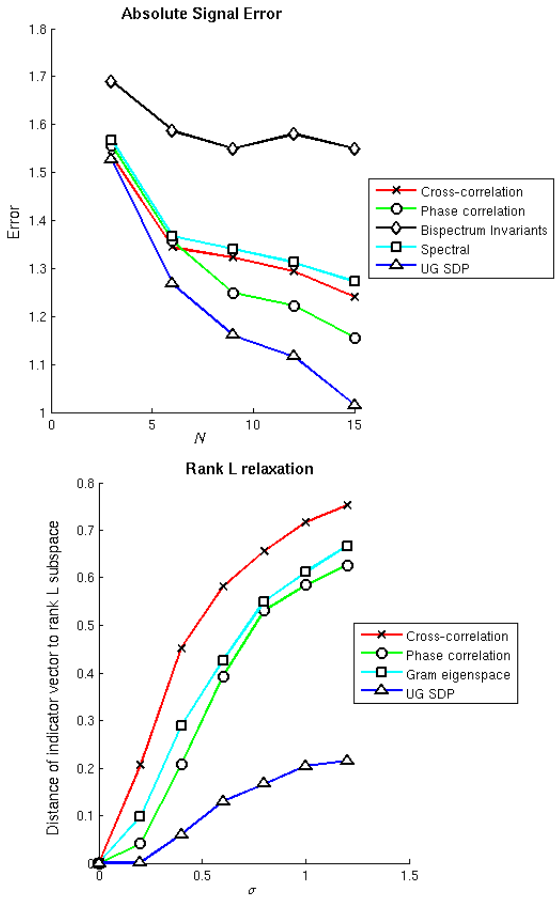

Figure 1: Averages of errors of several alignment methods across 500 iterations. The top plot has parameters $\sigma = 1$ and $L = 5$, and the bottom plot has parameters $N = 12$ and $L = 5$.

Another, perhaps more robust method, is to apply the theory of sparse recovery. The desired indicator vector $u \in \mathbb{R}^{NL}$ is a sparse vector lying near the column space of W. As suggested by Spielman, et al, in [33], such a sparse vector could be found by solving the ℓ_1-minimization problem $\min_{b \in \mathbb{R}^L} \|Wb\|_1$ and setting $u = Wb$. To prevent the solution from collapsing to zero, the search space can be limited by an affine constraint $b^T c = 1$ for an appropriate choice of $c \in \mathbb{R}^L$. In [33], this ℓ_1-minimization is done by choosing c from the rows of W.

6. GENERALIZATIONS AND FUTURE WORK

It is worth noting that the discrete multireference alignment problem naturally generalizes from shifts over $\mathbb{Z}_L$ to actions of finite groups G over finite spaces S. In this case, the analogue of phase correlation [25] is defined in terms of the generalized Fourier transform $\mathcal{F}_{G,S}(\cdot, \rho) : \mathbb{C}^S \to \mathbb{C}^{d_\rho \times d_\rho}$, where $\rho : G \to \mathbb{C}^{d_\rho \times d_\rho}$ are irreducible matrix representa-

tions. The Unique Games SDP may also be generalized in a natural manner, in rough analogy with SDP variants studied for Γ-MAX-2LIN instances. In the case of an abelian group $G = S$, the proof of Theorem 3.1 follows in the same fashion, and hence the alignment UG-based SDP without positivity constraints will remain equivalent to phase correlation.

The symmetry of the SDP in this case can be described naturally by the theory of $\mathbb{C}*$-matrix algebras. Symmetries in semidefinite programs can be written with respect to linear combinations of $\{0, 1\}$ matrices, which form a basis of a $*$-matrix algebra. Under sufficiently nice conditions, this basis can be block diagonalized, for example by Wedderburn's decomposition or by regular $*$-representations [28]. Hence, the $*$-matrix algebra can be represented by a lower-dimensional block diagonalized version. From a computational perspective, this can make highly symmetric SDPs much more amenable to sparse SDP solvers [9]. In our case, diagonalization by the block DFT allows the SDP to be rewritten as an optimization problem across L PSD matrices of size $N \times N$ instead of one PSD matrix of size $NL \times NL$.

It would be interesting to see how the performance of the SDP changes when we study the alignment problem across more difficult groups, especially non-abelian ones. For example, [32] applies angular synchronization to resolve an alignment problem over $SO(3)$.

The numerical simulations in Section 5 suggest that the UG-based SDP achieves exact recovery with high probability for sufficiently high SNR. That is, the resulting SDP matrix is integral, and by solving the SDP we obtain the solution to the quasi-ML estimator. Indeed, as the SNR decreases, there appears to be a phase transition during which the SDP almost always recovers an integral solution. Our analysis of the stability of the semidefinite program does not fully explain this phenomenon. The authors believe this to be an interesting direction for future work, especially since guarantees of exact recovery are attainable in high SNR settings for a few semidefinite relaxations, for example for the MIMO problem [26].

Another important question is to understand the sample complexity of our approach to the alignment problem. Since the objective is to recover the underlying signal x, a larger number of observations N should yield a better recovery. The question can be formalized as: for a given value of L and σ, how large does N need to be in order to allow for a reasonably accurate recovery? Methods like the bispectral invariants would be expected to require $N = \Omega(\sigma^2 L^2 \log L)$ observations. We would hypothesize, on the strength of our numerical results, that the UG-based SDP requires fewer observations for meaningful recovery.

