

Dynamic angular synchronization under smoothness constraints

Ernesto Araya

ARAYA@MATH.LMU.DE

Department of Mathematics

Ludwig-Maximilians-Universität München

Geschwister-Scholl-Platz 1, Munich, 80539, Germany

Mihai Cucuringu

MIHAI@MATH.UCLA.EDU

Department of Mathematics

University of California Los Angeles

520 Portola Plaza, Los Angeles, CA 90095, USA

Department of Statistics & Oxford-Man Institute of Quantitative Finance

University of Oxford

24-29 St Giles', Oxford OX1 3LB, United Kingdom

Hemant Tyagi

HEMANT.TYAGI@NTU.EDU.SG

Division of Mathematical Sciences,

School of Physical and Mathematical Sciences,

Nanyang Technological University,

21 Nanyang Link, 637371, Singapore

Editor: Jie Peng

Abstract

Given an undirected measurement graph $\mathcal{H} = ([n], \mathcal{E})$, the classical angular synchronization problem consists of recovering unknown angles $\theta_1^*, \dots, \theta_n^*$ from a collection of noisy pairwise measurements of the form $(\theta_i^* - \theta_j^*) \bmod 2\pi$, for all $\{i, j\} \in \mathcal{E}$. This problem arises in a variety of applications, including computer vision, time synchronization of distributed networks, and ranking from pairwise comparisons. In this paper, we consider a dynamic version of this problem where the angles, and also the measurement graphs evolve over T time points. Assuming a smoothness condition on the evolution of the latent angles, we derive three algorithms for joint estimation of the angles over all time points. Moreover, for one of the algorithms, we establish non-asymptotic recovery guarantees for the mean-squared error (MSE) under different statistical models. In particular, we show that the MSE converges to zero as T increases under milder conditions than in the static setting. This includes the setting where the measurement graphs are highly sparse and disconnected, and also when the measurement noise is large and can potentially increase with T . We complement our theoretical results with experiments on synthetic data.

Keywords: angular synchronization, dynamical networks, smoothness constrained estimators

1. Introduction

The group synchronization problem, which consists of recovering group elements g_1^*, \dots, g_n^* from a potentially sparse set of noisy pairwise ratios $g_i^* g_j^{*-1}$, has attracted considerable attention in the last decade, starting with the seminal paper of Singer (Singer, 2011). The motivation for this problem originated from the cryo-EM (Shkolnisky and Singer, 2012) and sensor network localization (Cucuringu et al., 2012a) literatures, but instantiations of this problem occur in a variety of application domains, including computer vision (Arie-Nachimson et al., 2012) in the context of the structure-from-motion problem (Özyeşil et al., 2017), ranking from pairwise comparisons (Cucuringu, 2016), community detection in social networks (Cucuringu, 2015), NMR spectroscopy (Cucuringu et al., 2012b), temporal ordering and registration of images in image processing applications (Dsilva et al., 2015), and time synchronization of distributed networks (Giridhar and Kumar, 2006).

In the most typical setting, the group elements g_1^*, \dots, g_n^* are represented as the nodes of a simple undirected graph $\mathcal{H} = ([n], \mathcal{E})$, where an edge $\{i, j\}$ is present if and only if the observer has access to a noisy version of $g_i^* g_j^{*-1}$. In the particular case where \mathcal{H} is connected and the exact value $g_i^* g_j^{*-1}$ is observed, the recovery problem has a straightforward solution. Indeed, it suffices to consider a spanning tree of \mathcal{H} , assign any value to an arbitrarily chosen root, and the remaining $n - 1$ values are found by traversing the edges of tree, solving each time a simple equation. Consequently, these $n - 1$ values are uniquely determined up to an global “shift” defined by the chosen value of the root node. On the other hand, if \mathcal{H} is disconnected, there is insufficient information to determine a global solution, since there are no comparisons between the elements of different connected components.

1.1 Angular synchronization

The simplest groups for which the synchronization problem has been studied are \mathbb{Z}_2 , in the context of community detection and signed clustering with two clusters (Cucuringu, 2015; Liu et al., 2023), and the non-compact group \mathbb{R} arising in the synchronization of clocks in a distributed network (Giridhar and Kumar, 2006). The special orthogonal group $\text{SO}(d)$ is the most commonly studied group with a particular emphasis on $\text{SO}(2)$ – often referred to as angular synchronization – which finds applications in various practical domains, including 2D sensor network localization (Cucuringu et al., 2012a) and ranking from pairwise comparisons (Cucuringu, 2016).

In the angular synchronization problem, which is the focus of this paper, the goal is to recover n unknown angles $\theta_1^*, \dots, \theta_n^* \in [0, 2\pi)$ from a set of m noisy measurements of their pairwise offsets $\Theta_{ij} = (\theta_i^* - \theta_j^*) \bmod 2\pi$. Equivalently, we can represent each angle θ_i^* by a unit modulus complex number $g_i^* = e^{i\theta_i^*}$ and the pairwise offsets by $g_i^* g_j^{*H}$, where for any $z \in \mathbb{C}$, z^H represent its complex conjugate. As explained above, the available pairwise measurements can be encoded into the edges of a measurement graph \mathcal{H} , which in most typical applications can be modeled as a sparse random geometric graph (e.g., disc graph), a k -nearest-neighbor graph, or an Erdős-Rényi graph. The difficulty of the problem is amplified, on one hand, by the amount of **noise** in the available pairwise offset measurements, and on the other hand by the **sparsity** of the measurements, since typically, $|\mathcal{E}| \ll \binom{n}{2}$, i.e., only a small subset of all possible pairwise offsets are available to the user.

Aiming to reconstruct the angle offsets as best as possible, the initial approach of Singer (Singer, 2011) considered a quadratic maximization problem over the angular domain, that aims to maximize the number of preserved angle offsets. Denote by A the measurement matrix of size $n \times n$ (which can be seen as a weighted adjacency matrix of $\mathcal{H} = ([n], \mathcal{E})$), where for each $\{i, j\} \in \binom{[n]}{2}$, one considers the angular embedding

$$A_{ij} = \begin{cases} e^{i\Theta_{ij}} & \text{if } \{i, j\} \in \mathcal{E} \\ 0 & \text{if } \{i, j\} \notin \mathcal{E} \end{cases},$$

with diagonal entries $A_{ii} = 0$. Note that A is Hermitian since $\Theta_{ji} = (-\Theta_{ij}) \bmod 2\pi$. One then considers the following maximization problem

$$\max_{\theta_1, \dots, \theta_n \in [0, 2\pi)} \sum_{i,j=1}^n e^{-i\theta_i} A_{ij} e^{i\theta_j} \equiv \max_{g_1, \dots, g_n \in \mathbb{C}; |g_i|=1} \sum_{i,j=1}^n g_i^* A_{ij} g_j, \quad (1)$$

which is in general NP hard to solve (Zhang and Huan, 2006). This intractability is usually circumvented by considering continuous relaxations of (1), leading typically to a spectral algorithm (Singer, 2011), or a semi-definite program (SDP) (Singer, 2011; Bandeira et al., 2017). Other algorithmic approaches include message passing algorithms (Perry et al., 2018; Bandeira et al., 2018), those based on graph neural networks (He et al., 2024), and also the generalized power method (Liu et al., 2017; Boumal, 2016). The statistical performance of algorithms is typically studied under the following two noise models.

- **Additive Gaussian noise (AGN) model (Bandeira et al., 2017).** Here, \mathcal{H} is the simple complete graph and the observations are of the form $g_i^* g_j^{*H} + \sigma W_{ij}$, where $(W_{ij})_{i < j}$ are independent complex standard Gaussian's, $W_{ii} = 0$ (for all i), and $\sigma > 0$ represents the noise level. Note that (1) is the maximum likelihood estimator (MLE) for the AGN model.
- **Outliers model (Singer, 2011).** Here, \mathcal{H} is modeled as an Erdős-Rényi graph with connection probability p . When p is small, the resulting measurement graph will be sparse and potentially disconnected (e.g., when $p = o(\frac{\log n}{n})$). Under this model, whenever $\{i, j\}$ is an edge in \mathcal{H} , we observe either the true value $g_i^* g_j^{*H}$ with probability η , or a value uniformly distributed on the complex unit circle with probability $1 - \eta$. The parameter η quantifies the level of noise.

Several results now exist pertaining to the statistical error analysis of the aforementioned algorithms for the above models. For instance, the SDP is known to be tight¹ for the AGN model if σ is not too large w.r.t n (Bandeira et al., 2017; Zhong and Boumal, 2018). In particular, this was first shown in (Bandeira and van Handel, 2016) for $\sigma = O(n^{1/4})$, and was later improved to $\sigma = O(\sqrt{n/\log n})$ in (Zhong and Boumal, 2018). Recovery guarantees for the spectral method exist for both the Outliers model (Singer, 2011), and the AGN model (e.g., (Boumal, 2016; Zhong and Boumal, 2018)). It was shown in (Zhong and Boumal, 2018) for the AGN model that the spectral method recovers g^* with an ℓ_2 error bound of $O(\sigma)$, and an ℓ_∞ error bound of $O(\sigma\sqrt{\log n/n})$, provided $\sigma = O(\sqrt{n/\log n})$.

1. This means that its global solution recovers that of (1).

For this regime of σ , it was also shown in (Zhong and Boumal, 2018) that (with spectral initialization) the iterates of the generalized power method (GPM) converge linearly to the global solution of (1). This improved upon previous convergence results for the GPM (Boumal, 2016; Liu et al., 2017) which required stronger conditions on σ .

Towards dynamic group synchronization. In many applications of group synchronization, the underlying measurement graph, along with the associated pairwise data evolve with time. This is the case, e.g., for the ranking problem where the strengths of players, and hence the pairwise comparison outcomes evolve with time (e.g., sports tournaments (Cattelan et al., 2013; Lopez et al., 2018), recommender systems), and has been the subject of recent work (Bong et al., 2020; Araya et al., 2023; Karlé and Tyagi, 2023). In (Cucuringu, 2016), this problem was formulated as an instance of the angular synchronization problem, along with an algorithm SYNCRANK which leads to state-of-art performance empirically. It is also the case for the signed community detection problem (with two communities), where the network and the underlying communities can evolve with time (e.g., (Chen et al., 2020)). This motivates us to initiate a principled study of group synchronization in a dynamic setting where the latent group elements, and the measurement graphs can both evolve with time.

1.2 Dynamic angular synchronization

In this paper, we introduce a dynamic version of the angular synchronization problem. Here, we are given a sequence of measurement graphs of the form $\mathcal{H}_k = ([n], \mathcal{E}_k)$, over T time points $k = 1, \dots, T$. At each k , we are given noisy pairwise information about the latent signal $g^*(k) \in \mathbb{C}^n$ as per the AGN model, or the Outliers model; see Section 2.2. While the measurements are assumed to be independently generated over time, we assume that $g^*(k)$'s evolve smoothly in the sense that

$$\sum_{k=1}^{T-1} \|g^*(k) - g^*(k+1)\|_2^2 \leq S_T, \quad (2)$$

for a suitably small smoothness parameter S_T . The above condition essentially stipulates that the quadratic variation of $g^*(k)$ w.r.t the path graph on T vertices is small. A similar assumption was recently used in (Araya et al., 2023) in the context of dynamic translation synchronization, which is also the main motivation for the present paper.

Our goal then is to recover the vector $g^* \in \mathbb{C}^{nT}$, obtained by column-stacking $g^*(k)$'s. Specifically, we aim to output an estimate \hat{g} of g^* that is weakly-consistent w.r.t T in terms of the mean squared error (MSE), i.e.,

$$\frac{1}{T} \|\hat{g} - g^*\|_2^2 = \frac{1}{T} \sum_{k=1}^T \|\hat{g}(k) - g^*(k)\|_2^2 \xrightarrow{T \rightarrow \infty} 0. \quad (3)$$

The following points are useful to keep in mind.

1. When all the graphs are connected, one could estimate each $g^*(k)$ by only using the pairwise information at time k . However this will typically lead to $\frac{1}{T} \|\hat{g} - g^*\|_2^2 = \Omega(1)$ as $T \rightarrow \infty$. Exploiting the smoothness information over k is then, in a sense, necessary to ensure (3). This is particularly relevant in the setting where the noise level is large.

2. If each measurement graph is disconnected, then estimating $g^*(k)$'s individually at each k will be meaningless. However, one might still expect, at least intuitively, that a weakly consistent estimate of g^* can be obtained by leveraging smoothness. For instance, if $S_T = 0$, then this would be possible provided the union graph $([n], \cup_k \mathcal{E}_k)$ is connected, where we would simply take the average of all the pairwise information.

Main contributions. We summarize our main contributions as follows.

- We propose three algorithms for this problem, based on different relaxation-based approaches.
 1. *Global sphere relaxation* (GTRS-DynSync; Algorithm 1): Optimizes a smoothness-penalized data fidelity term with solutions restricted to a sphere in \mathbb{C}^{nT} . This leads to a trust-region subproblem (TRS) which is efficiently solvable.
 2. *Local sphere relaxation with global smoothing* (LTRS-GS-DynSync; Algorithm 2): Solves sphere-relaxed synchronization problems at each time point k (via a TRS formulation), stacks the local solutions, and projects the stacked vector onto the low-frequency eigenspace of a smoothness operator (motivated by the Laplacian eigenmaps estimator (Sadhanala et al., 2016)).
 3. *Global matrix denoising and local synchronization* (GMD-LTRS-DynSync; Algorithm 3): Projects the measurement matrix onto the low-frequency eigenspace of a smoothness operator (matrix denoising), followed by locally solving (at each time k) the sphere-relaxation of the angular synchronization problem via a TRS formulation.
- As the TRS estimator is hard to analyze theoretically, we analyze a slight modification of Algorithm 3, namely Algorithm 4, where the TRS estimator for the local synchronization step is replaced by an alternative spectral approach (see also Remark 3). We provide theoretical guarantees for Algorithm 4, in the form of non-asymptotic high probability bounds for the global ℓ_2 estimation error, under the dynamic versions of the AGN and Outliers model (see Corollaries 17 and 19 respectively). In particular, under mild smoothness assumptions, we prove that (3) holds for the output of Algorithm 4, and give precise non-asymptotic bounds from which the convergence rates can be deduced.
 1. In the case of the Outliers model, our results apply even when the measurement graphs are very sparse – each being potentially disconnected (see Remarks 15 and 20).
 2. For the AGN model, our results show that Algorithm 4 succeeds even when the noise level σ is large w.r.t n , and is allowed to grow with T as $O(T^\alpha)$ for $\alpha \in [0, 1/4)$ (see Remarks 10 and 18).
- Finally, we support our theoretical findings with extensive experiments on synthetic data, for the aforementioned AGN and Outliers models. Notably, our experiments demonstrate that Algorithm 3 usually outperforms other competing methods in terms of accuracy.

While we have focused on the $\text{SO}(2)$ group for concreteness, the framework in the present paper can be extended to other compact groups such as \mathbb{Z}_2 and $\text{SO}(d)$.

1.3 Related work

Ranking and angular synchronization. The work of (Cucuringu, 2016) considered the problem of ranking n items given a set of pairwise comparisons on their rank offsets, over a single measurement graph. The SYNCRANK algorithm, introduced in (Cucuringu, 2016), formulates the above problem as an instance of the angular synchronization problem, following a group compactification procedure of the real line. Concretely, the approach in (Cucuringu, 2016) starts by mapping the noisy rank offsets $C_{ij} \in \{-(n-1), \dots, n-1\}$ to angles $\Theta_{ij} \in [0, 2\pi\delta)$ via the transformation $\Theta_{ij} := 2\pi\delta \frac{C_{ij}}{n-1}$, where $\delta \in [0, 1)$. The resulting problem is then treated as an instance of the angular synchronization problem, and solved via the tools therein. A post-processing step is applied to mod out the best circular permutation, and recover the final rankings. While we do not integrate our algorithmic framework with SYNCRANK in this paper, it would be interesting to do so as future work, in order to evaluate its performance on dynamic ranking problems. More recently, (Huang et al., 2017; d’Aspremont et al., 2021) investigated algorithms for recovering the latent strengths $r_i^*, i = 1, \dots, n$ from noisy versions of $r_i^* - r_j^*$. This is an instance of the group synchronization over the real line, also dubbed as *translation synchronization*. The work (Huang et al., 2017) proposed an iterative procedure which involves solving a truncated least-squares problem at each iteration. Conditions were derived under which the iterates were shown to converge to the ground truth. In (d’Aspremont et al., 2021), an algorithm based on the singular value decomposition was introduced, along with a detailed theoretical analysis.

Ranking from dynamic pairwise information. A recent line of work initiated a systematic study of ranking from dynamic pairwise information, along with statistical guarantees (Bong et al., 2020; Karlé and Tyagi, 2023; Araya et al., 2023). Here, the latent strengths of the items are assumed to evolve smoothly over T time points, with the pairwise data generated independently over time. These works are the main motivation behind the present paper.

- The setup in (Bong et al., 2020; Karlé and Tyagi, 2023) assumed a dynamic Bradley-Terry-Luce (BTL) model where the strength vector satisfies a Lipschitz condition in the (continuous) time domain $[0, 1]$. Pointwise estimation error bounds were derived for a given $t \in [0, 1]$ via a two stage procedure: the first stage involved suitably “smoothing” the pairwise data while the second stage involved applying an existing ranking method from the static setup. In particular, (Karlé and Tyagi, 2023) proposed a nearest-neighbor based averaging procedure in conjunction with Rank Centrality (Negahban et al., 2017), while (Bong et al., 2020) proposed a kernel based smoothing procedure in conjunction with the MLE for the static BTL model. The results in both papers imply pointwise consistency w.r.t T , with (Karlé and Tyagi, 2023) showing this even when the individual graphs are very sparse (Erdős-Rényi graphs) and potentially disconnected.

- The work (Araya et al., 2023) considered the dynamic translation synchronization problem which is related to the BTL model for ranking. Here, the latent strength vector $r^*(k) \in \mathbb{R}^n$ ($k = 1, 2, \dots, T$) satisfies a similar smoothness assumption as (2). Given a measurement graph $\mathcal{H}_k = ([n], \mathcal{E}_k)$ at time k , we are given noisy information $r_i^*(k) - r_j^*(k) + \epsilon_{ij}(k)$ for each $\{i, j\} \in \mathcal{E}_k$, where $\epsilon_{ij}(k)$ is zero-mean subgaussian noise. Two estimators were proposed, the first of which involves solving an unconstrained smoothness-penalized least squares problem. The second estimator involves first estimating $r^*(k)$ via least-squares minimization, at each k , and then projecting the vector of estimated solutions on to the low-frequency space of a smoothness operator. High-probability bounds were obtained for the MSE, establishing their consistency when T is large and the rest of parameters are fixed. In particular, these results imply that the proposed estimators can tolerate larger levels of noise compared to the static case. Our Algorithms 1 and 2 are motivated from these two methods, the key difference being that we now need to incorporate the (relaxations of) the constraints, imposed by the group $\text{SO}(2)$. This in particular leads to solving a TRS-based sphere-relaxation which is challenging to analyze theoretically. We also remark that the analysis in (Araya et al., 2023) requires each measurement graph to be connected due to technical reasons. Our analysis for Algorithm 4, however, does not suffer from these issues, and applies even when the graphs are very sparse and disconnected. We believe that a similar procedure can be applied for the dynamic translation synchronization problem, leading to similar guarantees as in the present paper (Theorem 13 and Corollary 19).

The work (Li and Wakin, 2021) considered a generalization of the translation synchronization model where the time-dependent model parameters evolve in a Markovian manner over time. Theoretical results were obtained for estimating a latent pairwise comparison matrix at time T . In general, however, existing results for the dynamic ranking problem are essentially empirical works (Fahrmeir and Tutz, 1994; Glickman and Stern, 1998; Lopez et al., 2018; Maystre et al., 2019; Cattelan et al., 2013; Jabin and Junca, 2015) with limited theoretical guarantees.

1.4 Paper structure

The remainder of this paper is structured as follows. In Section 2 we present the setup for dynamic synchronization. In Section 2.2 we introduce the dynamic versions of the AGN and Outliers model, together with our global smoothness assumptions. Our proposed estimators and the algorithms to compute them are described in Section 2.3. Section 3, which consists of the theoretical analysis of Algorithm 4, is divided into three parts. In Sections 3.1 and 3.2 we analyse the matrix denoising stage for the AGN and Outliers model respectively. In Section 3.3 we provide an analysis for the local synchronization stage of Algorithm 4, which is valid for both noise models. Section 4 contains numerical experiments on synthetic data and Section 5 contains concluding remarks.

2. Problem setup and algorithms

In Section 2.1 we describe general notation used throughout the paper. Section 2.2 contains a formal description of the problem, while Section 2.3 formally outlines our proposed algorithms. In particular, Section 2.3 details additional notation which is needed for describing the algorithms and is also used in the analyses later in Section 3.

2.1 Notation

For any $n \in \mathbb{N}$, define $[n] = \{1, \dots, n\}$. For $A \in \mathbb{C}^{n \times p}$, A^H denotes its Hermitian conjugate transpose. We define the operators Re and Im , from $\mathbb{C}^{n \times p}$ to $\mathbb{R}^{n \times p}$, which for any matrix give their real and imaginary parts, respectively. Denote $e_1, \dots, e_n \in \mathbb{R}^n$ to be the canonical vectors. We represent the vector of all ones as $\mathbb{1}$, with the dimension being clear from the context.

Let $\|z\|_2 = \sqrt{|z_1|^2 + \dots + |z_n|^2}$ denote the Euclidean norm for any $z \in \mathbb{C}^n$. If $Z, Z' \in \mathbb{C}^{n \times m}$, their Frobenius inner product is defined by $\langle Z, Z' \rangle_F = \text{Tr}(Z^H Z')$, where $\text{Tr}(\cdot)$ denotes the trace of a matrix. The Frobenius norm $\|\cdot\|_F^2$ is the norm induced by this inner product. We denote the Kronecker product of Z and Z' by $Z \otimes Z'$, and their Hadamard product by $Z \circ Z'$. For a matrix $Z \in \mathbb{C}^{n \times m}$, $\|Z\|_{op}$ is the operator norm of Z , which is defined as its largest singular value.

Define, for $n \in \mathbb{N}$, the set $\mathcal{C}_n := \{z \in \mathbb{C}^n : |z_i| = 1, \text{ for all } i \in [n]\}$. Recall that the special orthogonal group $\text{SO}(2)$ can be represented by the group of complex numbers z , such that $|z| = 1$, with the usual complex multiplication. The concatenation operator is defined for matrices X in $\mathbb{C}^{n \times m}$ and $Y \in \mathbb{C}^{n' \times m}$, denoted as $\text{concat}(X, Y) = \begin{bmatrix} X \\ Y \end{bmatrix} \in \mathbb{C}^{(n+n') \times m}$. This definition extends to any sequence $(X_k)_{k=1}^K$, with $K \in \mathbb{N}$, and $X_k \in \mathbb{C}^{n_k \times m}$, through the recurrence $\text{concat}(X_1, \dots, X_k) = \text{concat}(\text{concat}(X_1, \dots, X_{k-1}), X_k)$. Similarly, for any natural number m , $\text{blkdiag}(X_1, \dots, X_m)$ denotes the block-diagonal matrix with square matrices X_1, \dots, X_m on the main diagonal.

Recall that the subgaussian norm of a random variable X is given by

$$\|X\|_{\psi_2} := \sup_{p \geq 1} p^{-1/2} (\mathbb{E}|X|^p)^{1/p},$$

see e.g., (Vershynin, 2012) for other equivalent definitions. Note that if $|X| \leq M$ a.s, then $\|X\|_{\psi_2} \leq M$. When X follows a normal distribution with mean μ and variance σ^2 , we write $X \sim \mathcal{N}(\mu, \sigma^2)$. We say $X \sim \mathcal{CN}(0, 1)$ if X is complex-valued and

$$\text{Re}(X), \text{Im}(X) \stackrel{\text{ind.}}{\sim} \mathcal{N}(0, 1/2).$$

We will use the classic asymptotic notation, defined as follows. We say that $f(t) = O(g(t))$ if there exists positive real numbers t_0, M , such that for all $t \geq t_0$ we have $f(t) \leq M g(t)$. Similarly, we write $f(t) = o(g(t))$ if for every constant $c > 0$, there exists a constant t_0 , such that, $f(t) \leq c g(t)$, for all $t \geq t_0$. Additionally, we say $f(t) = \Omega(g(t))$, if $f(t) = O(g(t))$, and $f(t) = \omega(g(t))$, if $g(t) = o(f(t))$. Symbols used for denoting constants (e.g., c, C, c_1 etc.) may change from line to line.

2.2 Problem setup

Consider a set of items $[n]$ and a sequence of time-indexed undirected comparison graphs $\mathcal{H}_k = ([n], \mathcal{E}_k)$ for $k = 1, \dots, T$. The set of edges \mathcal{E}_k determines which items are being compared at time k . For each $i \in [n]$, we associate time-dependent latent phases $g_i^*(k) \in \text{SO}(2)$ which will constitute the ground truth signal that we aim to recover from noisy pairwise measurements. More formally, at time $k \in [T]$, we observe for every pair $\{i, j\} \in \mathcal{E}_k$ a noisy version of $g_i^*(k)g_j^*(k)^H$. Notice that even without the addition of noise, the recovery of the ground truth phases $g_i^*(k)$, from the observation of $g_i^*(k)g_j^*(k)^H$, can only be accomplished up to a global phase-shift. That is, given angles $\theta_1, \dots, \theta_T \in [0, 2\pi)$, the quantities $(g_i^*(k) \cdot e^{i\theta_k})_{k=1}^T$ define the same relative phases as $(g_i^*(k))_{k=1}^T$. To avoid identifiability issues, we will make the following assumption on the phase of the first element of each block.

Assumption 1 (Anchoring) *For all $k \in [T]$, we assume that $g_1^*(k) = 1$.*

For each $k \in [T]$, denote $g^*(k) = \text{concat}(g_1^*(k), \dots, g_n^*(k)) \in \mathcal{C}_n$, and g^* to be formed by column-wise stacking as

$$g^* = \text{concat}(g^*(1), \dots, g^*(T)) = \begin{pmatrix} g^*(1) \\ \vdots \\ g^*(T) \end{pmatrix} \in \mathcal{C}_{nT}. \quad (4)$$

We now describe two statistical models for generating the noisy pairwise measurements at each k .

Additive Gaussian noise (AGN) model. Denoting for each $k \in [T]$, $G^*(k) := g^*(k)g^*(k)^H$, we are given the measurements $A(k) \in \mathbb{C}^{n \times n}$ where

$$A(k) = (G^*(k) - I_n) + \sigma W(k) \quad \text{for } k \in [T], \quad (5)$$

where $W(k)$ is a random complex Hermitian (noise) matrix with independent entries above the diagonal and $W_{ii}(k) = 0$ (for all i). In particular, for each $i \neq j$, $W_{ij}(k) \sim \mathcal{CN}(0, 1)$. The parameter $\sigma \geq 0$ denotes the noise level. Implicit in (5) is the assumption that the comparison graphs \mathcal{H}_k are all complete. We remark that in (5), $A_{ii}(k) = 0$ for each i , in contrast to the Spiked Gaussian Wigner model discussed in Section 1. Furthermore, we assume that $W(1), \dots, W(T)$ are independent.

Outliers model. In this model, we assume that the observation graphs are generated independently from Erdős-Rényi distributions, i.e., $\mathcal{H}_k \stackrel{\text{ind.}}{\sim} ER(n, p(k))$, where $p(k)$ represents the connection probability at time k . For each $k \in [T]$ and edge $\{i, j\} \in \mathcal{E}_k$, we observe the exact relative phase $g_i^*(k)g_j^*(k)^H$, with probability $1 - \eta$, and pure noise (chosen uniformly at random in $\text{SO}(2)$), with probability η . The matrix of observations $A(k)$ can then be written as

$$A_{ij}(k) = \begin{cases} g_i^*(k)g_j^*(k)^H & \text{with probability } (1 - \eta)p(k), \\ e^{i\varphi_{ij}(k)} & \text{with probability } \eta p(k), \\ 0 & \text{with probability } (1 - p(k)), \end{cases}$$

where $\phi_{ij}(k) \stackrel{\text{ind.}}{\sim} \text{Unif}[0, 2\pi)$. Since there are no observations when $i = j$, we choose to set $(A(k))_{ii} = 0$, for all $i \in [n]$. Note that $(A_{ij}(k))_{i \leq j}$ are independent, and $A(k)$ can be expressed as

$$A(k) = \underbrace{(1 - \eta) p(k)(G^*(k) - I_n)}_{=\mathbb{E}[A(k)]} + R(k), \quad (6)$$

where $R(k) := A(k) - \mathbb{E}[A(k)]$. It is easy to see that $(R_{ij}(k))_{i \leq j}$ are independent centered sub-Gaussian random variables.

Smoothness assumption. As discussed in the introduction, we will assume that the latent phases $g_i^*(k)$ do not change too quickly with time as outlined below.

Assumption 2 (Global ℓ_2 smoothness) For a given parameter $S_T > 0$,

$$\sum_{k=1}^{T-1} \|g^*(k) - g^*(k+1)\|_2^2 \leq S_T. \quad (7)$$

Denoting $M \in \{0, 1, -1\}^{T \times (T-1)}$ to be the incidence matrix of any directed path graph² on T vertices, this can be written equivalently as

$$\|(M^\top \otimes I_n)g^*\|_2^2 = g^{*\text{H}}((MM^\top) \otimes I_n)g^* \leq S_T,$$

where MM^\top corresponds to the Laplacian matrix of the undirected path graph.

The above assumption states that $g^*(k)$ does not change too quickly with k on “average”, and resembles the notion of quadratic variation of a function. It can also be viewed as stating that g^* lies “close” (depending on S_T) to the subspace spanned by the “smallest few” eigenvectors of $(MM^\top) \otimes I_n$. The closed-form expressions of the eigenvectors (and also eigenvalues) of MM^\top are well-known (e.g., (Brouwer and Haemers, 2012)), and can be viewed as the Fourier basis for signals residing on that graph (with the smallest eigenvector corresponding to the lowest frequency signal, and so on). Since Assumption 1 is in place, we do not need to worry about the invariance of (7) with respect to constant shifts.

2.3 Smoothness constrained estimators

We now describe algorithms to estimate g^* , each of which incorporates Assumption 2 in different ways. We first recall that in the static setting (for some fixed time $k \in [T]$), the SO(2)-synchronization problem can be formulated as solving

$$\max_{g(k) \in \mathcal{C}_n} g(k)^\text{H} A(k) g(k), \quad (8)$$

which, under the AGN model, corresponds to finding the MLE. As a sanity check, one could solve (8) for each k , and stack the solutions to produce an estimate of g^* . However, this suffers from the drawback of treating each time block independently. In light of Assumption 2, there is a temporal smoothness in the observations, which can be exploited to improve the estimation. Before proceeding, it will be useful to introduce some notation.

2. The rows of M correspond to the T vertices, and the columns correspond to the $T - 1$ directed edges. For the column corresponding to edge $i \rightarrow j$, we set its i th entry to $+1$, the j th entry to -1 , and the rest of the entries in that column to 0.

- We will use the diacritical mark \sim to denote the removal of information relative to the first item of each time block. Thus $\tilde{g}^*(k)$ and $\tilde{g}(k)$, are defined as

$$g^*(k) = \text{concat}(1, \tilde{g}^*(k)) = \begin{bmatrix} 1 \\ \tilde{g}^*(k) \end{bmatrix}, \quad \text{where } \tilde{g}^*(k) \in \mathcal{C}_{n-1},$$

$$g(k) = \text{concat}(1, \tilde{g}(k)) = \begin{bmatrix} 1 \\ \tilde{g}(k) \end{bmatrix}, \quad \text{where } \tilde{g}(k) \in \mathbb{C}^{n-1}.$$

Similarly, $\tilde{A}(k) \in \mathbb{C}^{(n-1) \times (n-1)}$ and $b(k) \in \mathbb{C}^{n-1}$ are defined as

$$A(k) = \begin{bmatrix} A_{11}(k) & b^H(k) \\ b(k) & \tilde{A}(k) \end{bmatrix}.$$

- Let us define the set

$$\mathcal{G} := \left\{ \begin{bmatrix} 1 \\ \tilde{g} \end{bmatrix} \begin{bmatrix} 1 & \tilde{g}^H \end{bmatrix} : \tilde{g} \in \mathcal{C}_{n-1} \right\}.$$

Note that $G^*(k) \in \mathcal{G}$ for each k .

- Whenever we do not specify the time index, it will mean that we stack all time blocks, as

$$\begin{aligned} A &= \text{concat}(A(1), \dots, A(T)) = \begin{bmatrix} A(1) \\ A(2) \\ \vdots \\ A(T) \end{bmatrix}, \quad G^* = \text{concat}(G^*(1), \dots, G^*(T)) = \begin{bmatrix} G^*(1) \\ G^*(2) \\ \vdots \\ G^*(T) \end{bmatrix}, \quad \text{and} \\ G &= \text{concat}(G(1), \dots, G(T)) = \begin{bmatrix} G(1) \\ G(2) \\ \vdots \\ G(T) \end{bmatrix} \in \mathbb{C}^{nT \times n}. \end{aligned} \tag{9}$$

- For any $z = (z_1, \dots, z_m) \in \mathbb{C}^m$ we define its projection onto \mathcal{C}_m by

$$\mathcal{P}_{\mathcal{C}_m}(z)_i := \begin{cases} \frac{z_i}{|z_i|} & \text{if } z_i \neq 0 \\ 1 & \text{otherwise.} \end{cases} \tag{10}$$

- Finally, for any $\tau \in [T]$, define

$$\mathcal{L}_{\tau,n} := \text{span} \{v_{T-k} \otimes e_j\}_{k=0, j=1}^{\tau-1, n},$$

where $\{v_k\}_{k=1}^T$ are the eigenvectors of MM^\top , the Laplacian of the path graph on T vertices, with corresponding eigenvalues $\lambda_1 \geq \lambda_2 \dots > \lambda_T (= 0)$. Denote $\mathcal{P}_{\tau,n}$ to be the projection matrix onto the space $\mathcal{L}_{\tau,n}$, i.e.,

$$\mathcal{P}_{\tau,n} = \left(\sum_{k=0}^{\tau-1} v_{T-k} v_{T-k}^\top \right) \otimes I_n. \tag{11}$$

The following simple observation shows that Assumption 2 can be equivalently stated (up to factors depending only on n) as a condition on the matrices $G^*(k)$. Note that, for all $k \in [T]$,

$$\|\tilde{g}^*(k) - \tilde{g}^*(k+1)\|_2^2 = \|g^*(k) - g^*(k+1)\|_2^2,$$

and

$$\|G^*(k) - G^*(k+1)\|_F^2 = 2\|\tilde{g}^*(k) - \tilde{g}^*(k+1)\|_2^2 + \|\tilde{g}^*(k)\tilde{g}^*(k)^\mathsf{H} - \tilde{g}^*(k+1)\tilde{g}^*(k+1)^\mathsf{H}\|_F^2,$$

hence

$$2\|g^*(k) - g^*(k+1)\|_2^2 \leq \|G^*(k) - G^*(k+1)\|_F^2 \leq (2+4n)\|g^*(k) - g^*(k+1)\|_2^2. \quad (12)$$

This implies, in particular, that the matrices $G^*(k)$ do not evolve too quickly, as given by

$$\sum_{k=1}^{T-1} \|G^*(k) - G^*(k+1)\|_F^2 \leq (2+4n)S_T. \quad (13)$$

2.3.1 GLOBAL SPHERE RELAXATION

Denoising optimization problem. Since (13) holds, under Assumption 2, it is natural to consider the following denoising optimization problem, which promotes smoothness via penalization.

$$\min_{\substack{G=\text{concat}(G(1), \dots, G(T)) \\ \text{s.t. } \forall k: G(k) \in \mathcal{G}}} \underbrace{\sum_{k=1}^T \|A(k) - G(k)\|_F^2}_{\text{data fidelity term}} + \lambda \underbrace{\sum_{k=1}^{T-1} \|G(k) - G(k+1)\|_F^2}_{\text{smoothness term}}, \quad (14)$$

where $\lambda \geq 0$ is the parameter that controls the amount of regularization in (14). On the other hand, for $G(k) \in \mathcal{G}$, we have

$$\|A(k) - G(k)\|_F^2 = \|A(k)\|_F^2 + \underbrace{\|G(k)\|_F^2}_{=n^2} - \tilde{g}(k)^\mathsf{H} \tilde{A}(k) \tilde{g} - 2\text{Re}(b^\mathsf{H}(k) \tilde{g}(k)), \quad (15)$$

which will help us write (14) more compactly (with variables $\tilde{g}(k)$). Specifically, using (15) and (12), the following problem is equivalent to (14)

$$\max_{\substack{\tilde{g}=\text{concat}(\tilde{g}(1), \dots, \tilde{g}(T)) \\ \text{s.t. } \forall k: \tilde{g}(k) \in \mathcal{C}_{n-1}}} \underbrace{\tilde{g}^\mathsf{H} \tilde{A}_{\text{block}} \tilde{g} + 2\text{Re}(b^\mathsf{H} \tilde{g})}_{\text{data fidelity term}} - \underbrace{\lambda \tilde{g}^\mathsf{H} (MM^\mathsf{T} \otimes I_{n-1}) \tilde{g}}_{\text{smoothness term}}, \quad (16)$$

where $\tilde{A}_{\text{block}} := \text{blkdiag}(\tilde{A}(1), \dots, \tilde{A}(T))$.