Along with expanding the domain of the alignment problem, it would be interesting to attempt the style of analysis discussed in this paper for other maximum likelihood problems. Maximum likelihood estimators play an important role in many estimation problems, but often (as in our problem) computing or approximating the MLE is a challenging problem and semidefinite programming could perhaps provide a tractable alternative in an average case setting.

7. ACKNOWLEDGMENTS

The authors thank Yutong Chen for valuable assistance with the implementation of our algorithm. A. S. Bandeira was supported by AFOSR Grant No. FA9550-12-1-0317. M. Charikar was supported by NSF grants CCF 0832797, AF 0916218 and AF 1218687. A. Singer was partially supported by Award Number FA9550-12-1-0317 and FA9550-13-1-0076 from AFOSR, by Award Number R01GM090200 from the NIGMS, and by Award Number LTR DTD 06-05-2012 from the Simons Foundation. Parts of this work have appeared in A. Zhu's undergraduate thesis at Princeton University.

8. REFERENCES

[1] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp. Living on the edge: A geometric theory of phase transitions in convex optimization. *Available online at arXiv:1303.6672 [cs.IT]*, 2013.

[2] S. Arora, S. A. Khot, A. Kolla, D. Steurer, M. Tulsiani, and N. K. Vishnoi. Unique games on expanding constraint graphs are easy: extended abstract. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, STOC '08, pages 21–28, New York, NY, USA, 2008. ACM.

[3] A. S. Bandeira, A. Singer, and D. A. Spielman. A Cheeger inequality for the graph connection Laplacian. *to appear in SIAM Journal on Matrix Analysis and Applications (SIMAX)*, 2013.

[4] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.

[5] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59:1207–1223, 2006.

[6] E. J. Candès, T. Strohmer, and V. Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013.

[7] M. Charikar, K. Makarychev, and Y. Makarychev. Near-optimal algorithms for unique games. pages 205–214, 2006.

[8] R. Diamond. On the multiple simultaneous superposition of molecular structures by rigid body transformations. *Protein Science*, 1(10):1279–1287, October 1992.

[9] C. Dobre and E. de Klerk. Semidefinite programming approaches for structured combinatorial optimization problems, 2011.

[10] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52:1289–1306, 2006.

[11] I. Dryden and K. Mardia. *Statistical shape analysis*. Wiley series in probability and statistics. Wiley, Chichester [u.a.], 1998.

[12] H. Foroosh, J. B. Zerubia, and M. Berthod. Extension of phase correlation to subpixel registration. *IEEE Transactions on Image Processing*, 11(3):188–200, 2002.

[13] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefine programming. *Journal of the Association for Computing Machinery*, 42:1115–1145, 1995.

[14] J. L. Horner and P. D. Gianino. Phase-only matched filtering. *Appl. Opt.*, 23(6):812–816, Mar 1984.

[15] R. Kakarala and G. J. Iverson. Uniqueness of results for multiple correlations of periodic functions. *J. Opt. Soc. Am. A*, 10(7):1517–1528, Jul 1993.

[16] S. Khot. On the power of unique 2-prover 1-round games. *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 767–775, 2002.

[17] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for max-cut and other 2-variable csps? *SIAM J. Comput.*, 37(1):319–357, 2007.

[18] A. Kolla. Spectral algorithms for unique games. *25th Annual IEEE Conference on Computational Complexity*, pages 122–130, 2010.

[19] A. Kolla, K. Makarychev, and Y. Makarychev. How to play unique games against a semi-random adversary: Study of semi-random models of unique games. *Foundations of Computer Science, IEEE Annual Symposium on*, pages 443–452, 2011.

[20] A. Kolla and M. Tulsiani. Playing random and expanding unique games. Unpublished, 2007.

[21] P. Kosir, R. DeWall, and R. Mitchell. A multiple measurement approach for feature alignment. In *Aerospace and Electronics Conference, 1995. NAECON 1995., Proceedings of the IEEE 1995 National*, volume 1, pages 94–101 vol.1, 1995.

[22] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):pp. 1302–1338, 2000.

[23] G. Lerman, M. B. McCoy, J. A. Tropp, and T. Zhang. Robust computation of linear models, or how to find a needle in a haystack. *CoRR*, abs/1202.4044, 2012.

[24] W. Ligong, L. Xueliang, and H. C. Eigenvalues of a special kind of symmetric block circulant matrices. *Applied Math J. Chinese University Series B*, 19(1):17–26, 2004.

[25] A. Loneland. Non-commutative harmonic analysis: Generalization of phase correlation to the euclidean motion group. Master's thesis, University of Bergen, June 2010.

[26] A. Man-Cho So. Probabilistic analysis of the semidefinite relaxation detector in digital communications. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '10, pages 698–711, Philadelphia, PA, USA, 2010.

[27] A. Maurer. A bound on the deviation probability for sums of non-negative random variables. *J. Inequalities in Pure and Applied Mathematics*, 4(1):15, 2003.

[28] K. Murota. A numerical algorithm for block-diagonal decomposition of matrix *-algebras with application to semidefinite programming. *Japan Journal of Industrial and Applied Mathematics*, 27(1), 2007.