Remark 1 (Static anchored synchronization) *In the case $T = 1$, the smoothness penalty term will be absent from (14), and we recover the usual anchored synchronization MLE for the AGN model*

$$\min_{G \in \mathcal{G}} \|A - G\|_F^2 \equiv \max_{\tilde{g} \in \mathcal{C}_{n-1}} \tilde{g}^\mathsf{H} \tilde{A} \tilde{g} + 2\text{Re}(b^\mathsf{H} \tilde{g}). \quad (17)$$

Remark 2 (Complexity of (14)) *As mentioned in (Bandeira et al., 2017, Sec.2), the problem (17) is an instance of complex quadratic programming, which is known to be NP-Hard (Zhang and Huan, 2006, Prop.3.5). Without the penalty term, problem (14) amounts to solve T independent instances of (17), clearly remaining NP-Hard. However, it is unclear whether the hardness holds for all $\lambda > 0$. In particular, if one allows the value $\lambda = \infty$, the problem becomes trivial as any perfectly smooth G (that is $G(k) = G(1)$ for all $k \in [T]$) will be a solution. Nevertheless, the problem remains highly non-convex even for large values of λ , which makes the existence of an efficient algorithm for solving (14) (for an arbitrary A) unlikely.*

In light of Remark 2, we consider the following relaxation of (16)

$$\begin{aligned} \max_{\substack{\tilde{g} = \text{concat}(\tilde{g}(1), \dots, \tilde{g}(T)) \\ \text{s.t. } \|\tilde{g}\|_2^2 = (n-1)T}} \tilde{g}^H (\tilde{A}_{\text{block}} - \lambda(MM^\top \otimes I_{n-1})) \tilde{g} + 2\text{Re}(b^H \tilde{g}). \end{aligned} \quad (18)$$

Problem (18) is an instance of the trust region subproblem (TRS) with equality constraints, which consists of minimizing a general quadratic function (non-necessarily convex), under a sphere constraint. Although it will not be used in the sequel, it is worth mentioning that a version of TRS with ℓ_2 ball constraints exists (and it is, in some respect, more standard). There is a vast literature on the theory of TRS, including the characterization of its solutions (Hager, 2001), and several efficient algorithms that solve it, in both the spherical and ball constraints (Adachi et al., 2017). The complete procedure is outlined in Algorithm 1.

Algorithm 1 Global TRS for Dynamic Synchronization (GTRS-DynSync)

- 1: **Input:** Matrix $A \in \mathbb{C}^{nT \times n}$ of observations, smoothness parameter $\lambda > 0$.
 - 2: Form the matrix $\tilde{A}_{\text{block}} = \text{blkdiag}(\tilde{A}(1), \dots, \tilde{A}(T))$.
 - 3: Obtain $\hat{\tilde{g}} = \text{concat}(\hat{\tilde{g}}(1), \dots, \hat{\tilde{g}}(T)) \in \mathbb{C}^{(n-1)T} \leftarrow$ a solution of (18)
 - 4: **for** $k = 1 \dots T$ **do**
 - 5: Define $\hat{g}(k) = \begin{bmatrix} 1 \\ \mathcal{P}_{\mathcal{C}_{n-1}} \hat{\tilde{g}}(k) \end{bmatrix} \in \mathcal{C}_n$.
 - 6: **end for**
 - 7: **Output:** Estimator $\hat{g} \in \mathcal{C}_{nT}$.
-

2.3.2 LOCAL SPHERE RELAXATION

Instead of the “global relaxation” in (18), we now consider a different methodology which entails, as a first step, solving the following sphere relaxation of (17)

$$\hat{\tilde{g}}(k) := \arg \max_{\substack{\tilde{g}(k) \in \mathbb{C}^{n-1} \\ \|\tilde{g}(k)\|_2^2 = n-1}} \tilde{g}(k)^H \tilde{A}(k) \tilde{g}(k) + 2\text{Re}(b(k)^H \tilde{g}(k)). \quad (19)$$

This leads to a local estimate $\hat{\tilde{g}}(k)$ for each $k \in [T]$. Stacking them, we obtain

$$\hat{\tilde{g}} := \text{concat}(\hat{\tilde{g}}(1), \dots, \hat{\tilde{g}}(T)),$$

which satisfies the sphere constraint for each time block. The smoothness constraint has not been used yet, so as a second step, we consider an additional projection step onto the space of “low frequency” eigenvectors of $(MM^\top) \otimes I_{n-1}$. To achieve this, our proposed algorithm takes a regularization parameter $\tau \in [T]$ as input, and forms the projection matrix $\mathcal{P}_{\tau,n-1}$, which projects onto $\mathcal{L}_{\tau,n-1}$. It is worth noting that when τ is small, this projection step has a “smoothing” effect. For instance, for $\tau = 1$, the resulting projected-vector has equal block components. After the projection step onto $\mathcal{L}_{\tau,n-1}$, we finally need to use the projection $\mathcal{P}_{\mathcal{C}_{(n-1)T}}$ to ensure that the estimator lies in $\mathcal{C}_{(n-1)T}$. The complete procedure is outlined in Algorithm 2.

Algorithm 2 Local TRS + Global Smoothing for Dynamic Synchronization (LTRS-GS-DynSync)

- 1: **Input:** Matrix $A \in \mathbb{C}^{nT \times n}$ of observations, smoothness parameter $\tau \in \mathbb{N}$.
 - 2: **for** $k = 1, \dots, T$ **do**
 - 3: Obtain $\hat{g}(k) \leftarrow$ solution of (19).
 - 4: **end for**
 - 5: Form $\hat{g} = \text{concat}(\hat{g}(1), \dots, \hat{g}(T)) \in \mathbb{C}^{(n-1)T \times 1}$
 - 6: Define $\tilde{h} = \mathcal{P}_{\tau,n-1} \hat{g}$ where $\mathcal{P}_{\tau,n-1} \in \mathbb{C}^{(n-1)T \times (n-1)T}$ is the projection onto $\mathcal{L}_{\tau,n-1}$ as in (11).
 - 7: **for** $k = 1, \dots, T$ **do**
 - 8: Define $\hat{g}(k) = \begin{bmatrix} 1 \\ \mathcal{P}_{\mathcal{C}_{n-1}} \tilde{h}(k) \end{bmatrix}$, where $\tilde{h}(k)$ is such that $\tilde{h} = \text{concat}(\tilde{h}(1), \dots, \tilde{h}(T))$.
 - 9: **end for**
 - 10: **Output:** Estimator $\hat{g} \in \mathcal{C}_{nT}$.
-

2.3.3 DENOISING MEASUREMENT MATRICES

An alternative procedure involves denoising the measurement matrices A (recall (9)) by projecting its columns onto the low-frequency eigenspace of $(MM^\top) \otimes I_n$, in the spirit of the Laplacian eigenmaps estimator (see e.g., (Sadhanala et al., 2016)). Our proposed method consists of the following two steps.

- **Matrix denoising stage.** We project the columns of A onto the low frequency space $\mathcal{L}_{\tau,n}$, leading to the intermediate estimator $\hat{G} = \mathcal{P}_{\tau,n} A$.
- **Local synchronization stage.** For each $k \in [T]$ we first “Hermitianize” the block $\hat{G}(k)$ to obtain $\hat{G}'(k)$, and then solve a “local” synchronization problem, as in (17), but with input data $\hat{G}'(k)$. In Algorithm 3, we solve its TRS relaxation as in (19).

The matrix denoising stage can be regarded as a “global” denoising step, where the temporal smoothness captured by $(MM^\top) \otimes I_n$ is used to produce an estimation of $G^* \in \mathbb{C}^{nT \times nT}$. Notice that after this step, the obtained \hat{G} does not necessarily lie in \mathcal{G} . This constraint is enforced in the local anchored synchronization step, where we first solve a relaxation such as (19) and then project the solution onto \mathcal{C}_{nT} . The complete procedure is outlined in Algorithm 3.

Remark 3 (Spectral local synchronization) *It is possible to replace the TRS-based approach with a spectral method to solve the local synchronization problem in Algorithm 3. More specifically, lines 5 and 6 in Algorithm 3 are replaced by the following lines*

- 5: Find $\widehat{g}'(k) \leftarrow$ the eigenvector associated to the largest eigenvalue of $\widehat{G}'(k)$.
- 6: Define $\widehat{g}(k) = \mathcal{P}_{\mathcal{C}_n}(\widehat{g}'(k))e^{-i\widehat{\theta}_1(k)}$, where $\widehat{\theta}_1(k)$ is defined by $\widehat{g}'_1(k) = |\widehat{g}'_1(k)|e^{i\widehat{\theta}_1(k)}$ (with $\widehat{\theta}_1(k) = 0$ if $\widehat{g}'_1(k) = 0$).

While the spectral method may sacrifice some accuracy compared to the TRS-based approach, it offers simplicity in analysis. Our theoretical analysis in Section 3 is based on the spectral approach (outlined separately as Algorithm 4 for clarity), whereas in our experiments, we opt for the TRS-based method.

Remark 4 (Denoising smooth signals on a graph) *As alluded to earlier, the matrix denoising step in Algorithm 3 is motivated by the Laplacian eigenmaps estimator (see e.g., (Sadhanala et al., 2016)) for denoising smooth signals on a graph. To make the connection explicit using our notation, we recall that in this latter problem, we are given an undirected graph $\mathcal{H} = ([T], \mathcal{E})$, along with noisy measurements $y = x^* + \eta$ of a signal $x^* \in \mathbb{R}^T$ with η denoting the zero-mean noise. Denoting L to be the Laplacian matrix of \mathcal{H} , here x^* is assumed to be smooth w.r.t \mathcal{H} in the sense that*

$$(x^*)^\top L x^* = \sum_{\{i,j\} \in \mathcal{E}} (x_i^* - x_j^*)^2 \leq S_T. \quad (20)$$

The Laplacian eigenmaps estimator then estimates x^ by projecting y on to the space spanned by the τ smallest eigenvectors of L .*

In our setting, the aforementioned graph \mathcal{H} is a path graph³ and each vertex $k \in [T]$ has a latent signal $g^(k) \in \mathcal{C}_n$. We observe at each k a noisy matrix $A(k)$ which is also in the form of a signal (scalar times $G^*(k) - I_n$) plus zero-mean noise term; recall (5) and (6). Moreover, $g^*(k)$'s satisfy a smoothness condition (recall (7)) which, as discussed earlier, translates to a smoothness condition on $G^*(k)$'s (recall (13)) that is analogous to (20).*

As we will see in Section 3, the outline of the analysis is – at a top-level – similar to that in (Sadhanala et al., 2016) in the sense of bounding the bias and variance error terms, and finding the optimal τ . The difference lies in the technical details: we observe matrix-valued information (with complex-valued entries) at each node of \mathcal{H} which leads to more cumbersome calculations – especially for the Outliers model – in deriving high probability error bounds.

Remark 5 (Comparison with (Araya et al., 2023)) *As discussed in Section 1.3, the work of (Araya et al., 2023) considered an estimator for the dynamic translation synchronization problem, which also relied on the idea of projecting on to the low-frequency subspace. In that work, the estimator first solved the least-squares problem locally at each time-point, and then projected the stacked vector of individual-solutions onto the low-frequency space. This required each individual graph to be connected for the analysis to work. In Algorithm*

3. since we consider its vertices to denote time instants, although the analysis can be extended to general connected graphs as well.

3, however, we first project the (stacked) input data matrices onto the low-frequency space, and then solve the local synchronization problem. This allows the individual graphs to be disconnected in the Outliers model – something we will prove in Section 3 for a modified version of Algorithm 3, namely Algorithm 4.

Algorithm 3 Global Matrix Denoising + Local TRS for Dynamic Synchronization (GMD-LTRS-DynSync)

- 1: **Input:** Matrix $A \in \mathbb{C}^{nT \times n}$ of observations, parameter $\tau \in \mathbb{N}$.
 - 2: Define $\hat{G} = \mathcal{P}_{\tau,n}A$, where $\mathcal{P}_{\tau,n} \in \mathbb{C}^{nT \times nT}$ is the projection onto $\mathcal{L}_{\tau,n}$.
 - 3: **for** $k = 1 \dots, T$ **do**
 - 4: Hermitianize $\hat{G}'(k) := (\hat{G}(k) + \hat{G}(k)^H)/2$.
 - 5: Find $\hat{g}(k) \leftarrow$ solution of (19) with data $\hat{G}'(k)$.
 - 6: Form $\hat{g}(k) = \begin{bmatrix} 1 \\ \mathcal{P}_{\mathcal{C}_{n-1}} \hat{g}(k) \end{bmatrix} \in \mathcal{C}_n$.
 - 7: **end for**
 - 8: **Output:** Estimator $\hat{g} = \text{concat}(\hat{g}(1), \dots, \hat{g}(T)) \in \mathcal{C}_{nT}$.
-

3. Analysis

In this section, we present theoretical guarantees for recovering $g^*(1), \dots, g^*(T)$ under the AGN and Outliers models. Sections 3.1 and 3.2 contain results specifically addressing the *matrix denoising* stage of Algorithm 3 (as elaborated in Section 2.3.3). The main results therein are Theorem 8 for the AGN model, and Theorem 13 for the Outliers model. Then, Section 3.3 uses these estimates to provide guarantees for recovering $g^*(1), \dots, g^*(T)$ – it provides an analysis of the *local synchronization* stage by employing the spectral approach outlined in Remark 3. The main results therein are Corollary 17 for the AGN model, and Corollary 19 for the Outliers model.

3.1 Matrix denoising under AGN model

Given (5), we can write (c.f. the notation introduced in Section 2.3)

$$A = \underbrace{(G^* - \mathbb{1}_T \otimes I_n)}_{=: \bar{G}^*} + \sigma W,$$

where A, G^* are as in (9), and $W = \text{concat}(W(1), \dots, W(T))$ stacks the (Hermitian) noise matrices. We will focus on bounding the distance between the intermediate estimator $\hat{G} = \mathcal{P}_{\tau,n}A$ and the (shifted) ground truth \bar{G}^* . To this end, note that

$$\hat{G} = \mathcal{P}_{\tau,n}A = \mathcal{P}_{\tau,n}\bar{G}^* + \sigma \mathcal{P}_{\tau,n}W,$$

from which we deduce

$$\|\hat{G} - \bar{G}^*\|_F^2 = \|\mathcal{P}_{\tau,n}^\perp \bar{G}^*\|_F^2 + \sigma^2 \|\mathcal{P}_{\tau,n}W\|_F^2, \quad (21)$$

where $\mathcal{P}_{\tau,n}^\perp := I_{nT} - \mathcal{P}_{\tau,n}$. The first term on the right-hand side (RHS) of (21) represents the bias which depends on the smoothness level. Clearly \overline{G}^* has the same quadratic variation as G^* and hence satisfies (13) as well. Thus, we expect this term to decrease as τ increases. The second term therein corresponds to the variance introduced by noise, which we expect to increase with τ . Hence, selecting the optimal value for τ will achieve the right bias-variance trade-off. The following lemma gives a control over the bias term.

Lemma 6 (Bias bound) *Let $G^* \in \mathbb{C}^{nT \times n}$ be a matrix satisfying (13), for some $S_T \geq 0$. Then, it holds for all $1 \leq \tau \leq T$*

$$\|\mathcal{P}_{\tau,n}^\perp \overline{G}^*\|_F^2 \leq \frac{20n}{\pi^2} \frac{T^2 S_T}{\tau^2} \mathbb{1}_{\tau < T}. \quad (22)$$

Proof When $\tau = T$, we have $\mathcal{P}_{\tau,n}^\perp = 0$, from which the final bound follows. So assume now that $1 \leq \tau \leq T - 1$. Denote the j -th column of G^* by $G_{:,j}^*$. Since

$$\|\mathcal{P}_{\tau,n}^\perp G^*\|_F^2 = \sum_{j=1}^n \|\mathcal{P}_{\tau,n}^\perp G_{:,j}^*\|_2^2,$$

we have for each $j \in [n]$ that

$$\|\mathcal{P}_{\tau,n}^\perp \overline{G}_{:,j}^*\|_F^2 = \sum_{i=1}^n \sum_{k=\tau}^{T-1} \langle v_{T-k} \otimes e_i, \overline{G}_{:,j}^* \rangle^2.$$

On the other hand, it is easy to see that (13) can be rewritten as

$$\sum_{j=1}^n \sum_{i=1}^n \sum_{k=0}^{T-1} \lambda_{T-k} \langle v_{T-k} \otimes e_i, \overline{G}_{:,j}^* \rangle^2 \leq (2 + 4n) S_T,$$

from which we deduce (since $n \geq 2$)

$$\lambda_{T-\tau} \sum_{j=1}^n \sum_{i=1}^n \sum_{k=\tau}^{T-1} \langle v_{T-k} \otimes e_i, \overline{G}_{:,j}^* \rangle^2 \leq 5n S_T.$$

This implies (notice that $\lambda_{T-\tau} \neq 0$ since $1 \leq \tau \leq T - 1$)

$$\sum_{j=1}^n \|\mathcal{P}_{\tau,n}^\perp \overline{G}_{:,j}^*\|_2^2 \leq \frac{5n S_T}{\lambda_{T-\tau}}.$$

By the known expression for the eigenvalues of the Laplacian of the path graph (Brouwer and Haemers, 2012), we have $\lambda_{T-\tau} = 4 \sin^2(\frac{\tau\pi}{2T})$, and using that $\sin(x) \geq \frac{x}{2}$, for $x \in [0, \pi/2]$, we obtain

$$\sum_{j=1}^n \|\mathcal{P}_{\tau,n}^\perp \overline{G}_{:,j}^*\|_2^2 \leq \frac{20n}{\pi^2} \frac{T^2 S_T}{\tau^2}.$$

■

We now control the variance term.