[29] R. G. Pita, M. R. Zurera, P. J. Amores, and F. L. Ferreras. Using multilayer perceptrons to align high range resolution radar signals. In W. Duch, J. Kacprzyk, E. Oja, and S. Zadrozny, editors, *Artificial Neural Networks: Formal Models and Their Applications - ICANN 2005*, volume 3697 of *Lecture Notes in Computer Science*, pages 911–916. Springer Berlin Heidelberg, 2005.

[30] B. M. Sadler and G. B. Giannakis. Shift- and rotation-invariant object reconstruction using the bispectrum. *J. Opt. Soc. Am. A*, 9(1):57–69, Jan 1992.

[31] A. Singer. Angular synchronization by eigenvectors and semidefinite programming. *Applied and Computational Harmonic Analysis*, 30(1):20 – 36, 2011.

[32] B. Sonday, A. Singer, and I. G. Kevrekidis. Noisy dynamic simulations in the presence of symmetry: Data alignment and model reduction. *Computers & Mathematics with Applications*, 65(10):1535 – 1557, 2013.

[33] D. A. Spielman, H. Wang, and J. Wright. Exact recovery of sparsely-used dictionaries. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, IJCAI'13, pages 3087–3090. AAAI Press, 2013.

[34] D. L. Theobald and P. A. Steindel. Optimal simultaneous superpositioning of multiple structures with missing data. *Bioinformatics*, 28(15):1972–1979, 2012.

[35] D. P. Williamson and D. B. Shmoys. *The Design of Approximation Algorithms*. Cambridge University Press, New York, NY, USA, 1st edition, 2011.

[36] J. Zwart, R. van der Heiden, S. Gelsema, and F. Groen. Fast translation invariant classification of hrr range profiles in a zero phase representation. *Radar, Sonar and Navigation, IEE Proceedings*, 150(6):411–418, 2003.

APPENDIX

A. PHASE CORRELATION AND THE BISPECTRUM

Variations of the alignment problem crop up in many scientific fields, and thus many methods for alignment have been independently proposed. Such methods include iterative template aligning [21], zero phase representations [36], angular synchronization [32], and machine learning [29]. Unfortunately, to the authors' knowledge, no proposal exists for a multireference alignment method which fairly treats each input observation, makes use of all available information, and has rigorous performance discussion. We outline a couple of foundational techniques used in alignment literature in this section.

When aligning a pair of noisily shifted vectors ($N = 2$), a natural approach is to assign a score to each possible offset between the two vectors, and to estimate the best shift as the one with the highest score. A straightforward score is the inner product between the vectors, also known as cross-correlation. For two observations $y_i, y_j \in \mathbb{R}^L$, the selected shift is the maximal entry $\widehat{l} = \operatorname{argmax} v_l^{\mathsf{cross}}$ of the cross-correlation vector $v_l^{\mathsf{cross}} = \langle y_i, R_{-l} y_j \rangle$. More frequently used in practice is the phase correlation vector, whose components are

$$v_k^{\mathsf{phase}} = \frac{1}{\sqrt{L}} \sum_{l=0}^{L-1} e\left(\tfrac{-kl}{L}\right) \cdot \frac{\mathcal{F}(y_i, l)\mathcal{F}^*(y_j, l)}{|\mathcal{F}(y_i, l)\mathcal{F}^*(y_j, l)|}. \tag{17}$$

Essentially this measures the similarity of the observations in frequency space. For human-friendly images, phase correlation tends to be more appealing, since v^{phase} tends to

have a sharper peak as compared to v^{cross} [14]. The relation between cross- and phase correlation can be seen from the convolution theorem:

LEMMA A.1. *Convolution theorem:* $(\mathcal{F}^* v^{\text{cross}})_k = \sqrt{L} \cdot \mathcal{F}(y_i, k)\mathcal{F}^*(y_j, k)$.

PROOF. Let M_l denote the modulation operator $(M_l x)_k = x_k e(kl/L)$, defined such that $\mathcal{F}R_l x = M_l \mathcal{F} x$. By Parseval's Theorem,

$$\langle y_i, R_{-l} y_j \rangle = \langle \mathcal{F} y_i, M_{-l} \mathcal{F} y_j \rangle = \sum_{l=0}^{L-1} e\left(\tfrac{-kl}{L}\right) \mathcal{F}(y_i, l)\mathcal{F}^*(y_j, l).$$

□

Cross- and phase correlation may be directly applied for the problem of multireference alignment, say by aligning all of the observations against the first one. However, this is not a robust method.

Another relevant notion for characterizing alignments are the moment spectra, which are shift-invariant properties of the observations. The kth power spectrum of a real signal $x \in \mathbb{R}^L$ is defined by $|\mathcal{F}(x,k)|^2 = \mathcal{F}(x,k)\mathcal{F}^*(x,k)$. This can be extended to the d-th order bispectrum (or moment spectra)

$$b(k_1, k_2, \ldots, k_d) = \mathcal{F}\left(x, \Sigma_{i=1}^d k_i\right) \prod_{i=1}^d \mathcal{F}^*(x, k_i); \qquad (18)$$

its shift invariance can be seen since it is the Fourier transform of the d-th order autocorrelation function

$$a(k_1, k_2, \ldots, k_d) = \sum_{l=0}^{L-1} x_l \overline{x_{l-k_1} x_{l-k_2} \cdots x_{l-k_d}}, \qquad (19)$$

by the Wiener-Khinchin theorem. How much information do these invariants capture about the signal x? Sadler and Giannakis give iterative and least squares algorithms in [30] for reconstructing the Fourier phases from full knowledge of the bispectrum, and Kakarala in [15] show the uniqueness of real function given all of its higher order moment spectra. These arguments generally tend to be of a more number-theoretic flavor and may be tedious to implement in practice.