Lemma 7 (Variance bound) *Let $W(1), \dots, W(T)$ be independent, random Hermitian matrices as in the AGN model in Section 2.2. Then there exists a constant $C > 0$, such that for all $\tau \in [T]$ and $\delta \in (0, 1)$, it holds with probability larger than $1 - \delta$ that*

$$\|\mathcal{P}_{\tau,n}W\|_F^2 \leq C \left(n^2\tau + \log \left(\frac{4}{\delta} \right) \right).$$

Proof To simplify the analysis, we split W as follows

$$W = \frac{1}{\sqrt{2}}(W_1 + W_2),$$

where $W_1, W_2 \in \mathbb{C}^{nT \times n}$ satisfy $W_1(k) = W_2(k)^H$, for all $k \in [T]$, and the entries of $W_1(k)$ are i.i.d random variables from $\mathcal{CN}(0, 1)$, for all $k \in [T]$. Furthermore, we consider the decomposition

$$W_s = \underbrace{\text{Re}(W_s)}_{=:W_{s,R}} + i \underbrace{\text{Im}(W_s)}_{=:W_{s,I}},$$

for $s = 1, 2$. It will be enough to bound $\|\mathcal{P}_{\tau,n}W_1\|_F$, since the same bound will hold for $\|\mathcal{P}_{\tau,n}W_2\|_F$, analogously. Denoting $(W_{1,R})_{:,j}$ (resp. $(W_{1,I})_{:,i}$) to be the j -th column of $W_{1,R}$ (resp. $W_{1,I}$), we obtain

$$\begin{aligned} \|\mathcal{P}_{\tau,n}W_1\|_F^2 &= \|\mathcal{P}_{\tau,n}W_{1,R}\|_F^2 + \|\mathcal{P}_{\tau,n}W_{1,I}\|_F^2 \\ &= \sum_{j=1}^n \|\mathcal{P}_{\tau,n}(W_{1,R})_{:,j}\|_2^2 + \sum_{j=1}^n \|\mathcal{P}_{\tau,n}(W_{1,I})_{:,j}\|_2^2 \\ &= \underbrace{W_1'^H (I_{2n} \otimes \mathcal{P}_{\tau,n}) W_1'}_{=: \mathcal{V}_1}, \end{aligned}$$

where $W_1' \in \mathbb{R}^{2n^2T \times 1}$ is defined as

$$W_1' := \begin{bmatrix} (W_{1,R})_{:,1} \\ \vdots \\ (W_{1,R})_{:,n} \\ (W_{1,I})_{:,1} \\ \vdots \\ (W_{1,I})_{:,n} \end{bmatrix}.$$

Observe that W_1' has independent entries, with each entry either 0 or drawn from $\mathcal{N}(0, 1/2)$. To bound \mathcal{V}_1 , we will use the Hanson-Wright inequality (Rudelson and Vershynin, 2013, Thm. 1.1). It is easy to see that

$$\|I_{2n} \otimes \mathcal{P}_{\tau,n}\|_{op} = 1, \quad \|I_{2n} \otimes \mathcal{P}_{\tau,n}\|_F^2 = 2n^2\tau.$$

Thus, from Hanson-Wright inequality, we obtain, for all $t \geq 0$

$$\mathbb{P}(|\mathcal{V}_1 - \mathbb{E}[\mathcal{V}_1]| > t) \leq 2 \exp \left(-c_1 \left(\frac{t^2}{n^2\tau} \wedge t \right) \right).$$

On the other hand, notice that $\mathbb{E}[\mathcal{V}_1] \leq c_2 n^2 \tau$. We deduce that, for any $\delta > 0$, with probability larger than $1 - \delta$,

$$\mathcal{V}_1 \leq C \left(n^2 \tau + \sqrt{n^2 \tau \log \left(\frac{2}{\delta} \right)} + \log \left(\frac{2}{\delta} \right) \right) \leq 2C \left(n^2 \tau + \log \left(\frac{2}{\delta} \right) \right)$$

for a constant $C > 0$. To pass from the first to the second inequality, we used the fact $a^2 + ab + b^2 \leq 2(a^2 + b^2)$ for any $a, b \in \mathbb{R}$. Arguing in an analogous way, we obtain that

$$\|\mathcal{P}_{\tau,n} W_2\|_F^2 \leq 2C \left(n^2 \tau + \log \left(\frac{2}{\delta} \right) \right),$$

with probability greater than $1 - \delta$. Since $\|\mathcal{P}_{\tau,n} W\|_F^2 \leq \|\mathcal{P}_{\tau,n} W_1\|_F^2 + \|\mathcal{P}_{\tau,n} W_2\|_F^2$, the result follows via the union bound. \blacksquare

Optimal choice of τ . Combining (21) with Lemmas 6 and 7 we obtain the following bound (with probability larger than $1 - \delta$)

$$\|\widehat{G} - \overline{G}^*\|_F^2 \leq \frac{20n T^2 S_T}{\pi^2 \tau^2} \mathbb{1}_{\tau < T} + C \sigma^2 \left(n^2 \tau + \log \left(\frac{4}{\delta} \right) \right). \quad (23)$$

It remains to choose the optimal value of $\tau \in [T]$ which minimizes the RHS of (23), up to constant factors. To this end, using $\mathbb{1}_{\tau < T} \leq 1$, note that the global minimizer of the ensuing bound is given by $(\frac{T^2 S_T}{n \sigma^2})^{1/3}$. But this quantity does not necessarily lie in $[T]$. Hence we define

$$\tau^* := \min \left\{ 1 + \left\lfloor \left(\frac{T^2 S_T}{n \sigma^2} \right)^{1/3} \right\rfloor, T \right\} \in [T]. \quad (24)$$

Clearly τ^* is not the global minimizer of the bound in (23) but it will lead to an error bound which achieves the right scaling in terms of T and S_T . This is shown in the following theorem which is the first main result of this section. Its proof is deferred to Appendix A.

Theorem 8 (Main result: \overline{G}^* recovery under AGN model) *Let A be the stacked measurement matrix for the AGN model in (5), with a ground truth g^* satisfying Assumption 2. Then choosing $\tau = \tau^*$ with τ^* as in (24), the following holds for the estimator $\widehat{G} = \mathcal{P}_{\tau,n} A$. There exists a constant $C > 0$ such that for any $\delta \in (0, 1)$, it holds with probability at least $1 - \delta$ that*

$$\begin{aligned} \|\widehat{G} - \overline{G}^*\|_F^2 \leq C \Big(\max \left\{ \frac{n T^2 S_T}{\left(1 + \left(\frac{T^2 S_T}{n \sigma^2} \right)^{2/3} \right)}, n S_T \right\} \mathbb{1}_{\{\tau^* < T\}} + \sigma^2 n^2 \min \left\{ 1 + \left(\frac{T^2 S_T}{n \sigma^2} \right)^{1/3}, T \right\} \\ + \sigma^2 \log \left(\frac{4}{\delta} \right) \Big). \end{aligned} \quad (25)$$

Furthermore, there exist constants $C_1, C_2 > 0$ such that if

$$T \geq C_1 \max \left\{ \left(\frac{n \sigma^2}{S_T} \right)^{1/2}, \left(\frac{S_T}{n \sigma^2} \right), \sigma^4 S_T n^5 \right\} \quad (26)$$

then (25) implies

$$\|\widehat{G} - \overline{G}^*\|_F^2 \leq C_2 \left(\sigma^{4/3} T^{2/3} S_T^{1/3} n^{5/3} + \sigma^2 \log \left(\frac{4}{\delta} \right) \right). \quad (27)$$

The following remarks are in order for interpreting Theorem 8.

Remark 9 When $\sigma = 0$ then we obtain $\tau^* = T$ and $\|\widehat{G} - \overline{G}^*\|_F = 0$. This is also seen directly from (23). If $S_T = 0$, we obtain $\tau^* = 1$ and $\|\widehat{G} - \overline{G}^*\|_F^2 \lesssim \sigma^2 n^2 + \sigma^2 \log(4/\delta)$. This can also be obtained directly from (23). Note that $\frac{1}{\sqrt{T}} \|\widehat{G} - \overline{G}^*\|_F = O(1/\sqrt{T})$ which is the parametric rate.

Remark 10 The first part of Theorem 8 holds for all $\sigma, S_T \geq 0$. The interesting situation, however, is when $\sigma, S_T > 0$ – especially when both these quantities are sufficiently bounded away from zero. In this case, the second part of Theorem 8 is meaningful. For instance, suppose $\sigma \asymp T^\alpha$ for some $\alpha \in [0, 1/4)$. Then provided

$$S_T \gtrsim \frac{n}{T^{2(1-\alpha)}} \quad \text{and} \quad S_T = o(T^{1-4\alpha}) \quad (28)$$

it is easy to see that (26) will be satisfied for large enough T . Moreover, (27) then implies

$$\frac{1}{T} \|\widehat{G} - \overline{G}^*\|_F^2 \xrightarrow{T \rightarrow \infty} 0.$$

Finally, we remark that the $O(T^{2/3} S_T^{1/3})$ term in (27) matches that obtained in (Araya et al., 2023, Theorem B) and also matches the rate for denoising smooth signals on a path graph (Sadhanala et al., 2016, Theorem 6).

3.2 Matrix denoising under the Outliers model

We now focus on the outliers model, defined in (6), which can be written in stacked matrix notation as

$$A = D(G^* - I_n) + R = D\overline{G}^* + R, \quad (29)$$

where $D = (1 - \eta) \text{blkdiag}(p(1)I_n, \dots, p(T)I_n)$ and $R = \text{concat}(R(1), \dots, R(T))$. Note that, for each $k \in [T]$, $R(k)$ is a Hermitian matrix with $(R_{ij}(k))_{i \leq j}$ being independent centered sub-Gaussian random variables, and in particular, $R_{ii}(k) = 0$ for each $i \in [n]$.

Our goal is to bound the error $\|\widehat{G} - D\overline{G}^*\|_F^2$. Since $\widehat{G} = \mathcal{P}_{\tau,n} A = \mathcal{P}_{\tau,n}(D\overline{G}^*) + \mathcal{P}_{\tau,n} R$, we obtain the bound

$$\|\widehat{G} - D\overline{G}^*\|_F^2 \leq 2 \underbrace{\|\mathcal{P}_{\tau,n}^\perp(D\overline{G}^*)\|_F^2}_{\text{Bias}} + 2 \underbrace{\|\mathcal{P}_{\tau,n}(R)\|_F^2}_{\text{Variance}}, \quad (30)$$

where we recall $\mathcal{P}_{\tau,n}^\perp = I_{nT} - \mathcal{P}_{\tau,n}$. The error bound decomposition in (30) is similar to the bias-variance decomposition in (21). In order to control the bias term $\|\mathcal{P}_{\tau,n}^\perp(D\overline{G}^*)\|_F^2$, we need to place an assumption on the level of fluctuation of $p(1), \dots, p(T)$. Indeed, consider the ideal scenario where $p(k) = p$ for all $k \in [T]$. Then, $\|\mathcal{P}_{\tau,n}^\perp(D\overline{G}^*)\|_F^2 = (1 - \eta)^2 p^2 \|\mathcal{P}_{\tau,n}^\perp \overline{G}^*\|_F^2$ which can then be bounded using Lemma 6. This motivates the following assumption on the smoothness of the sequence $(p(k))_{k=1}^T$, or equivalently, on the diagonal entries of matrix D .

Assumption 3 *The matrix D is μ -smooth, i.e., for some $\mu \geq 1$ it holds for all $\tau \in [T]$ that*

$$\|\mathcal{P}_{\tau,n}(DG^*)\|_F^2 \leq \mu \bar{d}^2 \|\mathcal{P}_{\tau,n}G^*\|_F^2,$$

where $\bar{d} := \frac{\sum_{k=1}^T d(k)}{T}$, and $d(k) := (1 - \eta)p(k)$.

If $p(k) = p$ for all k , then note that Assumption 3 holds with $\mu = 1$. Now, if D satisfies Assumption 3, we can readily bound the bias term using (22) as

$$\|\mathcal{P}_{\tau,n}(DG^*)\|_F^2 \leq C_1 \frac{\mu \bar{d}^2 n T^2 S_T}{\tau^2} \mathbb{1}_{\tau < T} \quad (31)$$

for some constant $C_1 > 0$. For the variance term we have the following analog to Lemma 7. The proof mirrors that of Lemma 7, with a technical difference arising due to the different distribution of the entries of R . Its proof is deferred to Appendix B.

Lemma 11 (Variance bound) *Let $R(1), \dots, R(T)$ be independent random Hermitian matrices arising from the Outliers model in Section 2.2. Denote $Q(p)$ to be the ψ_2 norm of a centered Bernoulli random variable with parameter p , and define*

$$\begin{aligned} p_{\max} &:= \max_{k \in [T]} p(k), \quad Q := \eta + Q(\eta) + \max_{k \in [T]} Q(p(k)) \\ V &:= \max_{k \in [T]} \{p(k)(1 - p(k))\}, \quad f(\eta, p_{\max}, V) := p_{\max}\eta + (1 - \eta)^2 V + p_{\max}V(1 - \eta). \end{aligned}$$

Then, there exist constants $C_1 \in (0, 1)$ and $C_2 > 0$ such that for any $\delta \in (0, C_1)$, it holds with probability at least $1 - \delta$ that

$$\|\mathcal{P}_{\tau,n}R\|_F^2 \leq C_2 \left(n^2 \tau f(\eta, p_{\max}, V) + Q^2 n \sqrt{\tau} \log \left(\frac{1}{\delta} \right) \right).$$

Using (31) and Lemma 11, we arrive at the following theorem which is the second main result of this section. The proof uses similar calculations as for that of Theorem 8 and is deferred to Appendix C.

Remark 12 *The exact expression for $Q(p)$ is known to be (Ostrovsky and Sirota, 2014)*

$$Q(p) = \sqrt{\frac{1 - 2p}{4 \log(\frac{1-p}{p})}}.$$

Moreover, $Q(\cdot)$ is non-negative continuous on $[0, 1]$, with $Q(0+0) = Q(1-0) = 0$, and $Q^2(1/2) = 1/8$ (see (Ostrovsky and Sirota, 2014)). It is easy to also verify that $Q(p) \leq \frac{1}{\sqrt{2 \log(1/p)}}$ if $p \leq 1/4$.

Theorem 13 (Main result: \bar{G}^* recovery under Outliers model) *Let A be the stacked measurement matrix under the outliers model (6), with a ground truth g^* satisfying Assumption 2, and the matrix D (c.f. (29)) satisfies Assumption 3 with parameter $\mu \geq 1$. Then there exist constants $C_1 \in (0, 1)$ and $C_2, C_3, C_4, C_5 > 0$ such that for any $\delta \in (0, C_1)$, the following is true (recall the notation of Lemma 11).*

1. Denote $\tilde{f}(\eta, p_{\max}, V, Q, \delta) := f(\eta, p_{\max}, V) + Q^2 \log(1/\delta)$, and choose

$$\tau = \tau^* := \min \left\{ 1 + \left\lfloor \left(\frac{\mu \bar{d}^2 T^2 S_T}{n \tilde{f}(\eta, p_{\max}, V, Q, \delta)} \right)^{1/3} \right\rfloor, T \right\}. \quad (32)$$

Then with probability at least $1 - \delta$,

$$\begin{aligned} \|\hat{G} - D\bar{G}^*\|_F^2 &\leq C_2 \max \left\{ \frac{\mu \bar{d}^2 n T^2 S_T}{1 + \left(\frac{\mu \bar{d}^2 T^2 S_T}{n \tilde{f}(\eta, p_{\max}, V, Q, \delta)} \right)^{2/3}}, \mu \bar{d}^2 n S_T \right\} \mathbb{1}_{\tau^* < T} \\ &\quad + C_3 \tilde{f}(\eta, p_{\max}, V, Q, \delta) n^2 \min \left\{ 1 + \left(\frac{\mu \bar{d}^2 T^2 S_T}{n \tilde{f}(\eta, p_{\max}, V, Q, \delta)} \right)^{1/3}, T \right\}. \end{aligned} \quad (33)$$

2. Furthermore, if T satisfies

$$T \geq C_4 \max \left\{ \left(\frac{n \tilde{f}(\eta, p_{\max}, V, Q, \delta)}{\mu \bar{d}^2 S_T} \right)^{1/2}, \frac{\mu \bar{d}^2 S_T}{n \tilde{f}(\eta, p_{\max}, V, Q, \delta)} \right\} \quad (34)$$

then (33) implies

$$\|\hat{G} - D\bar{G}^*\|_F^2 \leq C_5 \tilde{f}^{2/3}(\eta, p_{\max}, V, Q, \delta) (\mu \bar{d}^2)^{1/3} n^{5/3} T^{2/3} S_T^{1/3}. \quad (35)$$

Remark 14 If $S_T = 0$, then $\tau^* = 1$ and (33) simplifies to $\|\hat{G} - D\bar{G}^*\|_F^2 \lesssim \tilde{f}(\eta, p_{\max}, V, Q, \delta) n^2$. Hence $\frac{1}{\sqrt{T}} \|\hat{G} - D\bar{G}^*\|_F = O(1/\sqrt{T})$ which is the parametric rate. On the other hand, if $\eta = 0$ and $p(k) = 1$ for all k , then note that $Q = V = \tilde{f}(\eta, p_{\max}, V, Q, \delta) = f(\eta, p_{\max}, V) = 0$. This implies $\tau^* = T$ and $\|\hat{G} - D\bar{G}^*\|_F = 0$.

Remark 15 Suppose η is constant and let $p_{\max} \asymp 1/n$. This means that a given measurement graph \mathcal{H}_k is disconnected with high probability. Moreover, suppose $\bar{d} \asymp 1/n$, and $\mu \asymp 1$ (e.g., this will happen if $p(k) = p \asymp 1/n$ for all k). Then we have $Q \asymp 1$, $V \asymp 1/n$ which implies $f(\eta, p_{\max}, V) \asymp 1/n$ and $\tilde{f}(\eta, p_{\max}, V, Q, \delta) \asymp \log(1/\delta)$ (assuming δ is a constant for simplicity). Consequently, the condition in (34) simplifies to

$$T \gtrsim \max \left\{ \left(\frac{n^3 \log(1/\delta)}{S_T} \right)^{1/2}, \left(\frac{S_T}{n^3 \log(1/\delta)} \right) \right\} \quad (36)$$

and (35) translates to $\|\hat{G} - D\bar{G}^*\|_F^2 \lesssim n T^{2/3} S_T^{1/3} \log^{2/3}(1/\delta)$. Thus for the smoothness regime $S_T = \omega(1/T^2)$ and $S_T = o(T)$, we see that (36) will be satisfied for T suitably large w.r.t n .

3.3 Spectral local synchronization guarantee

We now proceed to establish guarantees for the local synchronization phase within Algorithm 3. While Algorithm 3 employs a TRS-based methodology at this stage, it is easier to analyze the spectral method, as delineated in Remark 3. This is outlined as a separate

Algorithm 4 Global Matrix Denoising + Local Spectral method for Dynamic Synchronization (GMD-LSPEC-DynSync)

- 1: **Input:** Matrix $A \in \mathbb{C}^{nT \times n}$ of observations, parameter $\tau \in \mathbb{N}$.
 - 2: Define $\hat{G} = \mathcal{P}_{\tau,n} A$, where $\mathcal{P}_{\tau,n} \in \mathbb{C}^{nT \times nT}$ is the projection onto $\mathcal{L}_{\tau,n}$.
 - 3: **for** $k = 1 \dots, T$ **do**
 - 4: Hermitianize $\hat{G}'(k) := (\hat{G}(k) + \hat{G}(k)^H)/2$.
 - 5: Find $\hat{g}'(k) \leftarrow$ the eigenvector associated to the largest eigenvalue of $\hat{G}'(k)$.
 - 6: Find $\hat{g}(k) = \mathcal{P}_{\mathcal{C}_n}(\hat{g}'(k))e^{-i\hat{\theta}_1(k)}$, where $\hat{\theta}_1(k)$ is defined by $\hat{g}'_1(k) = |\hat{g}'_1(k)|e^{i\hat{\theta}_1(k)}$ (with $\hat{\theta}_1(k) = 0$ if $\hat{g}'_1(k) = 0$).
 - 7: **end for**
 - 8: **Output:** Estimator $\hat{g} = \text{concat}(\hat{g}(1), \dots, \hat{g}(T)) \in \mathcal{C}_{nT}$.
-

method in Algorithm 4 for clarity. Our main results for recovering g^* are Corollaries 17 and 19 for the AGN and Outliers models respectively.

Our strategy involves utilizing the Davis-Kahan inequality “locally”, focusing on specific time instances where the error converges to zero. To begin with, denote for each k a scaling $s(k) \geq 0$ where $s(k) = 1$ for the AGN model, and $s(k) = d(k) = (1 - \eta)p(k)$ for the outliers model (recall Assumption 3). Then note that since $\bar{G}^*(k)$ is symmetric, we have for the Hermitian matrix $\hat{G}'(k)$ defined in line 4 of Algorithm 4 that

$$\|\hat{G}'(k) - s(k)\bar{G}^*(k)\|_F \leq \|\hat{G}(k) - s(k)\bar{G}^*(k)\|_F. \quad (37)$$

The next step is to use the Davis-Kahan theorem (Davis and Kahan, 1970) to relate the largest eigenvector of $\hat{G}'(k)$ to $g^*(k)/\sqrt{n}$, for each k . Since $s(k)\bar{G}^*(k) = s(k)(g^*(k)g^*(k)^H - I_n)$, the spectral gap between the largest and second largest eigenvalues of $s(k)\bar{G}^*(k)$ is $s(k)n$. Recall $\hat{g}'(k)$ as defined in line 5 of Algorithm 4. Then, by the Davis-Kahan theorem (Davis and Kahan, 1970), we have for each k that

$$\left\| \left(I - \frac{g^*(k)g^*(k)^H}{n} \right) \hat{g}'(k) \right\|_2 \leq \frac{4\|\hat{G}'(k) - s(k)\bar{G}^*(k)\|_{op}}{s(k)n}. \quad (38)$$

A standard calculation reveals (see e.g., (d’Aspremont et al., 2021, Appendix D.1)) that there exists $\phi(k) \in [0, 2\pi)$ (depending on $g^*(k), \hat{g}'(k)$) such that

$$\left\| \hat{g}'(k) - e^{i\phi(k)} \frac{g^*(k)}{\sqrt{n}} \right\|_2 \leq 2 \left\| \left(I - \frac{g^*(k)(g^*(k))^H}{n} \right) \hat{g}'(k) \right\|_2$$

which together with (38) implies

$$\left\| \hat{g}'(k) - e^{i\phi(k)} \frac{g^*(k)}{\sqrt{n}} \right\|_2 \leq 8 \frac{\|\hat{G}'(k) - s(k)\bar{G}^*(k)\|_{op}}{s(k)n} \leq 8 \frac{\|\hat{G}'(k) - s(k)\bar{G}^*(k)\|_F}{s(k)n}. \quad (39)$$

Using (39), (37), we arrive at the following result.

Proposition 16 *Let g^* be a ground truth vector satisfying Assumption 1 as defined in (4), and let $\widehat{g} = \text{concat}(\widehat{g}(1), \dots, \widehat{g}(T)) \in \mathcal{C}_{nT}$ be obtained from Algorithm 4. Then there exists a constant $C > 0$ such that*

$$\|\widehat{g}(k) - g^*(k)\|_2 \leq C \frac{\|\widehat{G}(k) - s(k)\overline{G}^*(k)\|_F}{s(k)}.$$

Denoting $s_{\min} := \min_k s(k)$, it then follows that

$$\|\widehat{g} - g^*\|_2^2 = \sum_{k=1}^T \|\widehat{g}(k) - g^*(k)\|_2^2 \leq C^2 \frac{1}{s_{\min}^2} \sum_{k=1}^T \|\widehat{G}(k) - s(k)\overline{G}^*(k)\|_F^2.$$

The proof is deferred to Appendix D. Combining Proposition 16 with Theorems 8 and 13 directly leads to error guarantees for Algorithm 3, with spectral local synchronization, for the AGN and Outliers models. Recall that for the AGN model, $s_{\min} = 1$, while in the outliers model, $s_{\min} = (1 - \eta) \min_k p(k)$.

Corollary 17 (Main result: g^* recovery under AGN model) *Let A be the stacked measurement matrix for the AGN model in (5), with a ground truth g^* satisfying Assumption 2. Choosing $\tau = \tau^*$ with τ^* as in (24), consider the estimator $\widehat{g} = \text{concat}(\widehat{g}(1), \dots, \widehat{g}(T))$ obtained in Algorithm 4. There exist constants $C_1, C_2, C_3, C_4 > 0$ such that for any $\delta \in (0, 1)$ the following is true.*

$$\begin{aligned} \|\widehat{g} - g^*\|_2^2 \leq C \Big(\max \left\{ \frac{nT^2 S_T}{\left(1 + \left(\frac{T^2 S_T}{n\sigma^2}\right)^{2/3}\right)}, nS_T \right\} \mathbb{1}_{\{\tau^* < T\}} + \sigma^2 n^2 \min \left\{ 1 + \left(\frac{T^2 S_T}{n\sigma^2}\right)^{1/3}, T \right\} \\ + \sigma^2 \log \left(\frac{4}{\delta}\right) \Big). \end{aligned} \quad (40)$$

Furthermore, there exist constants $C_1, C_2 > 0$ such that if

$$T \geq C_1 \max \left\{ \left(\frac{n\sigma^2}{S_T}\right)^{1/2}, \left(\frac{S_T}{n\sigma^2}\right), \sigma^4 S_T n^5 \right\}$$

then (40) implies

$$\|\widehat{g} - g^*\|_2^2 \leq C_2 \left(\sigma^{4/3} T^{2/3} S_T^{1/3} n^{5/3} + \sigma^2 \log \left(\frac{4}{\delta}\right) \right).$$

Remark 18 *The following points are useful to note regarding Corollary 17.*

1. *Following Remark 9, when $\sigma = 0$ then we obtain $\tau^* = T$ and $\|\widehat{g} - g^*\|_2 = 0$. If $S_T = 0$, we obtain $\tau^* = 1$ and*

$$\|\widehat{g} - g^*\|_2^2 \lesssim \left(\sigma^2 n^2 + \sigma^2 \log(4/\delta) \right).$$

2. *Following Remark 10, suppose $\sigma \asymp T^\alpha$ for some $\alpha \in [0, 1/4)$. Then provided S_T satisfies (28), part 2 of Corollary 17 implies that*

$$\frac{1}{T} \|\widehat{g} - g^*\|_2^2 \xrightarrow{T \rightarrow \infty} 0.$$

In particular, when $\sigma > 0$ is a constant, then $\|\hat{g} - g^*\|_2^2 = O(T^{2/3} S_T^{1/3})$ which matches the (optimal) rate for denoising smooth signals on a path graph on T vertices (Sadhanala et al., 2016, Theorems 5,6); recall the discussion in Remark 4.

Corollary 19 (Main result: g^* recovery under Outliers model) *Let A be the stacked measurement matrix under the outliers model (6), with a ground truth g^* satisfying Assumption 2, and the matrix D (c.f. (29)) satisfies Assumption 3 with parameter $\mu \geq 1$. For $\tau = \tau^*$ as in (32), consider the estimator $\hat{g} = \text{concat}(\hat{g}(1), \dots, \hat{g}(T))$ obtained in Algorithm 4 and recall the notation of Theorem 13. There exist constants $C_1 \in (0, 1)$ and $C_2, C_3, C_4 > 0$ such that for any $\delta \in (0, C_1)$, the following is true.*

1. Denoting $p_{\min} := \min_k p(k)$ it holds with probability at least $1 - \delta$,

$$\begin{aligned} \|\hat{g} - g^*\|_2^2 \leq & \frac{C_2}{(1 - \eta)^2 p_{\min}^2} \left(\max \left\{ \frac{\mu \bar{d}^2 n T^2 S_T}{1 + \left(\frac{\mu \bar{d}^2 T^2 S_T}{n \tilde{f}(\eta, p_{\max}, V, Q, \delta)} \right)^{2/3}}, \mu \bar{d}^2 n S_T \right\} \mathbb{1}_{\tau^* < T} \right. \\ & \left. + \tilde{f}(\eta, p_{\max}, V, Q, \delta) n^2 \min \left\{ 1 + \left(\frac{\mu \bar{d}^2 T^2 S_T}{n \tilde{f}(\eta, p_{\max}, V, Q, \delta)} \right)^{1/3}, T \right\} \right). \end{aligned} \quad (41)$$

2. Furthermore, if T satisfies (34) (with a constant $C_3 > 0$), then (41) implies that

$$\|\hat{g} - g^*\|_2^2 \leq C_4 \left(\frac{\tilde{f}^{2/3}(\eta, p_{\max}, V, Q, \delta)}{(1 - \eta)^2 p_{\min}^2} \right) (\mu \bar{d}^2)^{1/3} n^{5/3} T^{2/3} S_T^{1/3}.$$

Remark 20 *Let us note the following points regarding Corollary 19.*

1. Following Remark 14, if $S_T = 0$, then $\tau^* = 1$ and we obtain

$$\|\hat{g} - g^*\|_2^2 \lesssim \frac{\tilde{f}(\eta, p_{\max}, V, Q, \delta)}{(1 - \eta)^2 p_{\min}^2} n^2.$$

On the other hand, if $\eta = 0$ and $p(k) = 1$ for all k , then $\tau^* = T$ and $\|\hat{g} - g^*\|_2 = 0$.

2. Consider the setup of Remark 15 so that $p_{\min} \asymp 1/n$. Then if T satisfies (36) (which is the case for T large enough when $S_T = \omega(1/T^2)$ and $S_T = o(T)$ holds), we obtain the bound

$$\|\hat{g} - g^*\|_2^2 \lesssim n^3 T^{2/3} S_T^{1/3} \log^{2/3}(1/\delta).$$

The bound exhibits the same scaling⁴ w.r.t T, S_T as in the AGN model; recall Remark 18.

4. Experiments

We now provide some empirical results⁵ on synthetic data generated by the AGN and Outliers models. Apart from Algorithms GTRS-DynSync, LTRS-GS-DynSync and GMD-LTRS-DynSync, we also evaluate the block-wise spectral approach described in Remark 3.

4. Note that the rates in (Sadhanala et al., 2016, Theorems 5,6) are obtained for additive Gaussian noise. We believe the dependency on T should be unchanged for more general sub-Gaussian noise models, however this needs to be established formally.
5. The code is available at https://github.com/ErnestoArayaV/dynamic_synch.

Experiment setup. The following remarks are in order.

1. For specified values of n, T and smoothness S_T , we generate a smooth signal $g^* \in \mathcal{C}_{nT}$, with $g_1^*(k) = 1$ for each block k , as follows. Define $\tau' := \min \{ \lfloor 1 + \sqrt{T S_T} \rfloor, T \}$, define $\mathcal{P}_{\tau', n-1}$ as in (11). Generate a Gaussian $u \sim \mathcal{N}(0, I_{(n-1)T})$ and compute

$$u' = \sqrt{T} \frac{\mathcal{P}_{\tau', n-1} u}{\|\mathcal{P}_{\tau', n-1} u\|_2}.$$

Then denoting $\tilde{g}^* := \exp(\iota(0.5\pi)u')$ which is computed entry-wise, we form g^* by prepending each block of \tilde{g}^* with 1. It is a simple exercise to verify that g^* (almost surely) satisfies the smoothness condition in Assumption 2 for the specified S_T , up to a constant.

2. For an estimate $\hat{g} \in \mathcal{C}_{nT}$ of the ground truth $g^* \in \mathcal{C}_{nT}$, we will measure the Root Mean Square Error (RMSE) defined as $\frac{1}{\sqrt{nT}} \|\hat{g} - g^*\|_2$. We remark that an additional \sqrt{n} factor appears here as compared to the quantity being bounded in the analyses earlier.
3. In order to choose the appropriate value of the regularization parameters for the algorithms (λ for **GTRS-DynSync** and τ for **LTRS-GS-DynSync**, **GMD-LTRS-DynSync**), we consider a data-driven grid-search based heuristic. In particular, define a discretization of $[0, T]$, where (i) the first $\lfloor \sqrt{T} \rfloor + 1$ elements of the grid are the numbers $0, 1, \dots, \lfloor \sqrt{T} \rfloor$; (ii) the next five elements are equi-spaced numbers in the interval $[\lfloor \sqrt{T} \rfloor, \lfloor T^{2/3} \rfloor]$; (iii) the last five elements are equi-spaced numbers in the interval $[\lfloor T^{2/3} \rfloor, T]$. For each value on the grid, call it β , we set $\tau = \min \{ \beta + 1, T \}$ and $\lambda = \beta \times \lambda_{\text{scale}}$ for a fixed choice of the scaling parameter $\lambda_{\text{scale}} > 0$. For the corresponding output \hat{g} of each algorithm, we compute a ‘data-fidelity’ value using the input pairwise data, as for the data-fidelity term in (16) (with \tilde{g} therein replaced by \hat{g}). Then we select the optimal β (thus optimal τ and λ) for each algorithm as follows.
 - (a) For **GTRS-DynSync**, the optimal β is the value that leads to the largest data-fidelity.
 - (b) For **LTRS-GS-DynSync** and **GMD-LTRS-DynSync**, the optimal β is chosen to be the value that leads to the maximum change in the slope of the data-fidelity (w.r.t β).

The above data-driven rule for selecting the optimal β is motivated by empirical evidence as shown in Figs. 1 and 2. Notice that the data-fidelity for **LTRS-GS-DynSync** and **GMD-LTRS-DynSync** exhibits a sharp change in the slope at $\beta = 1$ (which is where the RMSE is minimized for these methods). But it is not clear why the data-fidelity keeps increasing beyond $\beta = 1$, this needs further investigation. For **GTRS-DynSync**, notice that the data fidelity peaks at the RMSE-minimizer in the outliers model (see Fig. 1), while for the Wigner model, it peaks at a slightly sub-optimal β (see Fig. 2).

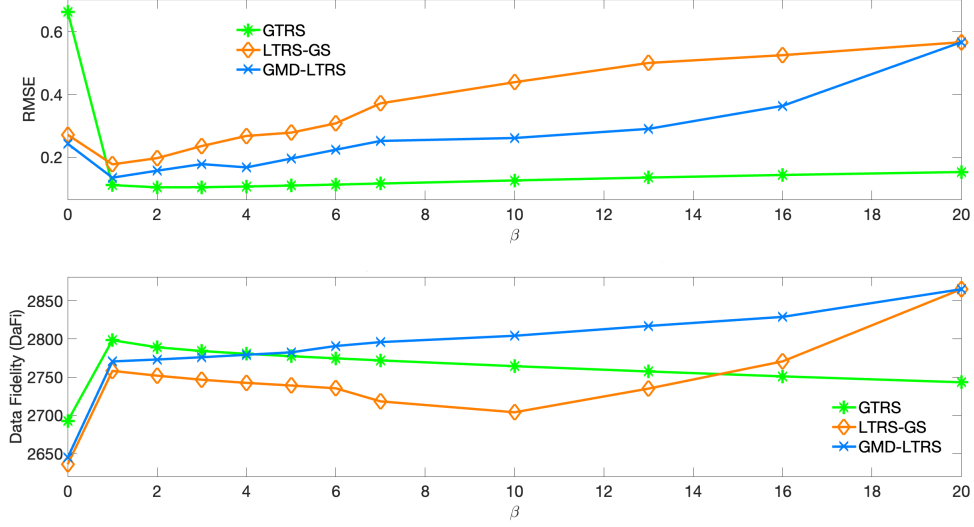


Figure 1: RMSE (top) and Data Fidelity (bottom) for the Outliers model for a single MC run ($n = 30, T = 20, S_T = 1/T, \eta = p = 0.2$). Notice a sharp change in slope of the data-fidelity for LTRS-GS-DynSync and GMD-LTRS-DynSync at $\beta = 1$, which is also where the RMSE is minimized for these two methods. For GTRS-DynSync (with $\lambda_{\text{scale}} = 10$) the data-fidelity peaks at $\beta = 1$ which is where the RMSE is minimized.

Experiment 1: RMSE versus T . In this experiment, we evaluate the RMSE versus T for all the above methods. Throughout, we fix $n = 30$, and vary T from 10 to 100. We perform 20 Monte Carlo (MC) runs for each T . In a single MC run, we randomly generate the ground-truth for a specified smoothness S_T (as described earlier), then generate the noisy pairwise data (as per specified noise model), and then compute the RMSE for each algorithm for the regularization parameters selected by the aforementioned data-fidelity rule. The RMSE values are then averaged over the 20 runs. The results are shown in Figs. 3 for the Wigner model, and Fig. 4 for the Outliers model, for different smoothness conditions S_T . As a sanity check, note that the RMSE of the spectral method remains high for all T . We can make some additional observations.

1. For the AGN model (Fig. 3), we see that GMD-LTRS-DynSync performs the best out of the three methods, while the RMSE for GTRS-DynSync does not shrink to zero with T . The performance of LTRS-GS-DynSync is seen to improve with T , albeit with a larger error than GMD-LTRS-DynSync.
2. For the Outliers model (Fig. 4), the RMSE for all three methods decreases with T . Interestingly, GTRS-DynSync typically performs the best out of the three methods (followed by GMD-LTRS-DynSync and then LTRS-GS-DynSync).

Experiment 2: RMSE versus noise-level. In this experiment, we evaluate the RMSE versus the noise-level for all the above methods. The noise-level is denoted by γ for ease of

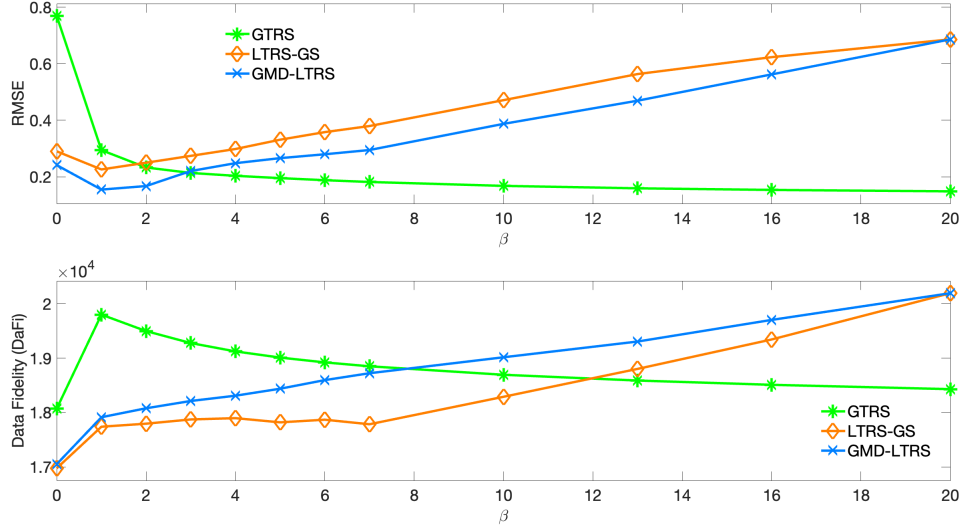


Figure 2: RMSE (top) and Data Fidelity (bottom) for the AGN model for a single MC run ($n = 30, T = 20, S_T = 1/T, \sigma = 3$). Notice a sharp change in slope of the data-fidelity for LTRS-GS-DynSync and GMD-LTRS-DynSync at $\beta = 1$, which is also where the RMSE is minimized for these two methods. For GTRS-DynSync (with $\lambda_{\text{scale}} = 10$) the data-fidelity peaks at $\beta = 1$, but the RMSE is minimized for $\beta = 20$.

exposition, with $\gamma = \sigma > 0$ for the AGN model (varied from 0 to 10), and $\gamma = \eta \in [0, 1]$ for the Outliers model (varied from 0 to 0.8). Throughout, we fix $n = 30, T = 20$, and perform 20 Monte Carlo (MC) runs for each value of γ . In a single MC run, we randomly generate the ground-truth for a specified smoothness S_T (as described earlier), then generate the noisy pairwise data (as per specified noise model), and then compute the RMSE for each algorithm for the regularization parameters selected by the aforementioned data-fidelity rule. The RMSE values are then averaged over the 20 runs. The results are shown in Fig. 5 for the Wigner model, and Fig. 6 for the Outliers model (with $p = 0.2$). The following observations can be made from these plots.