A simple special case occurs when we take $k_1 = k_2 = \ldots = k_d = 1$ for $d = 1, \ldots, L$. This gives us the set of bispectral invariants $\mathcal{F}(x,d)\mathcal{F}^*(x,1)^d$, which can be consistently estimated from our observations y_i. With an estimate for the phase of the first Fourier coefficient $\mathcal{F}(x,1)$ (say, sample over a discretization of the unit circle), we can recover the remaining Fourier phases $\mathcal{F}(x,d)$ and thus the signal x. This approach is one of the baseline methods used in Section 5.

B. RELAXATIONS FOR MAXIMUM LIKELIHOOD

LEMMA B.1. *The data Gram matrix C with entries $C_{ik;jl} = \langle R_{-k} y_i, R_{-l} y_j \rangle$ satisfies the following properties:*

1. *$C \succeq 0$ and has rank L, with non-zero eigenvalues $\lambda_k = L \sum_{i=1}^N |\mathcal{F}(y_i, k)|^2$.*
2. *There is a unitary matrix $\mathcal{P}$ for which $\mathcal{P}C\mathcal{P}^* = \operatorname{diag}(\mathcal{C}_0, \ldots, \mathcal{C}_{L-1})$ is block diagonal, where each $\mathcal{C}_k \in \mathbb{C}^{N \times N}$.*

PROOF. C has Cholesky decomposition

$$C = (Ry)(Ry)^T \in \mathbb{R}^{LN \times LN} \quad \text{where } Ry = \begin{pmatrix} \vdots \\ -\ R_{-k} y_i\ - \\ \vdots \end{pmatrix}_{i,k},$$

so $C \succeq 0$ has rank L. Since the inner product of two observations is invariant under a global shift of all the observations, C is composed of $N \times N$ circulant blocks of size $L \times L$. After permuting the rows and columns of C, one may write C as a block circulant matrix

$$P^T C P = \begin{pmatrix} C_0 & C_1 & \cdots & C_{L-1} \\ C_{L-1} & C_0 & \cdots & C_{L-2} \\ \vdots & \vdots & \ddots & \vdots \\ C_1 & C_2 & \cdots & C_0 \end{pmatrix},$$

where $(C_k)_{ij} = \langle y_i, R_{-k} y_j \rangle$ encodes the pairwise information between y_i and y_j, and P is the appropriate permutation matrix. $P^T C P$ is block diagonalized by the block Discrete Fourier Transform matrix $DFT_L \otimes I_N$ [24], where $\otimes$ denotes the Kronecker product and DFT_L is the matrix representation of the normalized Discrete Fourier Transform on $\mathbb{C}^L$. $\mathcal{P} = (DFT_L \otimes I_N)P^T \in \mathbb{C}^{NL \times NL}$ is thus a unitary transformation such that

$$\mathcal{P}C\mathcal{P}^* = \operatorname{diag}(\mathcal{C}_0, \ldots, \mathcal{C}_{L-1}); \quad \mathcal{C}_k \in \mathbb{C}^{N \times N} \qquad (20)$$

is block diagonal. These block diagonals are componentwise Discrete Fourier Transforms of the entries of $C_0, C_1, \ldots, C_{L-1}$:

$$\mathcal{C}_k = \sum_{l=0}^{L-1} e\left(\tfrac{kl}{L}\right) C_l = \left\{ L \cdot \mathcal{F}(y_i, k)\mathcal{F}^*(y_j, k) \right\}_{ij}$$

by Lemma A.1. Note also that each $\mathcal{C}_k$ is Hermitian and rank 1 (C is of rank L, and each of the blocks $\mathcal{C}_k$ has positive rank). The unique nonzero eigenvalue λ_k of $\mathcal{C}_k$ is given by

$$\lambda_k = \operatorname{tr}(\mathcal{C}_k) = L \sum_{i=1}^N |\mathcal{F}(y_i, k)|^2. \qquad (21)$$

□

For notation reference in the next lemma, consider a primal semidefinite problem of the form

$$\begin{aligned} &\text{minimize} && \operatorname{tr}(C_0 V) \\ &\text{subject to} && V \succeq 0 \quad \text{and} \quad \operatorname{tr}(C_i V) = b_i, \end{aligned} \qquad (22)$$

where V is the semidefinite matrix to optimize over, and $C_i \in \mathbb{R}^{n \times n}$ are data matrix constraints and $b = (b_1, \ldots, b_m)$ are given, for $i \in \{1, 2, \ldots, m\}$. For a finite group G acting on a finite space S, let P_g denote the permutation matrix associated with the action of $g \in G$. A matrix $A \in \mathbb{C}^{|S| \times |S|}$ is G-invariant if $P_g^T A P_g = A$.

LEMMA B.2. *In the semidefinite program (22), suppose the constraint matrices $C_0, C_1, \ldots, C_m$ are G-invariant. Then there is a solution V to the SDP which is also G-invariant.*

PROOF. If V' is a solution of (22), then the matrix $V \in \mathbb{C}^{n \times n}$ given by

$$V = \mathcal{R}_G(V) := \frac{1}{|G|} \sum_{g \in G} P_g^T V' P_g$$

is G-invariant, feasible, and has the same objective value as V'. The averaging map $\mathcal{R}_G$ is known as the Reynolds operator.

This is a special case of a more general theorem about symmetry reductions in semidefinite programs from the theory of $*$-matrix algebras [9]. Informally speaking, if all of the constraint matrices $C_0, C_1, \ldots, C_m$ satisfy a class of symmetries, then there is a real solution V to (22) which also satisfies the same symmetries (for details refer to [9], Corollary 2.5.2). $\square$

THEOREM B.3. *There is a matrix $V \in \mathbb{R}^{NL\times NL}$ composed of circulant $L \times L$ blocks which solves the program (11), excluding the positivity constraint $V \succeq 0$. V has rank L, corresponding to one eigenspace of eigenvalue N/L. This eigenspace contains the vector $v^{\mathsf{phase}} \in \mathbb{R}^{NL}$ satisfying*

$$v^{\mathsf{phase}}_{ik} = \frac{1}{\sqrt{L}} \sum_{l=0}^{L-1} e\left(\tfrac{-kl}{L}\right) \cdot \frac{\mathcal{F}(y_1, l)\mathcal{F}^*(y_i, l)}{|\mathcal{F}(y_1, l)\mathcal{F}^*(y_i, l)|}.$$

PROOF. The existence of a solution V composed of circulant blocks follows by a block-wise application of the Reynolds operator from Lemma B.2. Let $\mathcal{P}$ be the unitary matrix defined in Lemma 2.2. Then, $\mathcal{P}V\mathcal{P}^* = \mathrm{diag}(\mathcal{V}_0, \ldots, \mathcal{V}_{L-1})$ where $\mathcal{V}_k = \sum_{l=0}^{L-1} e\left(\frac{kl}{L}\right) V_l \in \mathbb{C}^{N\times N}$ is positive semidefinite. The SDP constraints

$$V \in \mathbb{R}_{NL\times NL}, \quad V \succeq 0, \quad V_{ik;il} = 0 \text{ for } k \neq l, \quad \sum_{k,l} V_{ik;jl} = 1$$

are respectively equivalent to the Fourier side constraints

$$\mathcal{V}_k = \overline{\mathcal{V}_{L-k}}, \quad \mathcal{V}_k \succeq 0, \quad (\mathcal{V}_k)_{ii} = 1/L, \quad (\mathcal{V}_0)_{ij} = 1/L. \quad (23)$$

Since $\mathcal{V}_k \succeq 0$, the magnitude of each entry of $\mathcal{V}_k$ is bounded by the maximum of its diagonal entries, which is $1/L$. Hence

$$\mathrm{tr}(CV) = \sum_{k\in\mathbb{Z}_L} \mathrm{tr}(\mathcal{C}_k\mathcal{V}_k) \leq \frac{1}{L} \sum_{k\in\mathbb{Z}_L} \sum_{i,j\in[N]} |(\mathcal{C}_k)_{ij}|, \quad (24)$$

with equality occuring when $\mathcal{V}_k$ has the entries of $\mathcal{C}_k$, normalized to magnitude $1/L$. Here, the $\mathcal{V}_k$'s are also rank-one matrices

$$(\mathcal{V}_k)_{ij} = \frac{\mathcal{F}(y_i, l)\mathcal{F}^*(y_j, l)}{|\mathcal{F}(y_i, l)\mathcal{F}^*(y_j, l)|}.$$

Define $w_i \in \mathbb{R}^L$ to be the vector with entries $(w_i)_k = (\mathcal{V}_k)_{i,1}$ (the choice of the index 1 is arbitrary). The concatenation of the w_i's is an eigenvector of the block diagonalization of V, and thus $v^{\mathsf{phase}} := \mathcal{P}^*w$ is an eigenvector of V. Its entries are given by

$$v^{\mathsf{phase}}_{ik} := \mathcal{F}^*(w_i, k) = \frac{1}{\sqrt{L}} \sum_{l=0}^{L-1} e\left(\tfrac{-kl}{L}\right) \cdot \frac{\mathcal{F}(y_1, l)\mathcal{F}^*(y_i, l)}{|\mathcal{F}(y_1, l)\mathcal{F}^*(y_i, l)|}.$$

Let $W^{\mathsf{phase}} \in \mathbb{R}^{NL\times L}$ be the matrix whose column vectors are generated by circulating the N blocks of L entries of v^{phase}. Its columns are linearly independent since the top $L \times L$ block of W^{phase} is the identity matrix I_L, and the column vectors span the eigenspace of V. $\square$

C. NP-HARDNESS OF QUASI MLE

A special case of the Unique-Games class of problems is that of MAX-2LIN(q). The labels $\{x_i\}$ can be thought of as equivalence classes modulo q, and each constraint is represented by a linear equation in two variables $a_{ij}x_i + b_{ij}x_j \equiv c_{ij} \pmod q$. The subclass Γ-MAX-2LIN(q) consists of the cases where each constraint has the form $x_i - x_j \equiv c_{ij} \pmod q$. It is known to be as hard as the general Unique-Games problem [17]. Associating each variable x_i as a vertex of a graph and each constraint as an edge, the instance is associated with a graph $G = (V(G), E(G))$, where $V(G) = [N]$ and $|E(G)| = M$.