1. For the AGN model (Fig. 5), note that the performance of the spectral method and GTRS-DynSync are similar for low noise levels (σ less than 1), both being worse than the other two methods in this noise regime. The performance of the spectral method degrades for $\sigma > 1$, while that of the other three methods is similar at moderate noise levels ($1 < \sigma < 5$). For larger values of σ , we see that GMD-LTRS-DynSync performs best, while LTRS-GS-DynSync and GTRS-DynSync have similar error values.
2. For the Outliers model (Fig. 6), we observe that the spectral method performs much worse than the other three methods for almost the whole noise regime. We further observe that GTRS-DynSync has a similar performance to GMD-LTRS-DynSync and LTRS-GS-DynSync up to moderate noise levels ($\eta < 0.35$), but its performance degrades for $\eta > 0.35$. On the other hand, GMD-LTRS-DynSync appears to be more robust to high noise levels ($\eta > 0.35$) as compared to the other two methods.

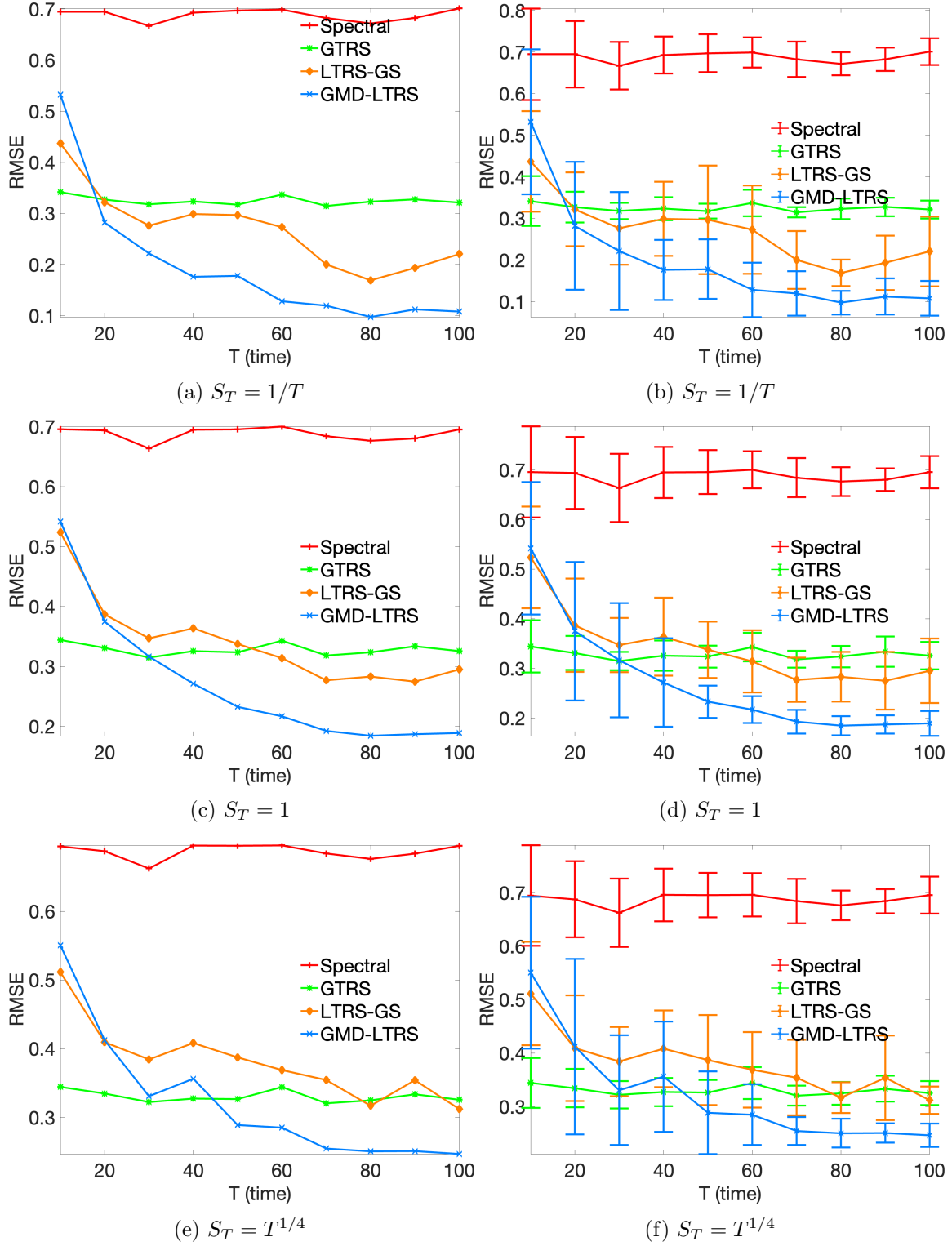


Figure 3: RMSE versus T for AGN model ($n = 30$, $\sigma = 3$) with $S_T \in \{1/T, 1, T^{1/4}\}$, results averaged over 20 MC runs. We choose $\lambda_{\text{scale}} = 10$ for GTRS-DynSync. Results on left show average RMSE, the right panel shows average RMSE \pm standard deviation.

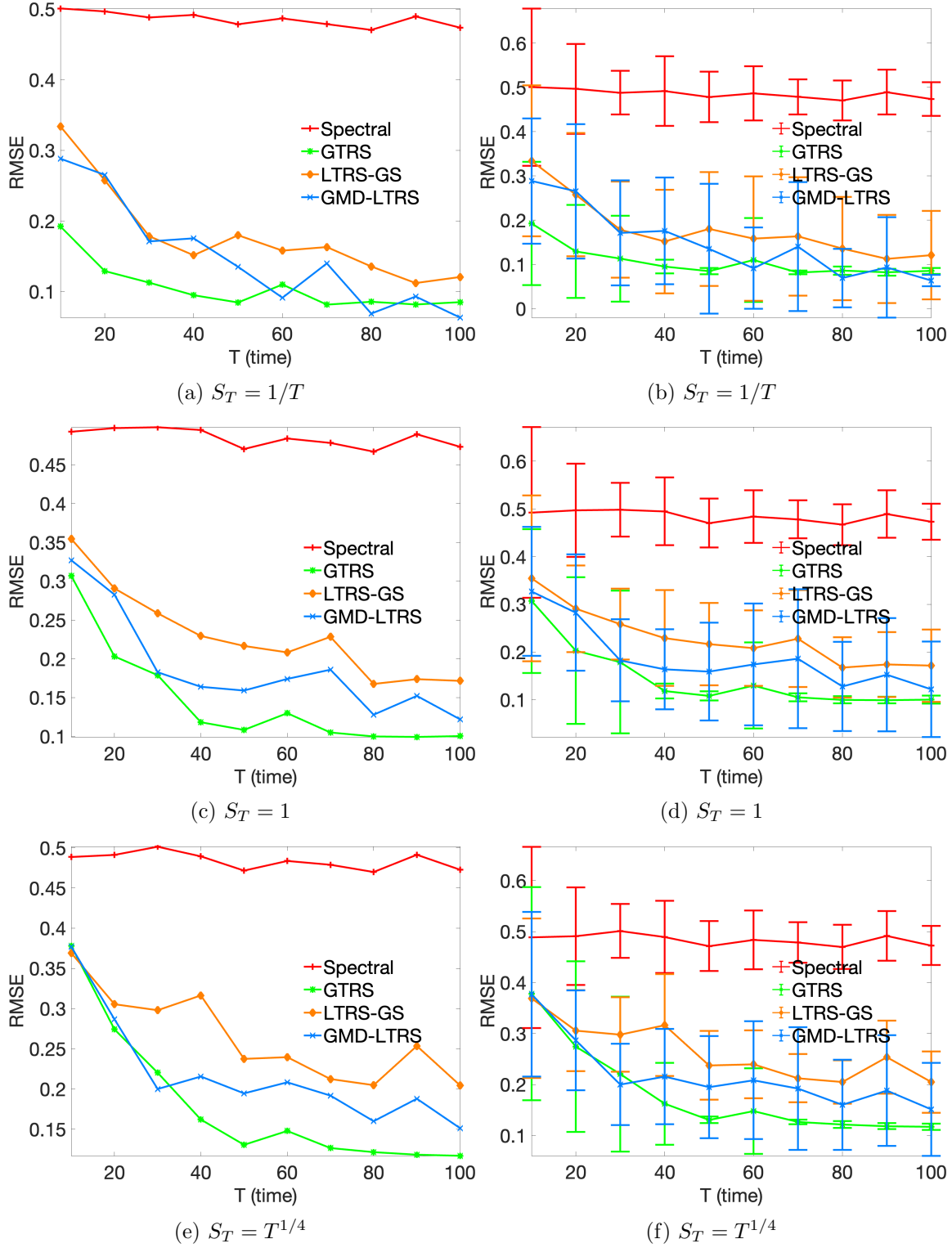


Figure 4: RMSE versus T for Outliers model ($n = 30$, $\eta = p = 0.2$) with $S_T \in \{1/T, 1, T^{1/4}\}$, results averaged over 20 MC runs. We choose $\lambda_{\text{scale}} = 10$ for GTRS-DynSync. Results on left show average RMSE, the right panel shows average RMSE \pm standard deviation.

5. Concluding remarks

In this work we introduced the problem of dynamic angular synchronization, where the underlying latent angles, and the measurement graphs, both evolve with time. Assuming a smoothness condition on the evolution of the angles, with the smoothness measured by the quadratic variation of the angles w.r.t the path graph on T vertices, we proposed three algorithms (namely, **GTRS-DynSync**, **LTRS-GS-DynSync** and **GMD-LTRS-DynSync**) for jointly estimating the angles at all time points. For Algorithm **GMD-LTRS-DynSync**, we derived non-asymptotic error bounds for the AGN and Outliers noise models, and established conditions under which the MSE goes to zero as T increases. In particular, this was shown to occur under much milder conditions than the static setup ($T = 1$). It includes the setting where the measurement graphs are highly sparse and disconnected, and also when the measurement noise is large and can potentially increase with T . We validated our theoretical results with experiments on synthetic data which show that **GMD-LTRS-DynSync** typically outperforms the other two methods.

An interesting direction for future work would be to extend our framework to other groups such as $SO(d)$. Another direction would be to consider evolution models where the measurements are statistically dependent over time – our analysis requires the measurements to be independent across time points.

Acknowledgments

Part of the work was done while E.A was at Inria Lille. The first version of this paper was prepared while H.T was affiliated to Inria Lille. Authors are listed in alphabetical order. Corresponding author: H.T.

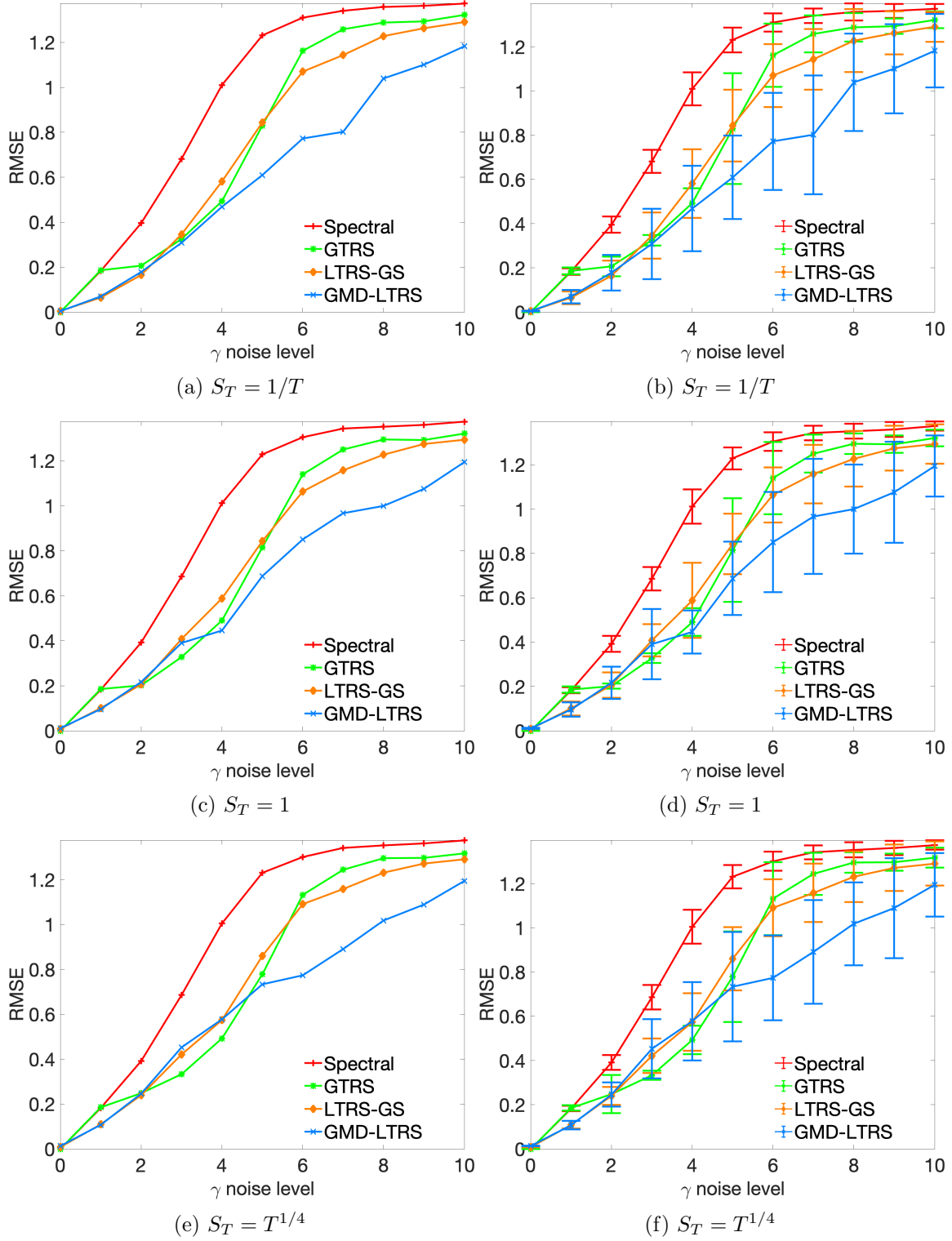


Figure 5: RMSE versus $\gamma(=\sigma)$ for AGN model ($n = 30$, $T = 20$) with $S_T \in \{1/T, 1, T^{1/4}\}$, results averaged over 20 MC runs. We choose $\lambda_{\text{scale}} = 10$ for GTRS-DynSync. Results on left show average RMSE, the right panel shows average RMSE \pm standard deviation.

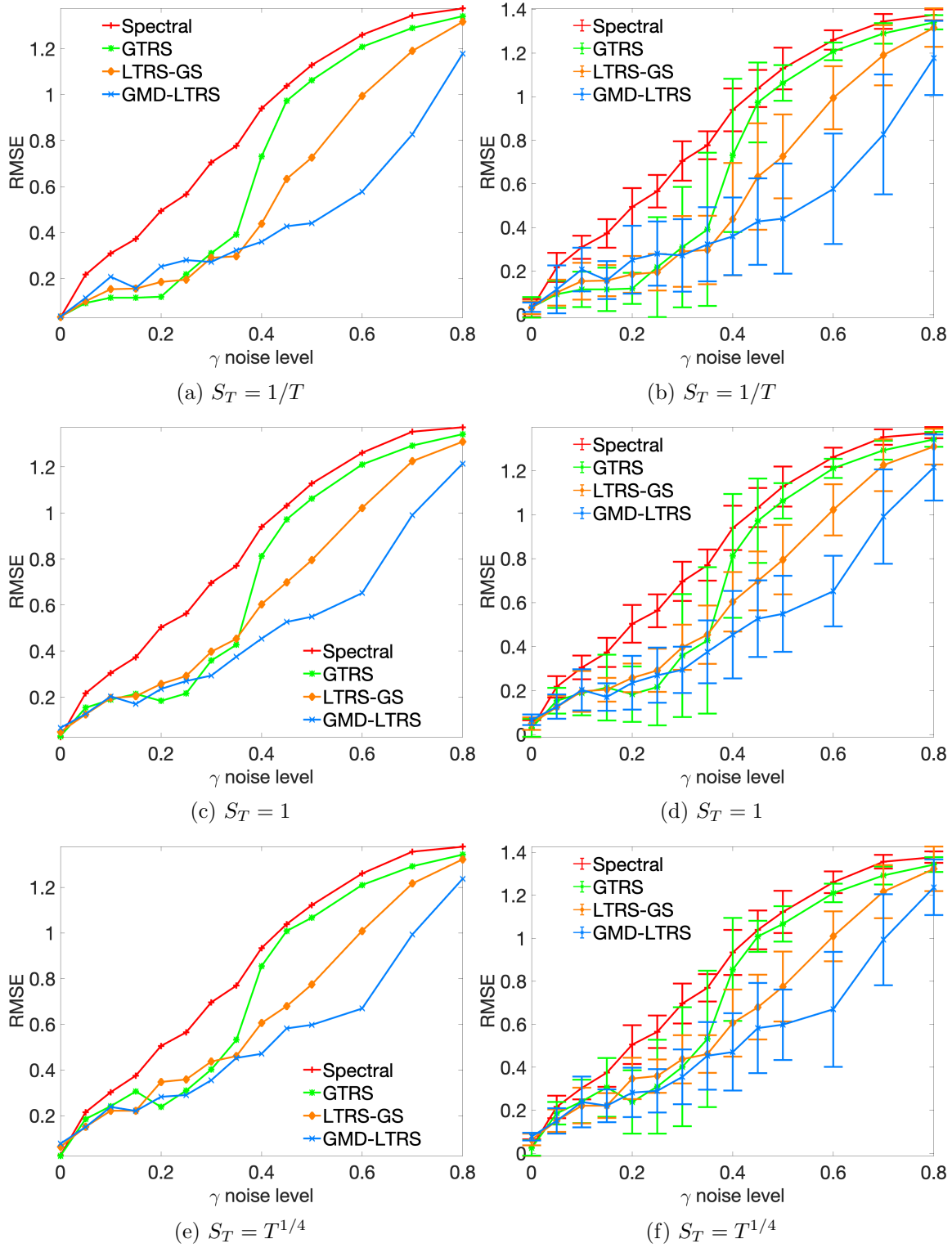


Figure 6: RMSE versus $\gamma(= \eta)$ for Outliers model ($n = 30$, $p = 0.2$, $T = 20$) with $S_T \in \{1/T, 1, T^{1/4}\}$, results averaged over 20 MC runs. We choose $\lambda_{\text{scale}} = 10$ for GTRS-DynSync. Results on left show average RMSE, the right panel shows average RMSE \pm standard deviation.

Appendix A. Proof of Theorem 8

The bound in (25) follows by plugging τ^* in (23), and using the bound $\tau^* \asymp \min \left\{ 1 + \left(\frac{T^2 S_T}{n\sigma^2} \right)^{1/3}, T \right\}$ along with the fact $(a+b)^2 \geq a^2 + b^2$ for $a, b \geq 0$.

To show the second part, observe that if $\left(\frac{T^2 S_T}{n\sigma^2} \right)^{1/3} \geq 1$ then (25) simplifies to

$$\|\hat{G} - \bar{G}^*\|_F^2 \lesssim \max \left\{ n^{5/3} \sigma^{4/3} T^{2/3} S_T^{1/3}, n S_T \right\} \mathbb{1}_{\{\tau^* < T\}} + \min \left\{ \sigma^{4/3} T^{2/3} S_T^{1/3} n^{5/3}, T \right\} + \sigma^2 \log \left(\frac{4}{\delta} \right)$$

and also $\tau^* \asymp \min \left\{ \left(\frac{T^2 S_T}{n\sigma^2} \right)^{1/3}, T \right\}$. Now the bound in (27) follows readily after minor simplifications provided T satisfies (26). We just remark that $T \gtrsim \frac{S_T}{n\sigma^2}$ implies $\tau^* \asymp \left(\frac{T^2 S_T}{n\sigma^2} \right)^{1/3}$, and in particular $\tau^* < T$.