THEOREM C.1. *Consider the problem* ALIGN$(y_1, \ldots, y_N)$*: for vectors $y_1, \ldots, y_N \in \mathbb{R}^L$, find the shifts $\boldsymbol{\ell} = (l_1, \ldots, l_N)$ which maximize*

$$\mathcal{A}(l_1, \ldots, l_N) = \sum_{i,j\in[N]} \langle R_{-l_i}y_i, R_{-l_i}y_j\rangle.$$

Let $\mathcal{A}^* = \max_{\boldsymbol{\ell}} \mathcal{A}(\boldsymbol{\ell})$. It is NP-hard (under randomized reductions) to esimate $\mathcal{A}^*$ within $16/17 + \varepsilon$ of its true value. It is UG-hard (under randomized reductions) to estimate $\mathcal{A}^*$ within any constant factor.

We demonstrate this by a poly-time approximation preserving reduction from (a connected instance of) Γ-MAX-2LIN(q). Suppose that at most ρ^* fraction of the M constraints $x_i - x_j \equiv c_{ij} \pmod q$ may be simultaneously satisfiable.

Let poly(M) be the space of integer functions which are bounded by polynomial order, i.e. $f \in \mathsf{poly}(M)$ iff there are constants C, k such that $f(M) \leq CM^k$ for all $M \geq 1$. We say that an event occurs w.h.p if it occurs with probability $1 - \epsilon(q, G)$, where $\frac{1}{\epsilon(q,G)} \notin \mathsf{poly}(q \cdot |E(G)|)$. Notice under this definition that if poly(qM) events occur w.h.p, then by an union bound the event that all occur will also be w.h.p.

Construct a parameter $s = s(q, M) \in \mathsf{poly}(qM)$. We take the vectors $y_1, \ldots, y_N$ to be of length $L = qMs$. The indices of the vector y_i can be expressed in mixed radix notation by elements of the tuple $(\mathbb{Z}_q, E(G), [s])$. For example, we would associate the tuple index (x, e_k, t) of y_i by the index $x \cdot qM + e_k \cdot M + t$, where e_k is the kth edge in some enumeration of $E(G)$. Note that a shift by $c \cdot qM$ takes this tuple index to $(x, e_k, t) + c \cdot qM \to (x + c, e_k, t)$.

For each edge constraint $x_i - x_j \equiv c_{ij}$, let z_{ij} be a vector uniformly at random chosen from $\{\pm 1\}^s$. Assign the entries $(0, \{i, j\}, \cdot)$ of y_i to z_{ij}, and the entries $(c_{ij}, \{i, j\}, \cdot)$ of y_j to z_{ij}. The remaining entries of the y_i's will be sampled from i.i.d Rademacher random variables ($\{\pm 1\}$ with probability $1/2$). Intuitively, a relative shift of $c_{ij} \cdot qM$ between y_i and y_j will produce a large inner product due to the overlapping of the z_{ij}'s, while any other shift between them would produce low inner products.

LEMMA C.2. *Suppose $\gamma \in \mathsf{poly}(qM)$. Consider two random vectors y_1, y_2 of length γ whose entries are i.i.d sampled from the Rademacher distribution. W.h.p, for any $0 < \epsilon \ll 1$, the inner product of every possible shift $R_\ell y_1$ of y_1 with y_2 is bounded in absolute value by $\sqrt{\gamma} \cdot m^\epsilon$.*

PROOF. By independence, each inner product is the sum of γ independent Bernoulli random variables which take values ± 1 with probability $1/2$. Hoeffding's inequality indicates

$$\Pr(\langle R_\ell y_1, y_2\rangle \geq \sqrt{\gamma} \cdot (qM)^\epsilon) \leq 2\exp\left\{-(qM)^\epsilon/2\right\},$$

so $\frac{1}{\Pr(\langle R_\ell y_1, y_2\rangle \geq \sqrt{\gamma}(qM)^\epsilon)} \notin \mathsf{poly}(qM)$. Union bounding over all $\gamma = \mathsf{poly}(qM)$ such inner products, w.h.p all of the inner products are simultaneously bounded by $\sqrt{\gamma} \cdot (qM)^\epsilon$. $\square$

We say an edge $\{i,j\} \in E(G)$ is c_{ij}-satisfied under a labelling $\boldsymbol{\ell}$ if $l_i - l_j \equiv c_{ij} qM \pmod{L}$. From Lemma C.2, we observe that w.h.p, for any choices of shifts l_i, l_j,

$$|\langle R_{-l_i} y_i,\ R_{-l_j} y_j\rangle - s \cdot \delta(\{i,j\} \in E(G) \text{ is } c_{ij}\text{-satisfied})| < (qM)^{\epsilon}\sqrt{L}.$$

It follows that if the labelling $\boldsymbol{\ell}$ induces exactly k c_{ij}-satisfied edges, w.h.p

$$\begin{aligned}\left|\mathcal{A}(\boldsymbol{\ell}) - 2ks\right| &< k(qM)^{\epsilon}\sqrt{L} + \left(N^2 - k\right)(qM)^{\epsilon}\sqrt{L} \\ &\le q^{\epsilon} M^{2+\epsilon}\sqrt{L}. \end{aligned} \tag{25}$$

THEOREM C.3. *From a given labelling $\boldsymbol{\ell}$, w.h.p one may in polynomial time construct $(x_1, \ldots, x_N) \in \mathbb{Z}_q^N$ satisfying at least $\left(\mathcal{A}(\ell) - q^{\epsilon} M^{2+\epsilon}\sqrt{L}\right)/(2s)$ edge-constraints $x_i - x_j \equiv c_{ij}$. Conversely, w.h.p there is a labelling $\boldsymbol{\ell}^{\mathsf{max}}$ w.h.p satisfying $\mathcal{A}\left(\boldsymbol{\ell}^{\mathsf{max}}\right) > 2s \cdot \rho^* M - q^{\epsilon} M^{2+\epsilon}\sqrt{L}$.*

PROOF. Consider the subgraph H of G with vertex set $V(G)$ and edge set comprising all edges c_{ij}-satisfied under $\boldsymbol{\ell}$. For each connected component of H, arbitrarily choose a vertex i of the component to have $x_i = 0$. Follow a spanning tree of each connected component and assign each neighbor j by $x_j \equiv x_i - c_{ij} \pmod{q}$. The first implication of the lemma follows immediately by applying (25). Conversely, construct the labelling $\boldsymbol{\ell}^{\mathsf{max}}$ by setting $l_i^{\mathsf{max}} = x_i \cdot qM$. This labelling induces at least $\rho^* M$ c_{ij}-satisfied edges, and (25) completes the lemma. □

It is UG-hard (and thus NP-hard assuming the Unique Games conjecture) to approximate Γ-MAX-2LIN(q) within any constant factor [17]. As a corollary to Theorem C.3, any poly-time approximation ratio for ALIGN will also be a poly-time approximation ratio for Γ-MAX-2LIN(q). Thus the same hardness of approximation results for Γ-MAX-2LIN(q) hold for ALIGN (under randomized reductions). MAX-CUT is a special case of Γ-MAX-2LIN(q). Since it is NP-hard to approximate MAX-CUT within $16/17+\varepsilon$ (independent of the Unique Games conjecture), it is NP-hard (under randomized reductions) to approximate ALIGN within $16/17 + \varepsilon$.

D. STABILITY

LEMMA D.1. $\mathbb{E}[\eta_{ij}|\xi_i] \le \mu_\eta \|\xi_i\|$ *and* $\mathbb{E}[\eta_{ij}^2|\xi_i] \le \sigma_\eta \|\xi_i\|^2$, *where* $\mu_\eta = 2\sigma(\log L + 1)$ *and* $\sigma_\eta^2 = 4\sigma^2(\log^2 L + 1)$.

PROOF. Take $\zeta \sim \mathcal{N}(0, I_L)$, and choose a unit vector $z \in \mathbb{R}^L$. Define $\eta(z) = \max_l \eta_l(z)$, where the distribution of $\eta_l(z) = |\langle R_l z, \zeta\rangle|$ is independent of the choice of z. Union bounding,

$$\begin{aligned}\Pr\left(\eta(z) \ge t\right) &\le \sum_{l \in \mathbb{Z}_L} \Pr\left(\eta_l(z) \ge \tfrac{t}{2}\right) \\ &= \sum_l \Pr\left(|\zeta_l| \ge \tfrac{t}{2}\right) \le \min\left\{1, \frac{2L}{t} \cdot \frac{e^{-t^2/2}}{\sqrt{2\pi}}\right\}.\end{aligned}$$

Thus, for $T > 1$,

$$\begin{aligned}\mu_\eta = \mathbb{E}\,\eta(z) &= \int_0^\infty \Pr\left(\eta(z) \ge t\right) dt \\ &\le T + \frac{2L}{\sqrt{2\pi}} \int_T^\infty t e^{-t^2/2}\, dt \\ &\le T + L e^{-T}\end{aligned}$$

and

$$\begin{aligned}\sigma_\eta^2 = \mathbb{E}\,\eta(z)^2 &= \int_0^\infty 2t \Pr\left(\eta(z) \ge t\right) dt \\ &\le T^2 + \frac{L}{\sqrt{2\pi}} \int_T^\infty t e^{-t^2/2}\, dt \\ &\le T^2 + L e^{-T}.\end{aligned}$$

The lemma follows by taking $T = \log L$, since $\eta_{ij} = 2\sigma\|\xi_i\| \cdot \eta\big(\xi_i/\|\xi_i\|\big)$ conditional on ξ_i. □

LEMMA D.2. *Let $t > 0$. With probability at least $1 - 2Ne^{-t^2 N}$,*

$$\sum_{i \ne j} (\eta_{ij} + \eta_{j0}) D_{ij} \le (\mu_\eta + \sigma_\eta t)\left(2\|x\| + \frac{1}{N}\sum_i \|\xi_i\|\right) N^2. \tag{26}$$

holds for all $D_{ij} \in [0,1]$.