Appendix B. Proof of Lemma 11

Let $H(k) \in \{0, 1\}^{n \times n}$ denote the (symmetric) adjacency matrix of \mathcal{H}_k . Hence $(H_{ij}(k))_{i < j}$ are independent Bernoulli random variables with parameter $p(k)$, and $H_{ii}(k) = 0$ for all i . Also denote $Z(k) \in \mathbb{C}^{n \times n}$ to be a Hermitian matrix with $Z_{ii}(k) = 0$ for each $i \in [n]$, and for each $i < j$

$$Z_{ij}(k) = \begin{cases} g_i^*(k) g_j^*(k)^H & \text{with probability } 1 - \eta, \\ e^{i\varphi_{ij}(k)} & \text{with probability } \eta \end{cases}$$

where $(\varphi_{ij}(k))_{i < j} \stackrel{\text{ind.}}{\sim} \text{Unif}[0, 2\pi)$. Then we can write $A(k) = Z(k) \circ H(k)$, and $A = Z \circ H$ where $Z \in \mathbb{C}^{nT \times n}$ (resp. $H \in \{0, 1\}^{nT \times n}$) is formed by stacking $Z(1), \dots, Z(T)$ (resp. $H(1), \dots, H(T)$). Then we can write $R = Z \circ H - D\bar{G}^*$. Now decompose $Z = Z_1 + Z_2$ where $Z_1(k)$ (resp. $Z_2(k)$) contains the upper triangular (resp. lower triangular) part of $Z(k)$. Write $H = H_1 + H_2$ and $\bar{G}^* = \bar{G}_1^* + \bar{G}_2^*$ in an analogous manner. Then R decomposes as

$$R = \underbrace{(Z_1 \circ H_1 - D\bar{G}_1^*)}_{=: R_1} + \underbrace{(Z_2 \circ H_2 - D\bar{G}_2^*)}_{=: R_2} \quad (42)$$

with the entries of R_1 (resp. R_2) being independent, centered subgaussian random variables.

Now, $\|\mathcal{P}_{\tau,n} R\|_F^2 \leq 2(\|\mathcal{P}_{\tau,n} R_1\|_F^2 + \|\mathcal{P}_{\tau,n} R_2\|_F^2)$ and it is sufficient to bound $\|\mathcal{P}_{\tau,n} R_1\|_F^2$, since the same bound will follow for $\|\mathcal{P}_{\tau,n} R_2\|_F^2$. Decomposing R_1 into its real and complex parts

$$R_1 = \underbrace{\text{Re}(R_1)}_{=: R_{1,\text{R}}} + i \underbrace{\text{Im}(R_1)}_{=: R_{1,\text{I}}},$$

we have $\|\mathcal{P}_{\tau,n} R_1\|_F^2 = \|\mathcal{P}_{\tau,n} R_{1,\text{R}}\|_F^2 + \|\mathcal{P}_{\tau,n} R_{1,\text{I}}\|_F^2$. Again, we focus only on $\|\mathcal{P}_{\tau,n} R_{1,\text{R}}\|_F^2$ as the same bound follows for $\|\mathcal{P}_{\tau,n} R_{1,\text{I}}\|_F^2$ in an analogous manner.

Denote $u \in \mathbb{R}^{n^2 T}$ to be the vector formed by stacking the columns of $R_{1,\text{R}} \in \mathbb{R}^{nT \times n}$. Then we obtain

$$\|\mathcal{P}_{\tau,n} R_{1,\text{R}}\|_F^2 = u^\top (I_n \otimes \mathcal{P}_{\tau,n}) u$$

which we will bound via the Hanson-Wright inequality (Rudelson and Vershynin, 2013, Thm. 1.1). To this end, we first derive an upper bound on $\mathbb{E}[u^\top (I_n \otimes \mathcal{P}_{\tau,n})u]$ as follows. Denoting $Z_{1,R}(k)$ (resp. $\bar{G}_{1,R}^*(k)$) to be the real part of $Z_1(k)$ (resp. $\bar{G}_1^*(k)$), we obtain

$$(R_{1,R}(k))_{ij} = (Z_{1,R}(k))_{ij}(H_1(k))_{ij} - (1 - \eta)p(k)(\bar{G}_{1,R}^*(k))_{ij}.$$

Hence we can bound $\mathbb{E}[(R_{1,R}(k))_{ij}^2]$ for all $i \neq j \in [n]$ as follows.

$$\begin{aligned} \mathbb{E}[(R_{1,R}(k))_{ij}^2] &= p(k)\mathbb{E}[(Z_{1,R}(k))_{ij} - (1 - \eta)p(k)(\bar{G}_{1,R}^*(k))_{ij}]^2 + (1 - p(k))p^2(k)(1 - \eta)^2(\bar{G}_{1,R}^*(k))_{ij}^2 \\ &\lesssim p(k)(\text{Var}((Z_{1,R}(k))_{ij}) + (1 - \eta)^2(1 - p(k))^2) + (1 - p(k))p^2(k)(1 - \eta)^2. \end{aligned}$$

It is not difficult to verify that $\text{Var}((Z_{1,R}(k))_{ij}) \lesssim \eta$, which upon plugging in the above bound readily implies

$$\mathbb{E}[(R_{1,R}(k))_{ij}^2] \lesssim \underbrace{p_{\max}\eta + (1 - \eta)^2V + p_{\max}V(1 - \eta)}_{=: f(\eta, p_{\max}, V)}.$$

Thus we have shown that

$$\mathbb{E}[u^\top (I_n \otimes \mathcal{P}_{\tau,n})u] \leq Cn^2\tau f(\eta, p_{\max}, V)$$

for some constant $C > 0$.

We next bound $\|(R_{1,R}(k))_{ij}\|_{\psi_2}$ for all $i \neq j$ as follows. To begin with, we have

$$\begin{aligned} \|(R_{1,R}(k))_{ij}\|_{\psi_2} &= \|(Z_{1,R}(k))_{ij}(H_1(k))_{ij} - (1 - \eta)p(k)(\bar{G}_{1,R}^*(k))_{ij}\|_{\psi_2} \\ &\leq \|(Z_{1,R}(k))_{ij} - (1 - \eta)(\bar{G}_{1,R}^*(k))_{ij}\|_{\psi_2} + \|(H_1(k))_{ij} - p(k)\|_{\psi_2} \end{aligned} \quad (43)$$

where we used triangle inequality and the fact $(H_1(k))_{ij}, |(\bar{G}_{1,R}^*(k))_{ij}| \leq 1$ (uniformly over i, j, k). Now the second term in (43) is bounded as

$$\|(H_1(k))_{ij} - p(k)\|_{\psi_2} \leq Q(p(k)) \leq \max_{k \in [T]} Q(p(k)).$$

To bound the first term in (43), we begin by writing

$$(Z_{1,R}(k))_{ij} - (1 - \eta)(\bar{G}_{1,R}^*(k))_{ij} = B_{ij}(k) \cos(\varphi_{ij}(k)) - (B_{ij}(k) - \eta) \cos(\theta_i(k) - \theta_j(k))$$

where $B_{ij}(k)$ is a Bernoulli random variable with parameter η . Then it is not difficult to verify

$$\|B_{ij}(k) \cos(\varphi_{ij}(k)) - (B_{ij}(k) - \eta) \cos(\theta_i(k) - \theta_j(k))\|_{\psi_2} \lesssim Q(\eta) + \eta.$$

Thus, we have shown that

$$\|(R_{1,R}(k))_{ij}\|_{\psi_2} \lesssim Q(\eta) + \eta + \max_{k \in [T]} Q(p(k)) =: Q.$$

Upon applying the above bounds in conjunction with the Hanson Wright inequality, we readily obtain that with probability at least $1 - \delta$,

$$\begin{aligned} \|\mathcal{P}_{\tau,n} R_{1,R}\|_F^2 &\leq C \left(n^2 \tau f(\eta, p_{\max}, V) + Q^2 n \sqrt{\tau \log(1/\delta)} + Q^2 \log(1/\delta) \right) \\ &\leq C' \left(n^2 \tau f(\eta, p_{\max}, V) + Q^2 n \sqrt{\tau \log(1/\delta)} \right) \end{aligned}$$

where we used $\sqrt{\log(1/\delta)} \leq \log(1/\delta)$ if $\delta \in (0, c)$ for a suitably small constant c . The proof is completed by using the same error bound on the other terms discussed earlier, applying a union bound, and adjusting the constants.

Appendix C. Proof of Theorem 13

Using (31) and Lemma 11 leads to the bound

$$\|\widehat{G} - D\overline{G}^*\|_F^2 \leq C_1 \frac{\mu \bar{d}^2 n T^2 S_T}{\tau^2} \mathbb{1}_{\tau < T} + C_2 \left(n^2 \tau f(\eta, p_{\max}, V) + Q^2 n \sqrt{\tau} \log \left(\frac{1}{\delta} \right) \right).$$

Following the same argument as in Section 3.1, we arrive at the choice $\tau = \tau^*$ which upon plugging in the RHS of the above bound leads to the stated error bound in part (i) (after minor simplification).

For part (ii), we simply note that if $1 \leq \left(\frac{\mu \bar{d}^2 T^2 S_T}{n \tilde{f}(\eta, p_{\max}, V, Q, \delta)} \right)^{1/3} \lesssim T$ then

$$\tau^* \asymp \left(\frac{\mu \bar{d}^2 T^2 S_T}{n \tilde{f}(\eta, p_{\max}, V, Q, \delta)} \right)^{1/3},$$

and consequently, (33) implies

$$\begin{aligned} \|\widehat{G} - D\overline{G}^*\|_F^2 &\lesssim n^{5/3} \tilde{f}^{2/3}(\eta, p_{\max}, V, Q, \delta) (\mu \bar{d}^2)^{1/3} T^{2/3} S_T^{1/3} \\ &\quad + \max \left\{ \tilde{f}^{2/3}(\eta, p_{\max}, V, Q, \delta) n^{5/3} (\mu \bar{d}^2)^{1/3} T^{2/3} S_T^{1/3}, \mu \bar{d}^2 n S_T \right\}. \end{aligned}$$

We then arrive at the simplified bound in (35) provided $T \geq \frac{\mu \bar{d}^2 S_T}{n \tilde{f}^{2/3}(\eta, p_{\max}, V, Q, \delta)}$.

Appendix D. Proof of Proposition 16

We will use the following useful inequality from (Liu et al., 2017).

Proposition 21 ((Liu et al., 2017)) *For any $q \in [1, \infty]$, $w \in \mathbb{C}^n$ and $g \in \mathcal{C}_n$, we have*

$$\|\mathcal{P}_{\mathcal{C}_n}(w) - g\|_q \leq 2\|w - g\|_q.$$

Proof [Proof of Proposition 16] For any $k \in [T]$ we obtain

$$\begin{aligned} \|\widehat{g}(k) - g^*(k)\|_2 &= \left\| \mathcal{P}_{\mathcal{C}_n} \left(\widehat{g}'(k) \right) e^{-\iota \hat{\theta}_1(k)} - g^*(k) \right\|_2 \\ &= \left\| \mathcal{P}_{\mathcal{C}_n} \left(\sqrt{n} \widehat{g}'(k) e^{-\iota \hat{\theta}_1(k)} \right) - g^*(k) \right\|_2 \\ &\leq 2 \left\| \sqrt{n} \widehat{g}'(k) e^{-\iota \hat{\theta}_1(k)} - g^*(k) \right\|_2 \quad (\text{using Proposition 21}) \\ &\leq 2 \left\| \sqrt{n} \widehat{g}'(k) e^{-\iota \phi(k)} - g^*(k) \right\|_2 + 2\sqrt{n} \left\| \widehat{g}'(k) \left(e^{-\iota \hat{\theta}_1(k)} - e^{-\iota \phi(k)} \right) \right\|_2, \quad (44) \end{aligned}$$

where (44) follows from the triangle inequality. Using (39), (37), we bound the first term as

$$\left\| \sqrt{n} \widehat{g}'(k) e^{-\iota \phi(k)} - g^*(k) \right\|_2 \leq C \frac{\|\widehat{G}'(k) - s(k) \overline{G}^*(k)\|_F}{s(k) \sqrt{n}} \leq C \frac{\|\widehat{G}(k) - s(k) \overline{G}^*(k)\|_F}{s(k) \sqrt{n}}. \quad (45)$$

For the second term, we note that

$$\begin{aligned}
 \sqrt{n} \left\| \widehat{g}'(k) \left(e^{-\iota \hat{\theta}_1(k)} - e^{-\iota \phi(k)} \right) \right\|_2 &= \sqrt{n} \left| e^{\iota \hat{\theta}_1(k)} - e^{\iota \phi(k)} \right| && (\text{since } \|\widehat{g}'(k)\|_2 = 1) \\
 &= \sqrt{n} |\mathcal{P}_{\mathcal{C}_1}(e^{\iota \hat{\theta}_1(k)}) - e^{\iota \phi(k)}| \\
 &= \sqrt{n} |\mathcal{P}_{\mathcal{C}_1}(\underbrace{\sqrt{n} |\widehat{g}'_1(k)| e^{\iota \hat{\theta}_1(k)}}_{=\widehat{g}'_1(k)}) - e^{\iota \phi(k)}| \\
 &\leq 2\sqrt{n} |\sqrt{n} \widehat{g}'_1(k) - e^{\iota \phi(k)}| && (\text{using Proposition 21}) \\
 &\leq 2n \left\| \widehat{g}'(k) - e^{\iota \phi(k)} \frac{g^*(k)}{\sqrt{n}} \right\|_2 && (\text{since } g_1^*(k) = 1) \\
 &\leq C \frac{\|\widehat{G}(k) - s(k) \overline{G}^*(k)\|_F}{s(k)} && (\text{using (39), (37)}). \quad (46)
 \end{aligned}$$

Plugging (45), (46) in (44) leads to the stated bound. \blacksquare

Appendix E. Additional simulations

E.1 RMSE versus noise-level

We show additional plots for RMSE versus noise-level (γ) for both the AGN (Fig. 7) and Outliers (Fig. 8) noise models but for $T = 100$ (recall we considered $T = 20$ in Figs. 5 and 6 in Section 4). We see for the AGN model that out of the three methods, **GMD-LTRS-DynSync** is most robust for moderate to high levels of noise. For the Outliers model, we find that **GTRS-DynSync** performs the best for low level of noise, while **GMD-LTRS-DynSync** is again the best method for moderate to high levels of noise.

E.2 Simulations for a Projected Power Method (PPM)

We now discuss an iterative projected power method for solving (16) that needs to be initialized with a “seed” $\tilde{g}^{(0)} \in \mathcal{C}_{(n-1)T}$ that is sufficiently close to the ground truth \tilde{g}^* . This method, also referred to as the generalized power method in the literature, has had success in tackling several non-convex problems such as orthogonal procrustes (Ling, 2024), multireference alignment (Chen and Candès, 2016), graph clustering (Wang et al., 2022), graph matching (Araya et al., 2024) and group synchronization (Boumal, 2016; Gao and Zhang, 2023; Liu et al., 2017), to name a few.

The version that we consider for dynamic synchronization is outlined in Algorithm 5, dubbed **PPM-DynSync**. At its core, it involves iteratively refining the seed by sequentially performing the following operations.

- **Power step.** Pre-multiply the current iterate by the gradient of the objective function in (16).
- **Projection step.** The result is projected onto the set $\mathcal{C}_{(n-1)T}$ using the operator $\mathcal{P}_{\mathcal{C}_{(n-1)T}}$.

For initialization, we will employ **GTRS-DynSync**, **GMD-LTRS-DynSync**, **LTRS-GS-DynSync** and the naive spectral algorithm considered in the experiments (Section 4). For **PPM-DynSync**, the parameter λ is chosen via the same grid search procedure as for **GTRS-DynSync** (see Section 4); we chose $\lambda_{scale} = 10$ for both **PPM-DynSync** and **GTRS-DynSync**. Surprisingly, we found that **PPM-DynSync** does not improve the performance of the seed, and in general worsens the estimation error. While a more detailed investigation of **PPM-DynSync** is left for future work, we illustrate its performance in Fig. 9 for the AGN model. Similar behaviour was observed for the Outliers model, and this was the case for different choices of λ_{scale} for **PPM-DynSync**.

Algorithm 5 Projected Power Method for Dynamic Synchronization (**PPM-DynSync**)

- 1: **Input:** Matrix $A \in \mathbb{C}^{nT \times n}$ of observations, parameter $\lambda > 0$, number of iterations S , initial point $\tilde{g}^{(0)}$.
 - 2: Form $\tilde{A}_{block} = \text{blkdiag}(\tilde{A}(1), \dots, \tilde{A}(T)) \in \mathbb{C}^{(n-1)T \times (n-1)T}$.
 - 3: **for** $s = 1, 2, \dots, S$ **do**
 - 4: $\tilde{g}^{(s)} = \mathcal{P}_{\mathcal{C}_{(n-1)T}} \left((\tilde{A}_{block} - \lambda M M^\top \otimes I_{n-1}) \tilde{g}^{(s-1)} + b \right)$.
 - 5: **end for**
 - 6: **for** $k = 1, \dots, T$ **do**
 - 7: Define $\hat{g}(k) = \begin{bmatrix} 1 \\ \tilde{g}^{(S)}(k) \end{bmatrix}$.
 - 8: **end for**
 - 9: **Output:** Estimator $\hat{g} = \text{concat}(\hat{g}(1), \dots, \hat{g}(T)) \in \mathcal{C}_{nT}$.
-

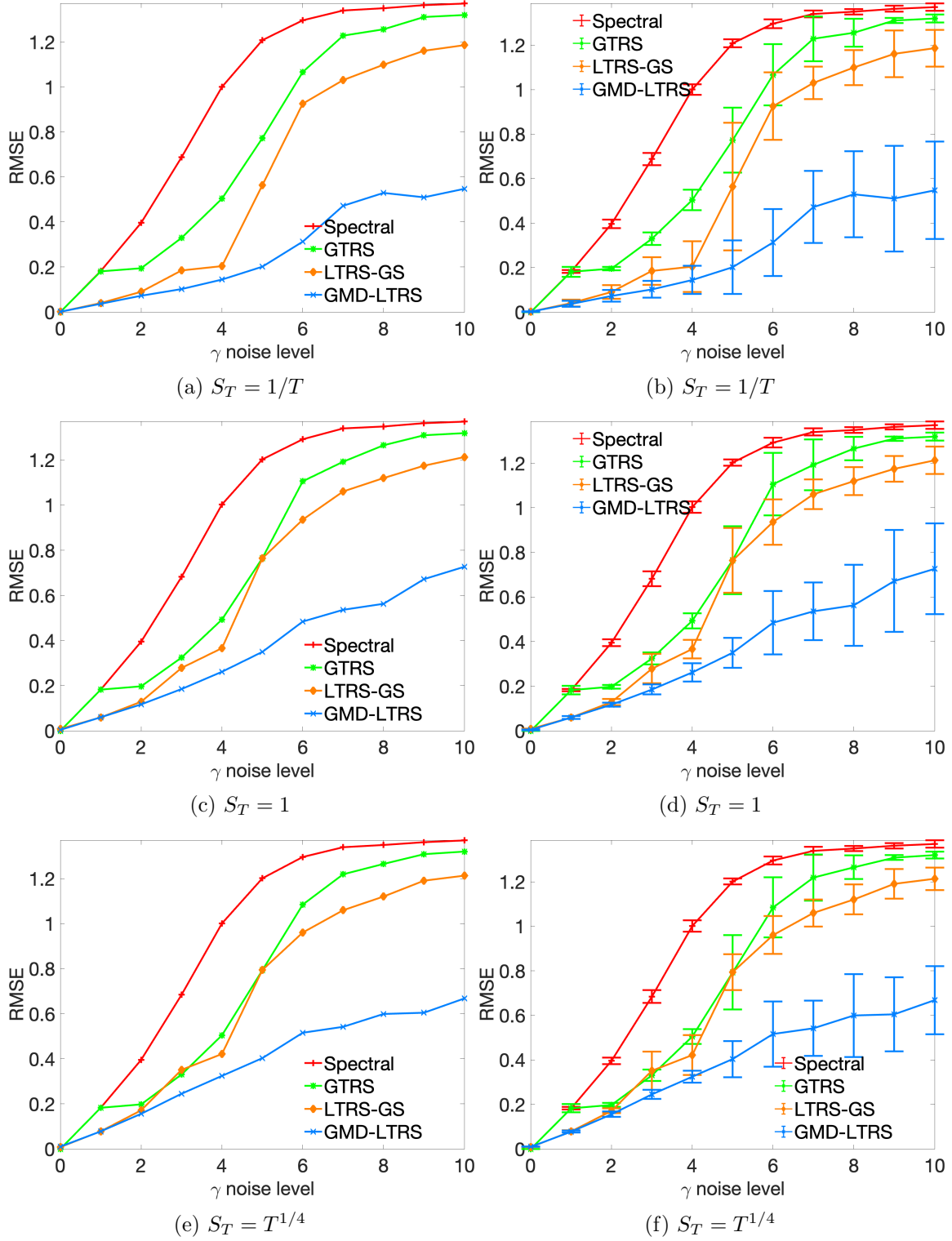


Figure 7: RMSE versus $\gamma(= \sigma)$ for the AGN model ($n = 30$, $T = 100$) with $S_T \in \{1/T, 1, T^{1/4}\}$, results averaged over 20 MC runs. We choose $\lambda_{\text{scale}} = 10$ for GTRS-DynSync. Results on left show average RMSE, the right panel shows average RMSE \pm standard deviation.

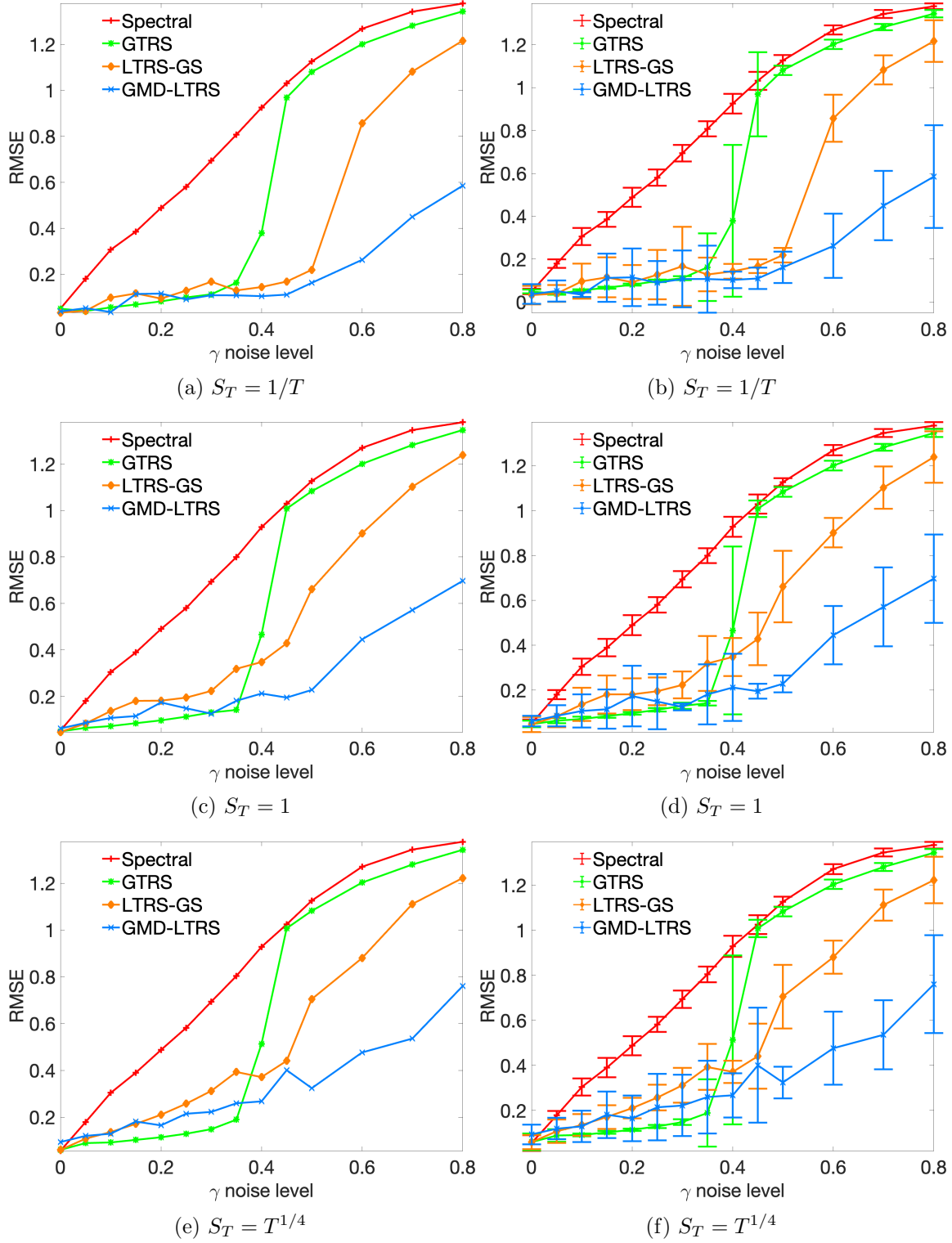


Figure 8: RMSE versus $\gamma(=\eta)$ for the Outliers model ($n = 30, p = 0.2, T = 100$) with $S_T \in \{1/T, 1, T^{1/4}\}$, results averaged over 20 MC runs. We choose $\lambda_{\text{scale}} = 10$ for GTRS-DynSync. Results on left show average RMSE, the right panel shows average RMSE \pm standard deviation.

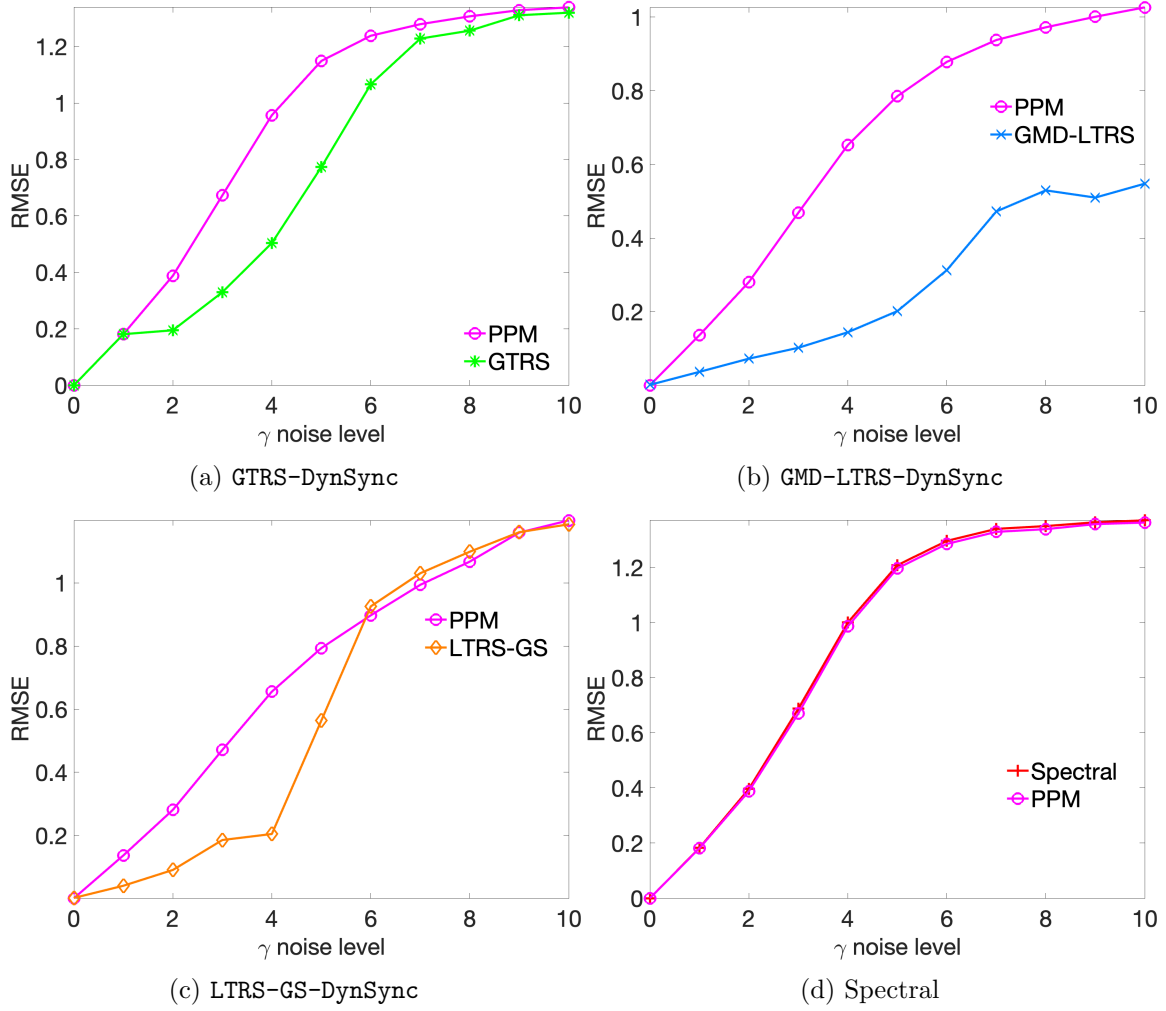


Figure 9: RMSE versus $\gamma(= \sigma)$ for PPM-DynSync for the AGN model ($n = 30$, $T = 100$) with $S_T = 1/T$ and different initializations. Results averaged over 20 MC runs. We choose $\lambda_{\text{scale}} = 10$ for GTRS-DynSync and PPM-DynSync.

References

- Satoru Adachi, Satoru Iwata, Yuji Nakatsukasa, and Akiko Takeda. Solving the trust-region subproblem by a generalized eigenvalue problem. *SIAM Journal on Optimization*, 27(1): 269–291, 2017.
- Ernesto Araya, Eglantine Karlé, and Hemant Tyagi. Dynamic ranking and translation synchronization. *Information and Inference: A Journal of the IMA*, 12(3):2224–2266, 2023.
- Ernesto Araya, Guillaume Braun, and Hemant Tyagi. Seeded Graph Matching for the Correlated Gaussian Wigner Model via the Projected Power Method. *Journal of Machine Learning Research*, 25(5):1–43, 2024.
- M. Arie-Nachimson, S. Z. Kovalsky, I. Kemelmacher-Shlizerman, A. Singer, and R. Basri. Global motion estimation from point matches. In *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization Transmission*, pages 81–88, 2012.
- Afonso S. Bandeira and Ramon van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Ann. Probab.*, 44(4):2479–2506, 07 2016.
- Afonso S. Bandeira, Nicolas Boumal, and Amit Singer. Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. *Mathematical Programming*, 163(1-2):145–167, 2017.
- Afonso S. Bandeira, Amelia Perry, and Alexander S. Wein. Notes on computational-to-statistical gaps: predictions using statistical physics. *arXiv:1803.11132*, 2018.
- Heejong Bong, Wanshan Li, Shamindra Shrotriya, and Alessandro Rinaldo. Nonparametric Estimation in the Dynamic Bradley-Terry Model. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, pages 3317–3326, 2020.
- Nicolas Boumal. Nonconvex phase synchronization. *SIAM Journal on Optimization*, 26(4): 2355–2377, 2016.
- A.E. Brouwer and W.H. Haemers. *Spectra of Graphs*. New York, NY, 2012.
- Manuela Cattelan, Cristiano Varin, and David Firth. Dynamic Bradley–Terry modelling of sports tournaments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(1):135–150, 2013.
- Jianrui Chen, Danwei Liu, Fei Hao, and Hua Wang. Community detection in dynamic signed network: an intimacy evolutionary clustering algorithm. *Journal of Ambient Intelligence and Humanized Computing*, 11:891–900, 2020.
- Yuxin Chen and Emmanuel J. Candès. The projected power method: An efficient algorithm for joint alignment from pairwise differences. *Communications on Pure and Applied Mathematics*, 71(8):1648–1714, 2016.
- Mihai Cucuringu. Synchronization over Z_2 and community detection in signed multiplex networks with constraints. *Journal of Complex Networks*, 3(3):469–506, 2015.

- Mihai Cucuringu. Sync-Rank: Robust Ranking, Constrained Ranking and Rank Aggregation via Eigenvector and Semidefinite Programming Synchronization. *IEEE Transactions on Network Science and Engineering*, 3(1):58–79, 2016.
- Mihai Cucuringu, Yaron Lipman, and Amit Singer. Sensor network localization by eigenvector synchronization over the Euclidean group. *ACM Trans. Sen. Netw.*, 8(3):19:1–19:42, 2012a.
- Mihai Cucuringu, Amit Singer, and David Cowlburn. Eigenvector synchronization, graph rigidity and the molecule problem. *Information and Inference*, 1(1):21–67, 2012b.
- Alexandre d’Aspremont, Mihai Cucuringu, and Hemant Tyagi. Ranking and synchronization from pairwise measurements via SVD. *Journal of Machine Learning Research (JMLR)*, 22(19):1–63, 2021.
- Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.
- Carmeline J. Dsilva, Bomyi Lim, Hang Lu, Amit Singer, Ioannis G. Kevrekidis, and Stanislav Y. Shvartsman. Temporal ordering and registration of images in studies of developmental dynamics. *Development*, 142(9):1717–1724, 2015.
- Ludwig Fahrmeir and Gerhard Tutz. Dynamic stochastic models for time-dependent ordered paired comparison systems. *Journal of the American Statistical Association*, 89(428):1438–1449, 1994.
- Chao Gao and Anderson Y Zhang. Optimal orthogonal group synchronization and rotation group synchronization. *Information and Inference: A Journal of the IMA*, 12:591–632, 2023.
- A. Giridhar and P. R. Kumar. Distributed clock synchronization over wireless networks: Algorithms and analysis. In *45th IEEE Conference on Decision and Control*, pages 4915–4920, 2006.
- Mark E. Glickman and Hal S. Stern. A state-space model for national football league scores. *Journal of the American Statistical Association*, 93(441):25–35, 1998.
- William W. Hager. Minimizing a quadratic over a sphere. *SIAM Journal on Optimization*, 12(1):188–208, 2001.
- Yixuan He, Gesine Reinert, David Wipf, and Mihai Cucuringu. Robust Angular Synchronization Using Directed Graph Neural Networks. *to appear at ICLR 2024; arXiv:2310.05842*, 2024.
- Xiangru Huang, Zhenxiao Liang, Chandrajit Bajaj, and Qixing Huang. Translation synchronization via truncated least squares. *Advances in Neural Information Processing Systems*, 30, 2017.
- Pierre-Emmanuel Jabin and Stéphane Junca. A continuous model for ratings. *SIAM Journal on Applied Mathematics*, 75(2):420–442, 2015.

- Eglantine Karlé and Hemant Tyagi. Dynamic Ranking with the BTL Model: A Nearest Neighbor based Rank Centrality Method. *Journal of Machine Learning Research*, 24 (269):1–57, 2023.
- Shuang Li and Michael B Wakin. Recovery guarantees for time-varying pairwise comparison matrices with non-transitivity. *arXiv:2106.09151*, 2021.
- Shuyang Ling. Generalized Orthogonal Procrustes Problem under Arbitrary Adversaries. *arXiv:2106.15493*, 2024.
- H. Liu, Man-Chung Yue, and A. Man-Cho So. On the estimation performance and convergence rate of the generalized power method for phase synchronization. *SIAM Journal on Optimization*, 27(4):2426–2446, 2017.
- Huikang Liu, Man-Chung Yue, and Anthony Man-Cho So. A unified approach to synchronization problems over subgroups of the orthogonal group. *Applied and Computational Harmonic Analysis*, 66:320–372, 2023.
- Michael J. Lopez, Gregory J. Matthews, and Benjamin S. Baumer. How often does the best team win? A unified approach to understanding randomness in North American sport. *The Annals of Applied Statistics*, 12(4):2483–2516, 2018.
- Lucas Maystre, Victor Kristof, and Matthias Grossglauser. Pairwise comparisons with flexible time-dynamics. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1236–1246, 2019.
- Sahand Negahban, Sewoong Oh, and Devavrat Shah. Rank centrality: Ranking from pairwise comparisons. *Operations Research*, 65(1):266–287, 2017.
- E. Ostrovsky and L. Sirota. Exact value for subgaussian norm of centered indicator random variable. *arXiv:1405.6749*, 2014.
- Amelia Perry, Alexander S. Wein, Afonso S. Bandeira, and Ankur Moitra. Message-Passing Algorithms for Synchronization Problems over Compact Groups. *Communications on Pure and Applied Mathematics*, 71(11):2275–2322, 2018.
- Mark Rudelson and Roman Vershynin. Hanson-Wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18:1 – 9, 2013.
- Veeranjaneyulu Sadhanala, Yu-Xiang Wang, and Ryan J. Tibshirani. Total variation classes beyond 1d: Minimax rates, and the limitations of linear smoothers. NIPS’16, page 3521–3529, 2016.
- Yoel Shkolnisky and Amit Singer. Viewing direction estimation in cryo-EM using synchronization. *SIAM journal on imaging sciences*, 5(3):1088–1110, 2012.
- Amit Singer. Angular synchronization by eigenvectors and semidefinite programming. *Appl. Comput. Harmon. Anal.*, 30(1):20–36, 2011.
- Roman Vershynin. *Introduction to the non-asymptotic analysis of random matrices*, page 210–268. Cambridge University Press, 2012.

- Peng Wang, Zirui Zhou, and Anthony Man-Cho So. Non-convex exact community recovery in stochastic block model. *Mathematical Programming*, 195(1):1–37, 2022.
- Shuzhong Zhang and Yongwei Huan. Complex quadratic optimization and semidefinite programming. *SIAM Journal on Optimization*, 16(3):871–890, 2006.
- Yiqiao Zhong and Nicolas Boumal. Near-optimal bounds for phase synchronization. *SIAM Journal on Optimization*, 28(2):989–1016, 2018.
- Onur Özyeşil, Vladislav Voroninski, Ronen Basri, and Amit Singer. A survey of structure from motion. *Acta Numerica*, 26:305–364, 2017.