PROOF. Define the random variables $Z_{ij} = (\eta_{ij} + \eta_{j0}) D_{ij} \le \eta_{ij} + \eta_{j0}$. For fixed i, the η_{ij}'s are independent positive random variables. Applying the one-sided tail bound for independent positive random variables by Maurer [27], the probability the event

$$\sum_j (\eta_{ij} + \eta_{j0}) - \sum_j \mathbb{E}\left(\eta_{ij} + \eta_{j0}\right) \ge t_i \sqrt{\sum_j \mathbb{E}\left(\eta_{ij}^2 + \eta_{j0}^2\right)}$$

occurs is bounded under $2e^{-t_i^2}$. Choose $t_i = t\sqrt{N}$. By union bound, with probability at least $1 - 2Ne^{-t^2 N}$,

$$\begin{aligned}\sum_{i,j} Z_{ij} &\le \sum_i \mu_\eta (2\|x\| + \|\xi_i\|) N + \sum_i t\sqrt{N^2\sigma_\eta^2(4\|x\|^2 + \|\xi_i\|^2)} \\ &\le (\mu_\eta + \sigma_\eta t)\left(2\|x\| + \frac{1}{N}\sum_i \|\xi_i\|\right) N^2.\end{aligned}$$

□

THEOREM D.3. *With probability $1 - e^{-N + o(N)}$, the solution to SDP (11) satisfies*

$$\sum_{i,j} D_{ij} \le \frac{(\|x\| + \sigma^2\sqrt{L}) \cdot 12 \log eL}{\Delta} \cdot N^2.$$

PROOF. Let V be a solution matrix to the SDP, so that $\operatorname{tr}(CV) \ge \operatorname{tr}(CV^{int})$. It follows from Lemma 4.1 that

$$\Delta \sum_{ij} D_{ij} \le \sum_{ij} (\eta_{ij} + \eta_{j0}) D_{ij}.$$

The χ^2-tail bound of Laurent-Massart [22] states

$$\Pr\left(\frac{1}{\sigma^2}\sum_i \|\xi_i\|^2 \le NL + 2\sqrt{NLt} + 2t\right) \ge 1 - \exp\{-t\}. \tag{27}$$

When $t = NL$ this gives

$$\sum_i \|\xi_i\| \le \sqrt{N \sum_i \|\xi_i\|^2} \le \sigma N \sqrt{5L}.$$

with probability at least $1 - e^{-NL}$. Union bounding with the tail bound of Lemma D.2, and applying Lemma D.1,

$$\begin{aligned}\Delta \sum_{ij} D_{ij} &\le (\mu_\eta + \sigma_\eta)\left(2\|x\| + \frac{1}{N}\sum_i \|\xi_i\|\right) N^2 \\ &\le 12 \log eL (\|x\| + \sigma\sqrt{L}) N^2\end{aligned}$$

with probability at least $1 - e^{-N+o(N)}$. $\square$

LEMMA D.4. *Let $0 < \delta, \varepsilon, \varepsilon' \ll 1$ be small constants, and define $\alpha_{JL}(x) = (1+\delta)x + \varepsilon$. There is a set of unit vectors $\mathfrak{N}$ of size at most*

$$\exp\left(\mathcal{O}\left(\delta^{-2}\log(1/\varepsilon)\log(1/\varepsilon')\right)\right)$$

such that for any set of unit vectors $\{v_i\}$, there is a map $\varphi : \{v_i\} \to \mathfrak{N}$ satisfying the inequality

$$\alpha_{JL}^{-1}\left(\|v_i - v_j\|^2\right) \le \|\varphi(v_i) - \varphi(v_j)\|^2 \le \alpha_{JL}\left(\|v_i - v_j\|^2\right)$$

for at least $1 - \varepsilon'$ fraction of the pairs $(i, j) \in [N] \times [N]$.

PROOF. This lemma appears frequently in SDP literature, and in particular is used to analyze adversarial semi-random Unique-Games instances in [19]. Notice that the size of the set $\mathfrak{N}$ is independent of the number of vectors in the set $\{v_i\}$.

Construct a $\varepsilon/32$-net $\mathfrak{N}$ of the unit hypersphere in a $\mathcal{O}(\delta^2 \log(1/\varepsilon'))$-dimensional space $\mathfrak{L}$. By the (strong version of the) Johnson-Lindenstrauss lemma, there is a randomized mapping $\varphi' : \{v_i\} \to \mathfrak{L}$ satisfying

$$(1-\delta/2)\|v_i - v_j\|^2 \le \|\varphi'(v_i) - \varphi'(v_j)\|^2 \le (1+\delta/2)\|v_i - v_j\|^2$$

with probability at least $1-\varepsilon'$. Define $\varphi(v_i)$ to be the closest point to $\varphi'(v_i)$ in $\mathfrak{N}$, and observe that

$$(1-\delta/2)x - \varepsilon \ge \frac{x-\varepsilon}{1+\delta} = \alpha_{JL}^{-1}(x), \quad (1+\delta/2)x + \varepsilon \le \alpha_{JL}(x)$$

for all $x > 0$, so φ satisfies the conditions of the lemma. $\square$